

The Global Impact of Open Data

**Key Findings from Detailed Case Studies
Around the World**



**Andrew Young
& Stefaan Verhulst**



San Jose



London



Beijing



New York



Singapore

Strata+ Hadoop WORLD

Make Data Work
strataconf.com

Presented by O'Reilly and Cloudera, Strata + Hadoop World helps you put big data, cutting-edge data science, and new business fundamentals to work.

- Learn new business applications of data technologies
- Develop new skills through trainings and in-depth tutorials
- Connect with an international community of thousands who work with data

The Global Impact of Open Data

*Key Findings from Detailed Case
Studies Around the World*

Andrew Young and Stefaan Verhulst



Supported by:



Beijing • Boston • Farnham • Sebastopol • Tokyo

O'REILLY®

The Global Impact of Open Data

by Andrew Young and Stefaan Verhulst

Copyright © 2016 O'Reilly Media Inc. All rights reserved.

Printed in the United States of America.

Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

O'Reilly books may be purchased for educational, business, or sales promotional use. Online editions are also available for most titles (<http://safaribooksonline.com>). For more information, contact our corporate/institutional sales department: 800-998-9938 or corporate@oreilly.com.

Editor: Shannon Cutt

Interior Designer: David Futato

Production Editor: Nicholas Adams

Cover Designer: Randy Comer

Copyeditor: Octal Publishing, Inc.

Illustrator: Rebecca Demarest

September 2016: First Edition

Revision History for the First Edition

2016-09-13: First Release

The O'Reilly logo is a registered trademark of O'Reilly Media, Inc. *The Global Impact of Open Data*, the cover image, and related trade dress are trademarks of O'Reilly Media, Inc.

While the publisher and the authors have used good faith efforts to ensure that the information and instructions contained in this work are accurate, the publisher and the authors disclaim all responsibility for errors or omissions, including without limitation responsibility for damages resulting from the use of or reliance on this work. Use of the information and instructions contained in this work is at your own risk. If any code samples or other technology this work contains or describes is subject to open source licenses or the intellectual property rights of others, it is your responsibility to ensure that your use thereof complies with such licenses and/or rights.

978-1-491-96467-5

[LSI]

Table of Contents

Acknowledgements.....	ix
Executive Summary.....	xi

Part I. Open Data's Impact—Lessons Learned

1. Understanding the Impact of Open Data.....	3
Introduction	3
I. What Is Open Data?	5
II. The Case Studies	6
III. What Is the Impact of Open Data on People's Lives?	14
IV. What Are the Enabling Conditions that Significantly Enhance the Impact of Open Data?	17
V. What Are the Challenges to Open Data Making an Impact?	20
VI. Recommendations: Toward a Next Generation Open- Data Roadmap	25
Key Remaining Questions	37

Part II. Case Studies: Improving Government

2. Brazil's Open Budget Transparency Portal.....	41
Making Public How Public Money Is Spent	
Context and Background	42

Project Description and Inception	46
Impact	50
Challenges	54
Looking Forward	57
3. Openaid in Sweden.....	59
Enhanced Transparency and Accountability	
Context and Background	60
Product Description and Inception	62
Impact	66
Challenges	69
Looking Forward	72
4. Open Contracting and Procurement in Slovakia.....	75
Establishing Trust in Government	
Context and Background	77
Impact	83
Challenges	90
Looking Forward	94
5. Indonesia's Kawal Pemilu.....	97
Elections: Free, Fair, and Open Data	
Context and Background	98
Project Description and Inception	102
Impact	105
Challenges	108
Looking Forward	110
6. Denmark's Open Address Data Set.....	113
Consolidating and Freeing-up Address Data	
Context and Background	114
Project Description and Inception	116
Impact	121
Challenges	129
Looking Forward	132

7. Opening Canada's T3010 Charity Information Return Data.....	135
Accountability of Charities through Open Data	
Context and Background	136
Project Description and Inception	138
Impact	141
Challenges	149
Looking Forward	156

Part III. Case Studies: Empowering Citizens

8. Open Education Information in Tanzania.....	161
A Tale of Two Dashboards	
Context and Background	162
Product Description and Inception	165
Impact	168
Challenges	172
Looking Forward	174
9. Kenya's Open Duka.....	177
Open Data for Transactional Transparency	
Context and Background	178
Project Description and Inception	182
Impact	187
Challenges	188
Looking Forward	190
10. Mexico's Mejora Tu Escuela.....	195
Empowering Citizens to Make Data-Driven	
Context and Background	196
Product Description and Inception	200
Impact	203
Challenges	208
Looking Forward	211
11. Uruguay's A Tu Servicio.....	215
Empowering Citizens to Make Data-Driven	
Context and Background	216

Product Description and Inception	218
Impact	220
Challenges	225
Looking Forward	227

Part IV. Case Studies: Creating Opportunity

12. Great Britain's Ordnance Survey.....	231
A Clash of Business Models	
Context and Background	232
Project Description and Inception	237
Impact	243
Challenges	250
Looking Forward	258
13. United States' New York City Business Atlas.....	261
Leveling the Playing Field	
Context and Background	262
Project Description and Inception	265
Impact	271
Challenges	275
Looking Forward	277
14. United States' NOAA: Opening Up Global Weather Data in Collaboration with Businesses.....	281
Creating a New Industry Through Access to Weather Data	
Context and Background	282
Project Description and Inception	285
Impact	293
Challenges	298
Looking Forward	302
15. United States Opening GPS for Civilian Use.....	307
Creating a Global Public Utility	
Context and Background	308
Project Description and Inception	310
Impact	318
Challenges	326

Looking Forward	330
-----------------	-----

Part V. Case Studies: Solving Public Problems

16. Battling Ebola in Sierra Leone.....	337
Data Sharing to Improve Crisis Response	
Context and Background	338
Project Description and Inception	340
Impact	349
Challenges	353
Looking Forward	356
17. New Zealand's Christchurch Earthquake Clusters.....	361
Open Data for Improving Emergency Response	
Context and Background	362
Project Description and Inception	364
Impact	374
Challenges	377
Looking Forward	379
18. Singapore's Dengue Cluster Map.....	381
Open Data for Public Health	
Context and Background	382
Project Description and Inception	387
Impact	390
Challenges	392
Looking Forward	394
19. United States' Eightmaps.....	397
The Unintended Negative Consequences	
Context and Background	398
Project Description and Inception	403
Impact	406
Challenges	410
Looking Forward	413

20. Kennedy v. City of Zanesville, United States.....	417
Open Data as Evidence	
Context and Background	418
Case Description and Inception	421
Impact	425
Challenges	428
Looking Forward	431
A. Appendix.....	433

Acknowledgements

The authors would like to thank Omdiyar Network for supporting this project. In particular, Laura Bacon, principal of policy investments, has been instrumental in developing the case studies. This work would not have been possible without the tireless efforts of our colleagues at the GovLab, including the case studies research team: Ali Clare, Auralice Graft, Juliet McMurren, Christina Rogawski, David Sangokoya, and Rebecca Young; Akash Kapur who provided crucial editorial support; and the GovLab developer team: Chris Wong, Claudio Mendonca, Batu Sayici, and Mark Adkins-Hastings who created odimpact.org where this research lives online. Thanks also to the GovLab's Beth Simone Noveck, Fred DeJohn, Lauren Yu, Dinorah Cantú, Julia Root, Irene Tello-Arista, and María Hermosilla for their input and support throughout the development of this research.

The research contained in this volume was developed thanks to the following experts and stakeholders who generously gave their time and provided essential on-the-ground knowledge about their inspiring and transformative work: Neil Ackroyd, Kim Alexander, Robert Andrews, Jay Bhalla, Uuf Brajawidagba, François Brouard, Iain Campion, John Carpenter, Daniel Carranza, Julian Carver, Pablo Clark, Izabela Corrêa, Jeff de la Beaujardiere, Leodelma de Marilac Felix, Vivien Deparday, Patrick DuFour, Ee-Peng Lim, Carl Elstam, Peter Elson, Felipe Estefan, Aidan Eyakuze, Stephen Ferris, Christian Fischer, Mike Flowers, Rafael García Aceves, Zachary Goldstein, Hanna Helquist, Thorhildur Jetzek, Al Kags, Jason Kim, Verena Luise Knippel, Daniel Kreiss, David Lasby, Michael Lenczner, Morten Lind, Andrew Loveless, Lindsay Marchessault, Arnold Minde, Lindsay Mollineaux, Oscar Montiel, Otávio Moreira de Cas-

tro Neves, Ainun Najib, Bitange Ndemo, Bo Overgaard, Tim Owen, Allan Parnell, Ed Parsons, Maria Patterson, Hilary Pearson, Karl Peterson, Florent Peyre, Peter Rabley, Tara Ramchandani, Ira Rubinstein, Fabrizio Scrollini, Diah Setiawaty, Rupert Simons, Singapore National Environment Agency, Gabriel Sipos, Diego Soria, Allison Soussi-Tanani, Javier Teran, Samhir Vasdev, Eva Vozárová, Neil Weston, Alyssa Wright, and Alexandra Zapata Hojel.

Finally, our gratitude to the following dedicated individuals who provided invaluable input during the open peer review process of this research: Karin Ahlin, Antonio Almansa Morales, Andi Argast, Jos Berens, Keitha Booth, J. Albert Bowden II, Mark Cardwell, Corinne Cath, Emmy Chirchir, Rafael García Aceves, Erik Holmlund, Brendan Kenny, Alessia Lefebure, Ulrich Mans, Valerie Moye, Alina Östling, Giuseppe Reale, Fathima Rifaa, Julian Singh, Rayna Stamboliyska, Julian Tait, Jamie Van Ymeren, Mario Velasco, Niki Virani, Johanna Walker, Andrew Weller, Ian White, Raymond Yee, and Maria Zuffova.

Executive Summary

Recent years have witnessed considerable enthusiasm over *open data*. Several studies have documented its potential to spur economic innovation and social transformation as well as to usher in fresh forms of political and government accountability. Yet for all the enthusiasm, we know little about how open data actually works and what forms of impact it is really having.

This report seeks to remedy that informational shortcoming. Supported by Omidyar Network, the GovLab has conducted 19 detailed case studies of open-data projects around the world. The case studies were selected for their sectoral and geographic representativeness. They were built in part from secondary sources (“desk research”), but also from a number of first-hand interviews with important players and key stakeholders. In this report, we consider some overarching lessons that we can learn from the case studies and assemble them within an analytical framework that can help us better understand what works—and what doesn’t—when it comes to open data.

The report begins ([“I. What Is Open Data?” on page 5](#)) with an overview of open data. Like many technical terms, open data is a contested and dynamic concept. The GovLab has conducted a study of nine widely used definitions to arrive at the following working definition, which guides our discussion here:

Open data is publicly available data that can be universally and readily accessed, used, and redistributed free of charge. It is structured for usability and computability.

[“II. The Case Studies” on page 6](#) includes a brief summary of our 19 case studies, each of which is detailed at considerably greater length,

in Parts II through V. Sections III through V represent the core of our analytical framework; they identify the key parameters and variables that determine the impact of open data.

“III. What Is the Impact of Open Data on People’s Lives?” on page 14 discusses what we have identified as the four most important dimensions of impact. Based on the case studies, GovLab has determined that open data projects are *improving government*, primarily by making government more accountable and efficient; *empowering citizens*, by facilitating more informed decision-making and enabling new forms of social mobilization; *creating new economic opportunities*; and *helping policymakers and others find solutions to big, previously intractable public problems* (e.g., related to public health or global warming).

These types of effects cannot be taken for granted. They are evident to varying degrees across our case studies, and sometimes not at all. Our research also identified four enabling conditions that allow the potential of open data to manifest (“IV. What Are the Enabling Conditions that Significantly Enhance the Impact of Open Data?” on page 17). Overall, we found that open data projects work best when they are *based on partnerships and collaborations* among various (often intersectoral) organizations; when they emerge within what we call an “*open data public infrastructure*” that enables the regular release of potentially impactful data; when they are accompanied by *clear open data policies*, including performance metrics; and when they address or attempt to solve a *well-defined problem or issue* that is an obvious priority to citizens and likely beneficiaries.

“V. What Are the Challenges to Open Data Making an Impact?” on page 20 identifies the key challenges that open data projects face. These include *a lack of readiness*, especially evident in the form of low technical and human capacity in societies or nations hosting open data initiatives; projects that are *unresponsive*—and thus inflexible—to user or citizen needs; projects that result in *inadequate protections for privacy or security*; and, finally, projects that suffer from *a shortage of resources*, financial and otherwise. None of the 19 initiatives we studied was immune to these obstacles; the most successful ones had found ways to surmount them and build applications or platforms that were nonetheless able to tap into the potential of open data.

“VI. Recommendations: Toward a Next Generation Open-Data Roadmap” on page 25 features a set of 10 recommendations directed at policymakers, entrepreneurs, activists, and others contemplating open-data projects. Each of these broad recommendations is accompanied by more specific and concrete steps for implementation. Together, these recommendations and steps for implementation add up to something of a toolkit for those working with open data. Although preliminary, they are designed to guide the open-data community in its ongoing efforts to launch new initiatives that achieve maximum societal, economic, political, and cultural change.

The report ends with each of our 19 in-depth case studies, presented in full and organized by their dimension of impact.

PART I

Open Data's Impact—Lessons Learned

CHAPTER 1

Understanding the Impact of Open Data

Introduction

Recent years have witnessed considerable enthusiasm over the opportunities offered by open data. Across sectors, it is widely believed today that we are entering a new era of information openness and transparency, and that this has the potential to spur economic innovation, social transformation, and fresh forms of political and government accountability. Focusing just on economic impacts, in 2013, for example, the consulting firm McKinsey estimated the possible global value of open data to be more than \$3 trillion per year.¹ A study commissioned by Omidyar Network has likewise calculated that open data could result in an extra \$13 trillion over five years in the output of G20 nations.²

Yet despite the evident potential of open data, and despite the growing amounts of information being released by governments and corporations, little is actually known about its use and impact. What kind of social and economic transformations has open data brought about, and what transformations might it effect in the future? How

¹ Manyika, James, Michael Chui, Diana Farrell, Steve Van Kuiken, Peter Groves, and Elizabeth Almasi Doshi. “[Open Data: Unlocking Innovation and Performance with Liquid Innovation](#).” McKinsey Global Institute. November 12, 2013.

² Gruen, Nicholas, John Houghton, and Richard Tooth. “[Open for Business: How Open Data Can Help Achieve the G20 Growth Target](#).” Omidyar Network. June 2014.

—and under what circumstances—has it been most effective? How have open-data practitioners mitigated risks (e.g., to privacy) while maximizing social good?

As long as such questions remain unanswered, the field risks suffering from something of a mismatch between the supply (or availability) of data and its actual demand (and subsequent use). This mismatch limits the impact of open data and inhibits its ability to produce social, economic, political, cultural, and environmental change. This report begins from the premise that in order to fully grasp the opportunities offered by open data, a more full and nuanced understanding of its workings is necessary.

Our knowledge of how and when open data actually works in practice is lacking because there have been so few systematic studies of its actual effect and workings. The field is dominated by conjectural estimates of open data's hypothetical influence; those attempts that have been made to study concrete, real-world examples are often anecdotal or suffer from a paucity of information. In this report, we seek to build a more systematic study of open data and its effect by rigorously examining 19 case studies from around the world. These case studies are chosen for their geographic and sectoral representativeness. They are built not simply from secondary sources (e.g., by rehashing news reports) but from extensive interviews with key actors and protagonists who possess valuable and thus far untapped on-the-ground knowledge. They go beyond the descriptive (what happened) to the explanatory (why it happened, and what is the wider relevance or impact).

To provide these explanations, we have assembled an analytical framework that applies across the 19 case studies and lets us present some more widely applicable principles for the use and impact of open data. Impact—a better understanding of how and when open data really works—is at the center of our research. Our framework seeks to establish a taxonomy of impact for open-data initiatives, outlining various dimensions (from improving government to creating economic opportunities) in which open data has been effective. In addition, the framework lays out some key conditions that enable impact, as well as some challenges faced by open-data projects.

I. What Is Open Data?

It is useful to begin with an understanding of what we mean by open data. Like many technical terms, open data is a contested concept. There exists no single, universally accepted definition. The GovLab recently undertook an analysis of competing meanings, with a view to reaching a working definition. The Appendix contains nine widely used definitions and our matrix of analysis.

Based on this matrix, we reached the following working definition, which guides our research and discussion throughout this report:

Open data is publicly available data that can be universally and readily accessed, used, and redistributed free of charge. It is structured for usability and computability.

It is important to recognize that this is a somewhat idealized version of open data. In truth, few forms of data possess all the attributes included in this definition. The openness of data exists on a continuum, and although many forms of information we discuss here might not be strictly open in the sense just described, they can nonetheless be shareable, usable by third parties, and capable of effecting wide-scale transformation. The 19 case studies included here therefore include a variety of different kinds of data, each of which is open in a different way, and to a different degree. Here are some examples:

- Brazil's Open Budget Transparency Portal is an example of the most "traditional" type of open-data project: a downloadable set of open government data accessible to the public.
- Mexico's Mejora Tu Escuela is the result of a nongovernmental organization compiling and presenting data (including open government data) in easily digestible forms.
- The Global Positioning System (GPS) is arguably not an "open data" system at all, but rather a means for providing access to a government-operated signal.
- The United Kingdom Ordnance Survey offers a combination of free and paid spatial data, suggesting the possibilities (and limitations) of a mixed model of open and closed data.

In each of these cases, "open" has different meanings and connotations. Many—but not all—of the cases, however, demonstrate the importance of shared and disseminated information, and highlight

open data's potential to enhance the social, economic, cultural, and political dimensions of our lives.

II. The Case Studies

Methodology

To select our case studies, we undertook a multistep process that involved several variables and considerations. To begin with, we examined existing repositories of open-data cases and examples in order to develop an initial universe of known open-data projects (see <http://odimpact.org/resources.html>). This initial scan of existing examples allowed us to identify gaps in representation—those sectors or geographies that often remain underrepresented in existing descriptions of open data and its effect (or lack thereof). To fill in some of these gaps (and more generally widen our list of case study candidates), we also reached out to a number of experts in relevant subject areas; for example, open data, open governance, civic technology, and other related fields. We also attended and conducted outreach at a number of open-data-related events, notably the 2015 International Open Data Conference in Ottawa, Canada and Con-Datos in Santiago, Chile.

Based on this process, we identified a long list of approximately 50 case studies from around the world. These included examples from the private sector, civil society, and government, and spanned the spectrum of openness just mentioned. The next step was to conduct a certain amount of preliminary research to arrive at our final list of 19 case studies. To do this, we took into account several factors: the availability and type of evidence in existence; the need for sectoral and geographic representativeness; and the type of impact demonstrated by the case study in question (if any). We also considered whether previous, detailed case studies existed; as much as possible, our goal was to develop case studies for previously unexplored and undocumented examples.

Having selected our 19 cases, we then began a process of more in-depth researching. This involved a combination of desk research (e.g., using existing media and other reports) and interviews (usually by telephone). For many of our examples, there existed very little existing research; the bulk—and certainly the most useful—of our evidence came from a series of in-depth interviews we conduced

ted with key participants and observers who had been involved in our various cases.

Upon completing drafts of each case study, and in the spirit of openness that defines the field under examination, we open-sourced the peer review process for each case and this paper. Rather than sharing drafts only with a select group of experts, we made our report and each of the case studies openly accessible for review in the interest of gaining broad input on our findings and collaboratively producing a common resource on open data's effects for the field. Through broad outreach at events like the 2015 Open Government Partnership Summit in Mexico City, Mexico, and through social media, more than 50 individuals from around the world signed up to peer review at least one piece.

During the month-long open-peer-review process, more than two dozen of those who signed up shared their input as Recognized Peer Reviewers through in-line comments and in-depth responses to the ideas and evidence presented in this report. Additionally, each element of the report was made openly accessible to the public, allowing anyone to share suggestions, clarifications, notes on potential inaccuracies and any other useful input prior to publishing. Much of this input was integrated into the final version of this report.

The 19 Cases

The standalone impact case studies (see Parts II through V) include detailed descriptions and analyses of the initiatives listed later in the report. In addition, the following table summarizes their main features and key findings. Here, we include a brief summary of each example:

Outcome: Improving Government

Brazil: Open Budget Transparency Portal

Sector: Public

Impact: Tackling corruption and transparency

Description: A tool that aims to increase fiscal transparency of the Brazilian Federal Government through open government budget data. As the quality and quantity of data on the portal have improved over the past decade, the Transparency Portal is now one of the country's primary anti-corruption tools, registering an aver-

age of 900,000 unique visitors each month. Local governments throughout Brazil and three other Latin American countries have modeled similar financial transparency initiatives after Brazil's Transparency Portal.

Sweden: openaid.se

Sector: Philanthropy and aid

Impact: Tackling corruption and transparency

Description: A data hub created by the Swedish Ministry of Foreign Affairs and the Swedish International Development Cooperation Agency (Sida) built on open government data. The website visualizes when, to whom, and why aid funding was paid out and what the results were. The reforms are seen to be an important force for enhanced transparency and accountability in development cooperation at an international level and increased cooperation and involvement of more actors in Swedish development policy.

Slovakia: Open contracting projects

Sector: Public sector

Impact: Tackling corruption and transparency

Description: In January 2011, Slovakia introduced a regime of unprecedented openness, requiring that all documents related to public procurement (including receipts and contracts) be published online, and making the validity of public contracts contingent on their publication. More than two million contracts have now been posted online, and these reforms appear to have had a dramatic effect on both corruption and, equally important for the business climate, perceptions of corruption.

Indonesia: Kawal Pemilu

Sector: Politics and elections

Impact: Tackling corruption and transparency

Description: A platform launched in the immediate aftermath of the contentious 2014 Indonesian presidential elections. Kawal Pemilu's organizers assembled a team of more than 700 volunteers to compare official vote tallies with the original tabulations from polling stations and to digitize the often handwritten forms, making the data more legible and accessible. Assembled in a mere two days,

with a total budget of just \$54, the platform enabled citizen participation in monitoring the election results, increased public trust in official tallies, and helped ease an important democratic transition.

Denmark: consolidation and sharing of address data

Sector: Geospatial services

Impact: Improving services

Description: In 2005, the Building and Dwelling Register of Denmark started to release its address data to the public free of charge. Prior to that date, each municipality charged a separate fee for access, rendering the data practically inaccessible. There were also significant discrepancies between the content held across different databases. A follow-up study commissioned by the Danish government estimated the direct financial benefits alone for the period 2005–2009 at €62 million, at a cost of only €2 million.

Canada: T3010 charity information return data

Sector: Philanthropy and aid

Impact: Improving services

Description: In 2013, the Charities Directorate of the Canada Revenue Agency (CRA) opened all T3010 Registered Charity Information Return data since 2000 via the government's data portal under a commercial open-data license. The resulting data set has been used to explore the state of the nonprofit sector, improve advocacy by creating a common understanding between regulators and charities, and create intelligence products for donors, fundraisers and grant-makers.

Outcome: Empowering Citizens

Tanzania: Shule and Education Open Data Dashboard

Sector: Education

Impact: Social mobilization

Description: Two recently established portals providing the public with more data on examination pass rates and other information related to school performance in Tanzania. Education Open Data Dashboard is a project established by the Tanzania Open Data Initiative; Shule was spearheaded by Arnold Minde, a programmer, entrepreneur, and open-data enthusiast. Despite the challenges

posed by Tanzania's low Internet penetration rates, these sites are slowly changing the way citizens access information and make decisions. They are encouraging citizens to demand greater accountability from their school system and public officials.

Kenya: Open Duka

Sector: Public sector

Impact: Informed decision-making

Description: A platform developed by the civil society organization, the Open Institute, that aims to address issues of opacity in governance in the private and public sectors, promoting corporate accountability and transparency by providing citizens, journalists, and civic activists with insight into the relationships, connections (and, to some extent, the dynamics) of those in and around the public arena. As a case study, it exemplifies the challenge for open-data initiatives to generate sufficient awareness and use necessary methods to achieve impact.

Mexico: Mejora Tu Escuela

Sector: Education

Impact: Informed decision-making

Description: A platform created by the Mexican Institute for Competitiveness (IMCO) that provides citizens with information about school performance. It helps parents choose the best option for their children, empowers them to demand higher-quality education, and gives them tools to get involved in their children's schooling. It also provides school administrators, policymakers, and NGOs with data to identify hotbeds of corruption and areas requiring improvement. Data available on the site was used in a report that uncovered widespread corruption in the Mexican education system and stirred national outrage.

Uruguay: A Tu Servicio

Sector: Health

Impact: Informed decision-making

Description: A platform that lets users select their location and then compare local health care providers based on a wide range of parameters and indicators, such as facility type, medical specialty,

care goals, wait times and patient rights. A Tu Servicio has introduced a new paradigm of patient choice into Uruguay's health care sector, enabling citizens not only to navigate through a range of options but also generating a healthy and informed debate on how more generally to improve the country's health care sector.

Outcome: Creating Opportunity

Great Britain's Ordnance Survey

Sector: Geospatial services

Impact: Economic growth

Description: Data from Ordnance Survey (OS), Britain's mapping agency, supports essentially any UK industry or activity that uses a map: urban planning, real estate development, environmental science, utilities, retail, and much more. OS is required to be self-financing and, despite the launch of its OS OpenData platform in 2010, uses a mixed-cost model, with some data open and some data paid. OS OpenData products are estimated to deliver between a net £13 million to £28.5 million increase in GDP over its first 5 years.

United States: New York City Business Atlas

Sector: Business

Impact: Economic growth

Description: Developed by the Mayor's Office of Data Analytics (MODA), the Business Atlas is a platform designed to alleviate the market research information gap between small and large businesses in New York City. The tool provides small businesses with access to high-quality data on the economic conditions in a given neighborhood to help them decide where to establish a new business or expand an existing one.

US: NOAA: Opening up global weather data in collaboration with businesses

Sector: Weather

Impact: Economic growth

Description: Opening up weather data through NOAA has significantly lowered the economic and human costs of weather-related damage through forecasts; enabled the development of a multibillion dollar weather derivative financial industry dependent on sea-

sonal data records; and catalyzed a growing million-dollar industry of tools and applications derived from NOAA's real-time data.

US: Opening GPS data for civilian use

Sector: Geospatial services

Impact: Economic growth

Description: Over the past 20 years, Global Positioning System (GPS) technology has led to a proliferation of commercial applications across industries and sectors, including agriculture, construction, transportation, aerospace and—especially with the proliferation of portable devices—everyday life. Were the system to be somehow discontinued, losses are estimated to be \$96 billion. In addition to creating new efficiencies and reducing operating costs, the adoption of GPS technology has improved safety, emergency response times and environmental quality, and has delivered many other less-readily quantifiable benefits.

Outcome: Solving Public Problems

Sierra Leone: Battling Ebola

Sector: Health

Impact: Data-driven engagement

Description: In 2014, the largest Ebola outbreak in history occurred in West Africa. At the beginning, information on Ebola cases and response efforts was dispersed across a diversity of data collectors, and there was little ability to get relevant data into the hands of those who could make use of it. Three projects—Sierra Leone's National Ebola Response Centre (NERC), the United Nations' Humanitarian Data Exchange (HDX), and the Ebola GeoNode—significantly improved the quality and accessibility of information used by humanitarians and policymakers working to address the crisis.

New Zealand: Christchurch earthquake GIS clusters

Sector: Emergency services

Impact: Data-driven engagement

Description: In February 2011, Christchurch was struck by a severe earthquake that killed 185 people and caused significant disruption and damage to large portions of a city already weakened by an earlier earthquake. In the response to the quake, volunteers and offi-

cials at the recovery agencies used open data, open source tools, trusted data sharing, and crowdsourcing to develop a range of products and services required to respond successfully to emerging conditions, including a crowdsourced emergency information web app that generated 70,000 visits within the first 48 hours after the earthquake, among others.

Singapore: Dengue cluster map

Sector: Health

Impact: Data-driven engagement

Description: In 2005, the Singapore National Environment Agency (NEA) began sharing information on the location of dengue clusters as well as disease information and preventive measures online through a website now commonly known as the “Dengue Website.” Since then, the NEA’s data-driven cluster map has evolved, and it became an integral part of the campaign against a dengue epidemic in 2013.

US: Eightmaps

Sector: Politics and elections

Impact: Data-driven engagement

Description: A tool, launched anonymously in 2009, that provided detailed information on supporters of California’s Proposition 8, which sought to bar same-sex couples from marrying. The site collected information made public through state campaign finance disclosure laws and overlaid that information onto a Google map of the state. Users could find the names, approximate locations, amount donated, and, where available, employers of individuals who donated money to support Prop 8. Eightmaps demonstrates how the increased computability and reusability of open data could be acted upon in unexpected ways that not only create major privacy concerns for citizens, but could also lead to harassment and threats based on political disagreements.

US: Kennedy v the City of Zanesville

Sector: Law

Impact: Data-driven assessment

Description: For more than 50 years, even though access to clean water from the City of Zanesville, Ohio, was available throughout the rest of Muskingum County, residents of a predominantly African American area of Zanesville were only able to use contaminated rainwater or drive to the nearest water tower. One of the key pieces of evidence used during the court case was a map derived from open data that showed significant correlation between the houses occupied by the white residents of Zanesville and the houses hooked up to the city water line. The case went in favor of the African-American plaintiffs, awarding them a \$10.9 million settlement.

III. What Is the Impact of Open Data on People's Lives?

What lessons can we learn from these examples of open-data applications, platforms, and websites? In this and the following sections, we outline some overarching insights derived from our 19 case studies. First, we focus on impact. What is the effect of open data on people's lives? What are the real, measurable, and tangible results of our case studies? And, just as important, who (which individuals, institutions, demographic groups) are most affected?

Taxonomy of Open-Data Impact

Determining impact requires taking certain nuances into account. In many cases, open-data projects show results in more than one dimension of impact. In addition, the effect of our case studies on people's lives is often indirect (and thus somewhat more subtle), mediated by changes in the way decisions are made or other broad social, political, and economic factors. Nonetheless, despite these nuances, our analysis suggests that there exist four main ways in which open data is having an influence on people's lives (Figure 1-1):

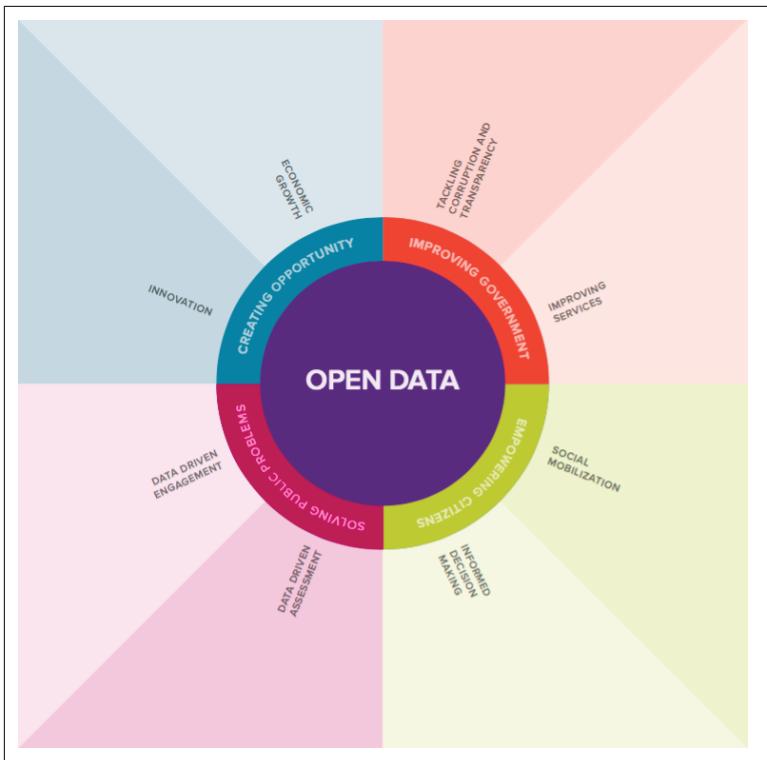


Figure 1-1. The open data taxonomy of impact

- First, open data is *improving government* around the world. It is doing so in various ways, but in particular by, a) making governments more accountable, especially by helping tackle corruption and adding transparency to a host of government responsibilities and functions (notably budgeting), and, b) making government more efficient, especially by enhancing public services and resource allocation.

Improvements in governance are evident in six of our 19 case studies. Notable examples include the Brazil Open Budget Transparency Portal, which brings accountability and citizen oversight to the country's budget processes; Slovakia's Central Registry, which is a global model for the open-contracting movement; and Canada's opening of tax return data submitted by charities, the first move in a broader global effort to increase the transparency and accountability of philanthropies.

- Open data is *empowering citizens* to take control of their lives and demand change by enabling more informed decision-making and new forms of social mobilization, both in turn facilitated by new ways of communicating and accessing information.

This dimension of impact plays a role in four case studies. Some notable examples in this category include Uruguay's A Tu Servicio, which empowers citizens to make more informed decisions about health care, and education dashboards in Mexico (Mejora Tu Escuela) and Tanzania (Shule and Education Open Data Dashboard), each of which enables parents to make more evidence-based decisions about their children's schools.

- Open data is creating *new economic opportunities* for citizens and organizations. Around the world, in cities and countries, greater transparency and more information are stimulating economic growth, opening up new sectors, and fostering innovation. In the process, open data is creating new jobs and new ways for citizens to prosper in the world.

This category of impact often follows from applications and platforms built using government data. It is evident in four of our case studies, each of which relies for its underlying data on information released by governments. Two notable examples include New York's Business Atlas, which lets small businesses use data to identify the best neighborhoods in which to open or grow their companies; and the various platforms and companies built around data released by the National Oceanic and Atmospheric Administration (NOAA) in the US.

- Finally, open data's effect is evident in the way it is helping *solve several big public problems*, many of which have until recently seemed intractable. Although most of these problems have not been entirely solved or eliminated, we are finally seeing pathways to improvements. Through open data, citizens and policymakers can analyze societal problems in new ways and engage in new forms of data-driven assessment and engagement.

Open data has created notable impacts during public-health crises and other emergencies. In Sierra Leone, open data helped to inform the actions of people working on the ground to fight Ebola. The government and citizens of Singapore are using a Dengue Fever Cluster Map to try to limit the spread of dengue fever during outbreaks like that experienced in 2013. The efforts

to rebuild following devastating earthquakes in Christchurch, New Zealand were also aided by open data. It is important to recognize, however, that attempts to solve problems can also have unintended consequences. We see this, for example, in the case of Eightmaps, where efforts to address discrimination and other issues unintentionally created new privacy (and even personal security) problems.

IV. What Are the Enabling Conditions that Significantly Enhance the Impact of Open Data?

Although our initial analysis told us what *types* of change open data was creating, a further round of analysis was required to understand *how* change comes about. In examining open data projects around the world, we are struck by the wide variability in outcomes. Some work better than others, and some simply fail. Eightmaps is an example of how open data can lead to unintended consequences, but there are many, many more examples that the GovLab did not select for this group of case studies due to the lack of meaningful, measurable effect to date. Some projects do well in a particular dimension of success while failing in others. If we are to achieve the believed potential of open data and scale the impact of the individual case studies included here, we need a better, more granular understanding of the enabling conditions that lead to success.

Based on our research, we identified four key enabling conditions, each of which allows us to articulate a specific “premise” for success:

Partnerships

The *power of collaboration* was evident in many of the most successful open-data projects we studied. Effective projects were built not from the efforts of a single organization or government agency, but rather from partnerships across sectors and sometimes borders. Two forms of collaboration were particularly important: partnerships with civil society groups, which often played an important role in mobilizing and educating citizens; and partnerships with the media, which informed citizens and also played an invaluable role in analyzing and finding meaning in raw open data. In addition, we saw an important

role played by so-called “data collaboratives,” which pooled data from different organizations and sectors.

Virtually all the case studies we examined were the products of some form of partnership. Uruguay’s A Tu Servicio was an important example of how civil society can work with government to craft more effective open-data initiatives. NOAA’s many offshoots and data initiatives are an equally important example of collaboration between the private and public sectors. New York City’s Business Atlas was similarly an illustration of a public-private partnership; its data set, built both from government and private-sector information (supplied by the company Placemeter), is an example of an effective data collaborative.

Premise 1: Intermediaries and data collaboratives allow for enhanced matching of supply and demand of data.

Public infrastructure

Several of the most effective projects we studied emerged on the back of what we might think of as an *open data public infrastructure*; that is, the technical backend and organizational processes necessary to enable the regular release of potentially impactful data to the public. In some cases, this infrastructure takes the form of an “open by default” system of government data generation and release. The team behind Kenya’s Open Duka, for example, is responding to its lack of impact to date by attempting to build such an infrastructure with county-level governments to improve the counties’ internal data capacity, improving the data available on Open Duka as a result.

An open-data public infrastructure does not, however, only involve technical competencies. As part of the push around Brazil’s Open Budget Transparency Portal, for example, organizers not only developed an interoperable infrastructure for publishing a wide variety of data formats, but also launched a culture-building campaign complete with workshops seeking to train public officials, citizens and reporters to create value from the open data.

Premise 2: Developing open data as a public infrastructure enables a broader impact across issues and sectors.

Policies and performance metrics

Another key determinant in the success of open-data projects is the existence of *clear open data policies, including well-defined performance metrics*. The need for clear policies (and more generally an enabling regulatory framework) is a reminder that technology does not exist in a vacuum. Policymakers and political leaders have an essential role to play in creating a flexible, forward-looking legal environment that, among other things, encourages the release of open data and technical innovation; and that spurs the creation of fora and mechanisms for project assessment and accountability.

In addition, high-level political buy-in is also critical. It is not sufficient simply to pass enabling laws that look good on paper. Policymakers and politicians must also ensure that the letter of the law is followed, that vested interests are adequately combated, and that there are consequences for working against openness and transparency.

Among the many case studies that benefited from a conducive policy environment, a few stand out. In Mexico, we can see how an open-data initiative (in this case, the Mejora Tu Escuela project) can benefit from high-level government commitments to opening data that trickles down to—and empowers—local and regional governments. Slovakia's Central Registry is another good example; it shows how laws can be redesigned, in this case to encourage transparency by default in contracting, and in the process greatly increase openness. The openness of GPS, though ingrained in daily life for many, was the subject of questions following the terrorist attacks of September 11, 2001; those questions were put to rest with the enactment of a new policy commitment in 2004 to maintain unfettered global access to the geospatial system.

Premise 3: Clear policies regarding open data, including those promoting regular assessments of open-data projects, provide the necessary conditions for success.

Problem definition

We have repeatedly seen how the most successful open-data projects are those that address *a well-defined problem or issue*. It is very challenging for open-data projects to try to change user behavior or convince citizens of a previously unfelt need. Effec-

tive projects identify an existing—ideally widely recognized—need, and provide new solutions or efficiencies to address that need.

Singapore's Dengue Fever Cluster Map is a good example in this regard. Its core area of activity (public health) has clear, tangible benefits; it seeks to limit the spread of an illness that policymakers widely recognize as a problem, and that citizens dread. Uruguay's A Tu Servicio is another good example: it provides clear, tangible benefits to citizens, giving them the means to take action that improves their health care. It is perhaps no coincidence that both of these examples are in the health sector: The most successful projects often touch on the most basic human needs (health, pocketbook needs, etc.). In a case involving one of the most essential human needs, the use of open data in Kennedy versus the City of Zanesville accomplished its singular goal: demonstrating beyond a reasonable doubt that water access decisions were being made on the basis of citizens' race.

Premise 4: Open data initiatives that have a clear target or problem definition have more effect.

V. What Are the Challenges to Open Data Making an Impact?

The success of a project is also determined by the obstacles and challenges it confronts. The challenges are themselves the function of numerous social, economic, and political variables. In addition, some regions might face more obstacles than others.

As with the enabling conditions, we found widespread geographic and sectoral variability in our analysis of challenges. Broadly, we identified four challenges that recurred the most frequently across our 19 case studies:

Readiness

Perhaps unsurprisingly, countries or regions with *overall low technical and human capacity or readiness* often posed inhospitable environments for open-data projects. The lack of technical capacity could be indicated by several variables, including low Internet penetration rates, a wide digital divide, or overall poor technical literacy. In addition, technical readiness can also be indicated by the existence of a group of individuals or entities

that are technically sophisticated and that believe in the transformative potential of technology, particularly of open data. Repeatedly, we have seen that such “data champions” or “technological evangelists” play a critical role in ensuring the success of projects.

Low technical capacity did not necessarily result in outright project “failures.” Rather, it often stunted the potential of projects, making them less impactful and successful than they could otherwise have been. In Tanzania, for instance, the Shule and Education Open Data Dashboard portals were limited by low Internet penetration rates and by a general low awareness about open data. Slovakia’s Central Registry was in many ways very successful; yet it, too, was restricted by a lack of technical capacity among government officials and others (particularly at the lower level). In these projects and others, we see that success is relative, and that even the most successful projects could be enhanced by greater attention to the overall technical environment or ecosystem.

Premise 5: The lack of readiness or capacity at both the supply and demand side of open data hampers its impact.

Responsiveness

Success is also limited when projects are *unresponsive to feedback and user needs*. As we saw in the previous section, the most successful projects address a clear and well-defined need. A corollary to this finding is that project sponsors and administrators need to be attuned to user needs; they need to be flexible enough to recognize and adapt to what their users want.

For Sweden’s OpenAid project, for example, user experience was not a core priority at launch, and much of the information found on the site is too complex for most citizens to digest. Despite this high barrier to entry, the site only offers limited engagement opportunities—namely, a button for reporting bugs on the site. Moreover, project titles found on the site often contain cryptic terms interpretable only to those with close familiarity with the project at hand.

NOAA, on the other hand, has some of the most mature and wide-reaching open-data efforts in any of the cases studied here. But given that breadth, for the agency’s essential information to remain useful to the evolving needs of its users, an increased

focus needs to be placed on customer analytics and user behaviors. The UK's Ordnance Survey has very sophisticated user analytics and prioritizes customer satisfaction; however, the separation of OS OpenData from its other data sets and products is potentially limiting.

Premise 6: Open data could be significantly more impactful if its release would be complemented with a responsiveness to act upon insights generated.

Risks

A major challenge arises from the trade-offs between the potential of open data and the *risks posed by privacy and security violations*. These risks are inherent to any open-data project—by its very nature, greater transparency exists in tension with privacy and security. When an initiative fails to take steps to mitigate this tension, it risks not only harming its own prospects, but more broadly the reputation of open data in general.

Concerns about privacy and security dogged many of the projects we studied. In Brazil, more than 100 legal actions were brought against the Open Budget Transparency Portal when it inadvertently published the salaries of public servants. In New York, despite steps being taken to mitigate such harms, there has been concern that citizen privacy might be violated as cameras collect data for the project in public spaces.

Without question, the clearest example of open data leading to privacy concerns (and even outright violations) can be found in the Eightmaps case study, which used public campaign finance disclosure laws to publish various identifying information about and home addresses for donors to California's Proposition 8, leading to instances of intimidation and harassment.

For all the very real—and legitimate—concerns, our case studies also show that the scope for privacy and security abuses can be mitigated. For example, NOAA stood out for its creation of a dedicated Cyber Security Division to address data security challenges when collecting and releasing data (the sole instance of such a dedicated division in our 19 case studies). Singapore, too, took proactive steps to anonymize data to protect the privacy of citizens. Addressing risks to privacy and security, though important, can also work against the goals of openness and transparency. For example, in the city of Zanesville, Ohio, secu-

rity concerns have been raised (controversially) to restrict access to data that has proven essential in addressing decades-old civil rights violations. Such examples are an important reminder of the tensions that exist between openness and security/privacy, and of the need for careful, judicious policymaking to achieve a balance.

Premise 7: Open data does pose a certain set of risks, notably to privacy and security; a greater, more nuanced understanding of these risks will be necessary to address and mitigate them.

Resource allocation

Finally, we found that *inadequate resource allocation* was one of the most common reasons for limited success or outright failure. Many of the projects we studied were “hackable”—that is, easily put together on a very limited budget, often created by idealistic volunteers. Indonesia’s Kawal Pemilu, for example, was assembled with a mere \$54. Over time, though, projects require resources to succeed; although they might emerge on the backs of committed (and cheap) idealists, they are fleshed out and developed with real financial backing.

The limited success of Kenya’s Open Duka is a good example. Although the project was well conceived and based on a sound premise, it has been limited by the unanticipated effort involved in data collection. More resources would almost certainly have helped address this challenge. In addition, Mexico’s Mejora Tu Escuela is just one project that relies on foundation funding to operate, leading to some level of uncertainty about the long-term sustainability of such projects should any of those funding streams be discontinued. The UK’s Ordnance Survey, meanwhile, is required to be self-financing, forcing the agency to rely heavily on private sector customers paying to access the more sophisticated data products not included in OS OpenData.

Even an initiative as central and widely used as GPS experiences funding challenges. In a government climate focused on budget cuts at every corner, new features and capabilities, even for a “global public utility,” can be difficult to finance through public money.

Premise 8: Even though open-data projects can often be launched cheaply, those projects that receive generous, sus-

tained, and committed funding have a better chance of success over the medium and long term.

The Eight Premises that Determine the Impact of Open Data

The following is a compilation of our eight premises:

Premise 1

Intermediaries and data collaboratives allow for enhanced matching of supply and demand of data.

Premise 2

Developing open data as a public infrastructure enables a broader impact across issues and sectors.

Premise 3

Clear policies regarding open data, including those promoting regular assessments of open-data projects, provide the necessary conditions for success.

Premise 4

Open data initiatives that have a clear target or problem definition have more effect.

Premise 5

The lack of readiness or capacity at both the supply and demand side of open data hampers its impact.

Premise 6

Open data could be significantly more impactful if its release would be complemented with a responsiveness to act upon insights generated.

Premise 7

Open data does pose a certain set of risks, notably to privacy and security; a greater, more nuanced understanding of these risks will be necessary to address and mitigate them.

Premise 8

Even though open-data projects can often be launched cheaply, those projects that receive generous, sustained and committed funding have a better chance of success over the medium and long term.

Supply Versus Demand Trajectories

In studying the ways in which open data has been made available, we've found consistent trajectories depending on whether the data is pushed from the government or made available by users in civil society or the general public extracting that data from reluctant institutions. Interestingly, we've found that as both open data push and pull trajectories advance, the optimal end point is the same: greater collaboration between data holders and data users.

Trajectory of open data push

Data release

Simply making some amount of data available

Open by default

Creating the infrastructure and processes needed for constant, automatic data release

Demand-driven collaboration

Working with users to make the most useful data available in the most useful ways

Trajectory of open data pull

Data audit and gap identification

Outside assessment of where data could have an effect if made accessible

Creation and demand

Through scraping, Freedom of Information requests, data leaks or other methods, data users finding ways to make government data accessible without the direct involvement (and often without the blessing) of the data holding institution

Collaboration

Working with government to craft impactful data release strategies

VI. Recommendations: Toward a Next Generation Open-Data Roadmap

Our case studies clearly indicate the tremendous potential and possibilities offered by open data. Around the world, open data has improved governments, empowered citizens, contributed solutions

to complex public problems, and created new economic opportunities for companies, individuals, and nations.

But despite this clear potential, the hurdles are also apparent. We outlined several of the particular issues faced by open-data projects in the preceding sections. In addition to these specific challenges, there is the more general problem of scaling: How do we move beyond a “points of light” narrative that celebrates individual case studies to a broader narrative about the social, economic, and political transformation that could result from a far broader deployment of open data? In this section, we outline 10 steps or recommendations for policymakers, advocates, users, funders, and other stakeholders in the open-data community that we believe could usher in such wholesale transformation ([Figure 1-2](#)). For each step, we describe a few concrete methods of implementation—ways to translate the broader recommendation into meaningful impact.

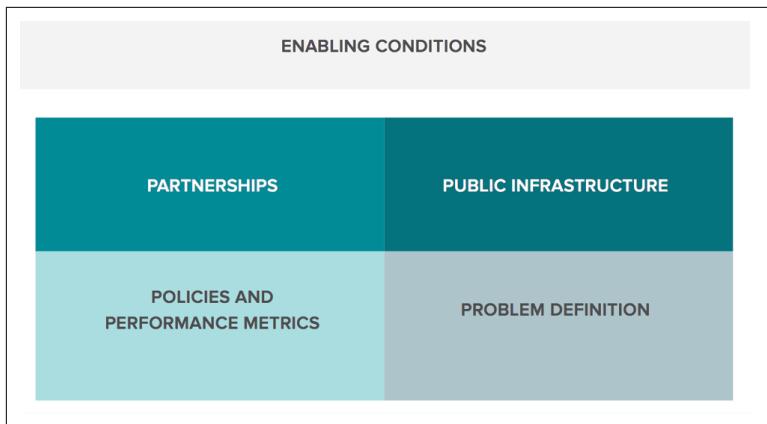


Figure 1-2. Recommendations for a next generation open data roadmap

Together, these 10 recommendations and their means of implementation amount to a *Next Generation Open Data Roadmap*. They let us better understand how the potential of open data can be fulfilled, across geographies, sectors, and demographics.

Recommendation 1: Focus on and Define Key Problem Areas for Which Open Data Can Add Value

A core premise offered by our case studies is that the impact of open data is often dependent on how well the problem it seeks to address is defined and understood. It is therefore essential for open-data advocates and practitioners to clearly define their goals, the problem they are seeking to address, and the steps they plan to take. Here are some possibilities for how this focus can be achieved:

- Set up a crowdsourced “Problem Inventory” to which users can contribute specific questions and answers, both of which can help define open-data projects. The UK Ordnance Survey’s GeoVation Hub is an interesting model focusing on the latter. It poses very specific questions (e.g., “How can we improve transport?” and “How can we feed Britain?”) for users to answer using OS OpenData.
- Facilitate user-led design exercises to help define important public and social problems and how open data can help solve them.
- To guide such exercises, it can be useful to establish “Problem and Data Definition toolkits”—potentially modeled on and informed by Freedom of Information requests—that help formulate clearly defined public issues and connect them with potentially useful open-data streams.

Recommendation 2: Encourage Collaborations Across Sectors (Especially Between Government, Private Sector, and Civil Society) to Better Match the Supply and Demand of Open Data

Large public problems are by definition cross-sectoral and interdisciplinary. They define boundaries and require a variety of expertise, knowledge, and data to be successfully addressed. It therefore stands to reason that the most successful open-data projects will similarly be collaborative and work across sectors and disciplines. Working in a collaborative manner can help draw on a diverse pool of talent and can also lead to innovative, out-of-the-box solutions. Perhaps most important, by allowing data users and data suppliers to work

together and interact, collaborative approaches can improve the match between data demand and supply, thus enhancing the overall efficiency of the demand-use-impact value chain for open data.

Here are some pathways to achieving the required collaborative and cross-sectoral approaches:

- Create data collaboratives to improve the efficiency and effectiveness of the demand-use-impact cycle. The value of data collaboratives is clearly illustrated by New Zealand's Canterbury Earthquake Recovery Authority's data sharing with construction companies, which is projected to deliver NZ\$40 million in savings. In addition, NOAA's Big Data Partnership, which formalized a sector partnership with five leading private-sector data and cloud technology companies, is also a good example.
- Engage and nurture data intermediaries, especially from civil society, to help spread awareness and disseminate data (and their findings) more widely. Data intermediaries play a particularly important role in countries with low technical capacity (e.g., as is evident in our Tanzanian case study); they offer a vital link between technology and society, helping citizens maximize and make real, effective use of data in their everyday lives.

Recommendation 3: Approach and Treat Data as a Form of Vital 21st-Century Public Infrastructure

Too often, policymakers and decision-makers focus solely on opening up data, as if open data on its own provides a silver bullet for a society's problems. In fact, as repeatedly evidenced in our case studies, data—in its raw form—needs to be supplemented by a host of other commitments: sustained and sustainable funding, skills training among those charged with data collection and use, and effective governance structures for every step of the data collection and use cycle. Approaching data in this broader, more holistic way means treating it as a vital form of public infrastructure. And this infrastructure is one that is at the heart of a society or nation, essential for its success, and embedded within wider social, economic, and political structures.

There are several steps policymakers can take to advance a “data-as-infrastructure” approach, including the following:

- Developing a systems design and mapping methodology. Mapping the public and private sector data infrastructure as well as local, national, and global data infrastructures that can affect the value creation of open data is a first and necessary step to approach data as infrastructure. A systems map could enable the more targeted, coordinated, and collaborative development of open-data technical standards and best practices across sectors.
- Embracing and implementing the Open Data Charter,³ which seeks to “foster greater coherence and collaboration” around open-data standards, practices, and, in particular, the following principles:
 - Open by default
 - Timely and comprehensive
 - Accessible and usable
 - Comparable and interoperable
 - Developed for improved governance and citizen engagement
 - Designed for inclusive development and innovation
- Taking advantage of existing public infrastructure, such as libraries, schools, and other cultural and education institutions, so that data is more firmly embedded into other forms of public investment and public life. Open Referral, for example, is creating a data backend for the social safety net, allowing pilot partners, including libraries, to tap into a wide, interconnected range of potentially impactful data on civic and social services.
- Developing skills and capacity around data collection, cleaning, and standardization to ensure better quality data is being released. This is especially important within agencies and organizations releasing data (to ensure its quality), but also, to the extent possible, within the community of users.
- Viewing and treating open data as a public good, something to which citizens and taxpayers are entitled. Moving toward a view

³ <http://opendatacharter.net/>

of open data as a public good requires as much of a cultural change as a policy change. As our case studies have repeatedly shown, the success of open data initiatives depends crucially on government stakeholders accepting that citizens, whether they be researchers, journalists, or just average individuals, have a right to demand access to government data.

Recommendation 4: Create Clear Open-Data Policies That Are Measurable and Allow for Agile Evolution

Our research illustrates the vital enabling role played by a national legal and regulatory framework that supports open data. Well-articulated internal rules and priorities are equally important when the releasing entity is a company or other organization. In both cases, clarity is essential: open data thrives when there is an unambiguous commitment to its cause. Importantly, open-data policies should include provisions to measure the success (or otherwise) of an initiative; systems for measurement and assessment are vital to ensuring accountability.

There are several steps policymakers can take to ensure the necessary clarity of open data policies. These include the following:

- Cocreating open-data policies with citizen and other groups, which can be an important way not only of drafting inclusive (and thus more legitimate) policies, but also of ensuring that policies are responsive to actual conditions and needs. Our research repeatedly shows that policies drafted without adequate public input and participation are less effective than those that draw on a wider range of experiences and expertise. Of course, attention must be paid to knowledge and power asymmetries involved in such cocreation processes.
- Engaging the public in defining and monitoring metrics of success: citizen participation in measuring the results of open-data initiatives is as important as in drafting policies, and for the same reasons. It is a vital part of ensuring accountability and in enhancing the legitimacy and effectiveness of open-data projects.
- Creating a “Metrics Bank” of important indicators, with input from stakeholders, researchers, and experts in the field. Such a

Metrics Bank could be built around the variety of categories of open data's effects, such as economic concerns (like return on investment or private sector economic revenues generated), public problem solutions (lives saved, increases in the efficiency of service delivery), and others. In line with the previous suggestion, the Metrics Bank should be reviewed on a regular basis by a citizens' group or panel created specifically for that purpose.

Recommendation 5: Take Steps to Increase the Capacity of Public and Private Actors to Make Meaningful Use of Open Data

Repeatedly, we have seen how open data initiatives are limited by a lack of capacity and preparedness among those who could potentially benefit most. Often, this manifests quite simply as a lack of awareness: those who do not know about the potential of open data are likely to use and benefit less from it. It is important to recognize that low capacity is a problem both on the demand side and supply side of the open-data value chain—policymakers and those tasked with releasing data are often as unprepared as intended beneficiaries.

Several steps can be taken to increase capacity and preparedness:

- Set up coaching and training centers to teach policymakers and key stakeholders among citizens about the potential benefits and applications of open data. Brazil's Open Budget Transparency Portal, for instance, benefited tremendously from TV campaigns and regular workshops designed to train citizens, reporters, and public officials on how to use the Open Budget Transparency Portal. In addition, a combined overview or searchable directory of coaching opportunities already in place and provided by, for instance, the GovLab Academy and the Open Data Institute, could enable easier navigation and matching of interests and needs worldwide.
- Establish mentor and expert networks for those seeking to use open data. Such networks can serve as valuable resources, providing guidance on the optimal uses of open data and helping citizens and policymakers overcome hurdles or navigate obstacles.

- Invest in and promote user-friendly data tools such as data visualizations and other analytic tools. Raw data can often be overwhelming for novice users; platforms and apps that include analytics and visualizations are often far more accessible. Notable examples from our case studies include the UK Ordnance Survey's OS OpenMap, NYC's Business Atlas, and Mexico's Mejora Tu Escuela.
- Use online and offline meet-ups and similar tools to create a culture that encourages knowledge sharing and collaboration. Many off-the-shelf tools already exist. If they are integrated within open-data initiatives or data labs—like the Justice Data Lab in the United Kingdom—they can provide a helpful online supplement to the types of training efforts and expert-mentor networks mentioned above.

Recommendation 6: Identify and Manage Risks Associated with the Release and Use of Open Data

As our case studies have shown, open data can be a force for good, but it is not without risks. Two of the most important risks involve potential violations of privacy and security that can result from widespread releases of data. Such risks were apparent in a number of our case studies, notably Eightmaps, Brazil's Open Budget Transparency Portal, and New York's Business Atlas. Mitigating such risks is essential not only for its inherent value, but also because privacy and security violations undermine trust in open data and, over the long run, limit its potential.

Several steps can be taken to mitigate risks:

- Develop data governance “decision trees” to help decision-makers track the potential risks and opportunities around certain types of data releases. These decision trees can also help weigh the pros and cons and relative risks of data releases.
- Create innovative, collaborative open-data risk-management frameworks so that governments and other institutions releasing data can draw on a clear, structured, step-by-step process to strategically respond to breaches of privacy, security or other risks. NOAA, for example, is working with outside experts to crowdsource new frameworks for data management.

- Involve all stakeholders (including citizen groups) in developing data quality and risk standards. A participatory, collaborative approach to mitigating risks can build trust and help achieve the right balance between social goods like innovation, on the one hand, and risks like privacy and security, on the other hand. Crowdsourcing can be a valuable tool here, giving policymakers a way to solicit a wide range of responses from diverse stakeholder groups.

Recommendation 7: Be Responsive to the Needs, Demands, and Questions Generated from the Use of Open Data

We have seen that public participation is essential in the drafting of open-data policies and in decisions about what data to release. It is equally important in understanding the impact of open data and in taking advantage of the opportunities it offers. For example, open data can generate insights that require government action; open data can likewise reveal inefficiencies that need concrete steps in order to be addressed. And as we have seen in the Brazilian case study on preventing government corruption, meaningful responsiveness requires the ability to take such steps and actions; what's required are communities focused on problem solving, not simply on releasing data.

Meaningful responsiveness can be achieved through the following methods:

- Develop open and online feedback mechanisms, including Q&As, ratings and feedback tools to gauge public opinion and solicit insights from citizens. For example, Denmark's Open Address Initiative has a single portal for users to correct data errors across all agencies. Simplified mechanisms such as this help establish a virtuous open-data cycle, allowing open data to generate insights and ensuring meaningful action on those insights.
- Designate an open-data ombudsman function to consistently track the usefulness of open data and whether necessary follow-up actions are being taken. This ombudsman should itself be

open and transparent, and ideally include a wide range of stakeholder inputs.

Recommendation 8: Allocate and Identify Adequate Resources to Sustain and Expand the Necessary Open-Data Infrastructure in a Participatory Manner

As noted, open-data initiatives are often cheap to get off the ground, but require resources and investment over time. Goals such as increased participation and transparency are laudable, but without resource commitments, they might remain unachievable. Kenya's Open Duka project is a good example of a laudable open-data initiative that has been limited by a lack of resources. Similarly, as of late 2015, Canada's Open Charity Initiative T3010 has not been updated since its original 2013 release, in part due to a lack of funding. This means that anyone seeking recent data on Canadian charities must now scrape information independently.

Adequate resource allocations can be achieved by doing the following:

- Implementing participatory budgeting initiatives, which let citizens choose their priorities and how public funds are allocated. Such initiatives can ensure that the most useful open-data initiatives receive the most funding.
- Undertaking more rigorous cost/benefit analyses of open-data initiatives, which would give policymakers and other stakeholders the means to assess the relative opportunities offered by projects against their costs and possible risks. Among our case studies, NOAA and the UK Ordnance Survey both commissioned cost/benefit studies before launching their projects. This played a vital role in bolstering support and long-term commitments from policymakers and government stakeholders.
- Exploring innovative avenues for funding, especially crowd-sourcing, which can offer the public (and other interested parties) an avenue not only for funding initiatives, but also for establishing and ensuring the sustainability of their priorities.

Recommendation 9: Develop a Common Research Agenda to Move Toward Evidence-Based Open Data Policies and Practices

The most effective avenue to understanding how open data works and how to achieve maximum positive effect, is through collaboration. Our knowledge of open data today is in many ways fragmentary, spread across organizations and individuals who are themselves scattered across the globe. There is a need for more communication and pooling of analysis (and resources). To achieve the potential of open data, we need a common research agenda, based on a wider evidential foundation. Importantly, this research framework should integrate a better understanding of impact into its core agenda. Too often, open data research focuses simply on the best ways of releasing data, with its effect—positive or negative—being simply an afterthought.

To achieve this common research agenda, we should do the following:

- Set up mechanisms for communication and interaction among various stakeholders (individuals and organizations) currently working in the field of open data. Such mechanisms could include annual meetings or conferences, listservs, monthly hangouts, and other offline and online tools. The goal of these interactions would be to trade insights and ideas, to share evidence, and to collaboratively develop best practices. Events like the Open Data Research Summit within the context of the International Open Data Conference can provide, for instance, the impetus toward improved exchange and collaboration among researchers in this field.
- Build on the taxonomy of impact developed through these 19 case studies and have other researchers test the premises we identified earlier. In addition, the open-data research community could consider further fine-tuning of the open-data common assessment framework⁴ GovLab developed together with Web Foundation and others in order to create a standardized tool for evaluating every stage of the open-data value chain.

⁴ <http://opendataresearch.org/sites/default/files/posts/Common%20Assessment%20Workshop%20Report.pdf>

- Create a directory (perhaps in wiki format) of various assessment frameworks (in addition to our own), spread across countries and sectors. Such a directory would also include a list of key contacts and organizations, and would help facilitate discussion by establishing a baseline of sorts toward achieving a common research agenda.

Recommendation 10: Keep Innovating

Open data fuels innovation, but how can we innovate open data? We need to recognize different forms and models of open data, including big and small data and text-based data, and encourage stakeholders to think broadly about what data is and what *open* really means. Even while we work to better understand open data and its effect (for example, through exercises such as this one), we should foster a culture of proactive experimentation and innovation.

There are many ways to foster such a culture. Here are a few:

- Institutionally, we can look at creating new entities or intermediaries, for example a global open-data innovation lab whose explicit purpose would be to think outside the box and research new models of open data that can be tested across sectors, regions and use cases.
- The need for collaborative research mentioned earlier can also be institutionally developed into a cross-border and interdisciplinary open-data innovation network. Such a network would draw on global expertise and ideas.
- Perhaps most important, we need to be open to new ideas and insights, and always remain in *question mode*. This report has outlined several recommendations and suggestions for how to maximize the value of open data. But we recognize that this is just a beginning. Our research has raised as many questions as it has suggested answers.

We end, therefore, with what we believe to be some of the most important questions we should be asking ourselves about open data: questions that can help direct future research, but, perhaps most important, fuel a culture of innovation and flexibility when it comes to how we think about open data.

Key Remaining Questions

The preceding findings and recommendations for policymakers and stakeholders in the open-data community are based on the examination of 19 case studies of open-data initiatives from around the world. Though this effort enabled a major step forward in our understanding of open data and its real and potential impacts, key questions remain, including the following:

- What are the optimal value propositions (e.g., fighting corruption, spurring economic activity, citizens' right to government information) to highlight in order to spur open-data activity in different contexts based on local priorities and needs?
- What are the conditions to scale the effect of open data?
- How can open data initiatives be made sustainable?
- What comparative insights are transferable in a universal manner?
- What is the optimal internal data infrastructure for enabling impactful open-data initiatives?

PART II

Case Studies: Improving Government

Our research shows that open data is helping to improve the functioning of government. By making public previously inaccessible data sources, open data is playing a role in both tackling corruption through increased transparency and accountability, and enhancing public service delivery and resource allocation.

The case studies in this section examine:

- The exposure and a reduction of corruption among public officials in Brazil;
- The discovery of tax fraud among charities in Canada and savings in how development aid is spent in Sweden;
- Procurement improvements in Slovakia; and
- Improved public services in Denmark as a result of greater accuracy in address data.

CHAPTER 2

Brazil's Open Budget Transparency Portal

Making Public How Public Money Is Spent

By Auralice Graft, Stefaan Verhulst and Andrew Young

Summary: In 2004, the Brazilian Office of the Comptroller General (CGU) created the Transparency Portal, a tool that aims to increase fiscal transparency of the Brazilian Federal Government through open government budget data. Developed in partnership with the Federal Data Processing Service, the Transparency Portal relies on the collaboration of diverse ministries and bodies of the Federal Public Administration to advance transparency and to offer a tool that stimulates citizen participation. As the quality and quantity of data on the portal have improved over the past decade, the Transparency Portal is now one of the country's primary anti-corruption tools, registering an average of 900,000 unique visitors each month. The project is regarded as one of the most important e-government initiatives with regard to control over public spending. Local governments throughout Brazil and three other Latin American countries have modeled similar financial transparency initiatives after Brazil's Transparency Portal.

Dimension of Impact: Improving Government—Tackling Transparency Corruption and Transparency

Key Takeaways:

- Open data initiatives can help uncover illegal or irresponsible government spending, engage citizens in anti-corruption cam-

paigns, and lead to meaningful public policy changes that might not occur without the public outrage created by disclosure.

- Anonymous feedback channels are an important way to supplement existing government-maintained databases, especially those focused on corruption. Giving whistleblowers a platform for sharing information can lead to more comprehensive corruption data—although the risk of false positives should not be overlooked.
- Some information, such as in-depth budgetary data, might require some level of translation to make it truly accessible and useful to average citizens in terms of technological availability and comprehensibility. Additional efforts, of course, must be made to ensure that any personally identifiable financial information is anonymized before being made public.

Context and Background

Corruption in Brazil

As with many of the countries included in this series, Brazil has long suffered from corruption. The country is ranked 69 among 175 countries in Transparency International's 2014 Corruption Perceptions Index, which also reports that the country's legal system is "plagued with inefficiencies and corrupt judges."¹ Brazilian elections are also believed to be infiltrated by corruption and, despite public outrage, politicians with prior convictions are often voted back to office. In Rio de Janeiro, for example, Brazil's third-largest state with a population of 11.9 million as of 2012,² "only one single politician in the Brazilian state of Rio de Janeiro is said to have never fallen foul of the law," according to one news report.³ Additionally, a 2009 World Bank and IFC1 Enterprise Survey found that 70 percent of global and domestic companies viewed corruption as a "major con-

1 Corruption Perceptions Index 2014: Results. Transparency International. <http://www.transparency.org/cpi2014/results>

2 Rio de Janeiro metropolitan area profile. Bookings Global Cities Initiative. Accessed October 13, 2015. <http://www.brookings.edu/about/programs/metro/~/media/455849BC16B24B7CA98FE5C5FDE3C0EF.ashx>

3 Barth, Philipp. "The persistence of corruption in Brazilian politics." DW. May 10, 2014. <http://www.dw.com/en/the-persistence-of-corruption-in-brazilian-politics/a-17974970>.

straint to doing business” in Brazil, while a Federation of the Industries of the state of São Paulo investigation found that corruption cost Brazil almost \$40 billion (about 2.3 percent of GDP) in 2008 alone.⁴

Somewhat paradoxically, Brazil is often cited for its strong legal framework aimed at limiting corruption, and is considered a role model among developing countries in terms of its legal and regulatory efforts. While the Penal Code criminalizes the embezzlement of public funds, extortion, public graft, breach of public duty and bribery (passive and active), a 2014 anti-bribery law makes companies liable for acts of corruption committed by their employees. The law applies to the bribing of domestic and foreign officials, and convicted companies must pay a fine of up to 20 percent of their gross annual income and may also face suspension of operations, confiscation of assets, or closure.⁵

Yet despite these formal mechanisms, corruption has proven deep-rooted and difficult to eradicate. Recent major scandals include the Petrobras scandal, which involves money laundering through a network of political corruption and that investigators believe cost state oil giant Petrobras more than \$2 billion.⁶ In addition, media headlines have been dominated by the 2005 Mensalao (“big monthly payment,” in Portuguese) political bribery scandal, where public funds were reportedly used to pay monthly bribes and buy votes during the *Luiz Inácio Lula da Silva* presidency. Additional scandals are described in “Major Incidences of Convicted Officials Returning to Politics in Recent Years” on page 44.

4 Cruz, Marcos and Alexandra Lazarow. “Innovation in government: Brazil.” *McKinsey & Company Insights & Publications*. September 2012. http://www.mckinsey.com/insights/public_sector/innovation_in_government_brazil.

5 “Business Corruption in Brazil.” Business Anti-Corruption Portal. Accessed October 13, 2015. <http://www.business-anti-corruption.com/country-profiles/the-americas/brazil/show-all.aspx>.

6 Savarese, Mauricio. “No End in Sight for Brazil’s Petrobras Scandal.” *Americas Society / Council of the Americans*. August 5, 2015. <http://www.as-coa.org/articles/no-end-sight-brazil%E2%80%99s-petrobras-scandal>.

Major Incidences of Convicted Officials Returning to Politics in Recent Years

Rio de Janeiro's governor Anthony Garotinho was convicted in 2010 for corruption and establishing a criminal organization. He was said to head a gang of militias consisting of former police officers who extort protection money and organized parallel systems of justice. In spite of conviction, his prison sentence was reduced to a community service order and in 2014 he was back in the gubernatorial race, briefly enjoying high ratings in the polls before losing.

In **São Paulo**, former governor Paul Maluf was appointed as member of the Brazilian parliament, despite being known to his fellow countrymen as “Corrupto, mas faz,” which translates as “Corrupt, but hands-on.” He is said to have embezzled hundreds of millions of euros, and, as of 2014, Interpol still had an international warrant out for his arrest. His name Maluf is also the inspiration for the newly minted verb “Malufar” which means “to steal from the state.”

Open Data in Brazil

Brazil is a co-founder of the Open Government Partnership, along with Indonesia, Mexico, Norway, Philippines, South Africa, the United Kingdom and the United States. The government has committed itself to actively developing initiatives that encourage citizen participation in government, and to using technology to create and promote openness. These efforts are led by the National Infrastructure Open Data (INDA), which aims to establish open data technical standards (i.e., common formats and machine-readable metadata, etc.), promote training, and support and encourage government publication of open data.⁷

Some of the open data efforts undertaken by the government include:

- The implementation of Brazil's Access to Information Law in 2011. This law regulates citizens' access to federal, state, provin-

⁷ <http://www.transparency.org/country#BRA>

cial and municipal public documents, as formally guaranteed by the 1988 Constitution.⁸

- The Open Data Portal (described further below) was launched in 2012 to provide citizens with a tool to find and use public information.⁹
- A nationwide conference on transparency (CONSOCIAL) was organized in 2012, involving more than 100,000 Brazilians and providing for a national platform to debate issues relating to transparency, citizen participation in public administration, and the fight against corruption.¹⁰
- A Freedom of Information Law was passed in 2012. While the law has notable advantages—like its applicability to all levels of government, unlike freedom of information laws in other countries—“there is still no set of regulations detailing how citizens can ask for data, and what municipal, state or federal officials must do to comply.”¹¹

In fact, the legal basis for these efforts can be dated back to 1988, when the country’s constitution established the need for “publicity” of administrative actions as one of its five core principles. The constitution also provides for direct citizen participation in overseeing public policies on health, social security and welfare; and freedom of speech and freedom of the press. A 2000 law on fiscal responsibility solidified the government’s commitment to openness, mandating the public availability of key budget documents.

Nonetheless, despite these legal mechanisms to ensure openness, public access to government data was patchy (at best), and scope for citizen participation in government oversight was somewhat limited

⁸ “Brazil: New Access to Information Law becomes effective today.” *Article 19*. May 16, 2012. <https://www.article19.org/resources.php/resource/3208/en/brazil:-new-access-to-information-law-becomes-effective-today>.

⁹ “New Brazilian Data Portal Dados.Gov.Br—Powered by CKAN.” *Open Knowledge Blog*. May 10, 2012. <http://blog.okfn.org/2012/05/10/new-brazilian-portal-dados-gov-br-made-by-citizens/>.

¹⁰ <http://15iacc.org/agenda/agenda-day-4/the-consocial-in-brazil-empowerment-and-governmental-agenda/>

¹¹ Barbassa, Juliana. “Brazil Freedom of Information Law Passed.” *The World Post*. May 17, 2012. http://www.huffingtonpost.com/2012/05/17/brazil-freedom-of-information_n_1525131.html.

for much of Brazil's recent history. The launch of Brazil's Transparency Portal in 2004 marked the country's first concrete steps toward true openness and hopefully, true government accountability.¹²

Project Description and Inception

Over the course of the 2002 presidential campaign, Luiz Inácio Lula da Silva (known as "Lula") ran as an outsider to the political system and made a number of promises to fight corruption and improve transparency. Almost immediately after he took office, civil society groups (and other organizations and even some government agencies) began putting pressure on him to fulfill those promises.

In 2003, the post of the Comptroller General of Brazil (*Controladoria-Geral da União*—CGU) was established; the CGU was an office close to the presidency and was mandated to be in charge of transparency. CGU Chief Minister Walder Pires (also known as the Minister of Control and Transparency) began pushing for several anti-corruption measures, including one to provide transparency on federal government spending. His team, headed by **Jorge Hage**, was assigned the responsibility for implementing this last initiative.

One of Hage's first suggestions was to open data that was already being compiled by the Ministry of Finance and kept in an Integrated System of Financial Administration (SIAFI). This system, which was not accessible to the public, contained fiscal information related to federal transfers to states and municipalities as well as information on funds directly received from public resources (for example, payments from entities with government contracts). However, the Head of Treasury at the time, Joaquim Levy, opposed the idea of opening SIAFI data, arguing that the information was stored in a format and manner that would be too technical for citizens to properly use or understand. According to Leodelma de Marilac Felix, then head of the General Auditing Coordination at the Ministry of Finance, which was in charge of the SIAFI data, Hage's desire to open up the data was otherwise received enthusiastically, and the key issue was finding a way to make it more accessible and usable for citizens.

¹² "OECD Integrity Review of Brazil: Managing Risks for a Cleaner Public Service." OECD Publishing. 2012. http://www.keepeek.com/Digital-Asset-Management/oecd/governance/brazil-oecd-integrity-review_9789264119321-en.

“Most of us involved were very excited. We thought, ‘Great idea,’” said Felix.¹³

It was from this need for a more accessible platform that the seeds of the Transparency Portal were planted. At the time, the CGU did not have the technical capacity to create such a portal, so it collaborated with Serpro, a public company linked to the Ministry of Finance, to develop the platform, which would then be maintained by the CGU. According to Izabela Corrêa, CGU Head of Transparency 2007-2012, opening up the data and making it publicly available was a complex process. The data was stored in a wide variety of formats, and was generally quite inaccessible (even many members of Congress had not been allowed to access it). In general, its readability was very poor, and the technical team faced serious challenges in understanding how the system stored the data.

In June 2003, the CGU set up a working group to study how to extract and combine the data. This working group, which consisted of officials from the CGU, the Secretariat of National Treasury, and the Federal Service of Data Processing, began making some headway, and the result was the Transparency Portal, launched in November 2004.¹⁴ Suggestions provided by citizens via the portal’s “contact us” link and at subsequent portal events were also adopted.¹⁵ The launch of the portal represented a major milestone in Brazil’s move toward openness and transparency. As Waldir Pires, former Minister of Defence, remarked at the time: “With this government project we take an important step toward full transparency of accounts, contributing in effect to societal control over the use of public resources.”¹⁶

¹³ GovLab interview with Leodelma de Marilac Felix, Finance and Control Analyst, Transparency and Public Oversight Directorate, Corruption Prevention and Transparency Secretariat, Office of the Comptroller General, Brazil, August 18, 2015.

¹⁴ “OECD Integrity Review of Brazil: Managing Risks for a Cleaner Public Service.” OECD Publishing. 2012. http://www.keepeek.com/Digital-Asset-Management/oecd/governance/brazil-oecd-integrity-review_9789264119321-en.

¹⁵ “OECD Integrity Review of Brazil: Managing Risks for a Cleaner Public Service.” OECD Publishing. 2012. http://www.keepeek.com/Digital-Asset-Management/oecd/governance/brazil-oecd-integrity-review_9789264119321-en.

¹⁶ <http://www.portaltransparencia.gov.br/noticias/DetalheNoticia.asp?noticia=1> Translated from Portuguese by Google.

Portal Description

According to the CGU website, today the portal is “a channel through which citizens can monitor the financial implementation of government programs at the federal level.” The goal is “to increase the transparency of public management, enabling citizens to track and monitor how public money is being used.”¹⁷ Care has been taken to make sure the information contained on the portal is easily accessible to the public, for example without requiring a username or password. In addition, specialized financial information like budgets is presented in a way that is easy for laypeople to analyze and understand. “Many legal and budget-related terms, in particular, are very complicated,” says portal coordinator Otávio Moreira de Castro Neves. “We try to present them in a language and format that the user will understand.”¹⁸

The portal’s organizers have also taken numerous steps to engage users and spread awareness about the site and how to use it. The site launch was complemented by TV campaigns and workshops designed to train citizens, reporters and public officials to use the portal. In addition, regular workshops and activities were held, including drawing and essay competitions for adults and children.¹⁹ These outreach activities continue to be well attended, and form an important part of the government’s strategy to engage and involve users in the portal.

When initially launched, the portal contained a limited amount of information: data on federal transfers to states and municipalities, payments to the government, information on public servants. Over the years, more information has been added, and today five broad categories of data are published: 1) direct spending by federal government agencies through contracts and tender processes; 2) all

¹⁷ <http://www.cgu.gov.br/assuntos/transparencia-publica/portal-da-transparencia> Translated from Portuguese by Google.

¹⁸ GovLab interview with Otávio Moreira de Castro Neves, Coordinator for Open Government and Transparency, Transparency and Public Oversight Directorate, Corruption Prevention and Transparency Secretariat, Office of the Comptroller General, Brazil. August 6, 2015.

¹⁹ Hage Sobrinho, Jorge. “Brazil’s Transparency Portal Freely Delivers Information to Citizens.” Transparency & Accountability Initiative. September 19, 2011. <http://www.transparency-initiative.org/news/brazils-transparency-portal-freely-delivers-information-to-citizens>.

financial transfers to states, municipalities and the federal district; 3) financial transfers to social program benefactors; 4) administrative spending, including staff salaries, staff travel expenses and per diems and office expenditures; and 5) information on all government official credit card spending. The final decision on what information to upload rests with the CGU.

One important component of the portal is a National Debarment List, which highlights contractors to avoid. The list consolidates in one place all companies and individual suppliers that have been sanctioned for acts of corruption or committing fraud, allowing public officials to exercise some discretion before hiring or granting public tenders. As of September 2011, the Debarment List held 5,018 debarred entities. While this tool was designed primarily to be used by public officials, it has also proved useful for private sector entities. However, this component is not without its flaws. As there is no public register of beneficial ownership, it does not prevent corrupt business owners from launching and registering businesses under new names.

The screenshot shows the 'Portal da Transparéncia' homepage with a green header. Below the header, a search bar is visible with the placeholder 'Acesso rápido: Selecione...'. To the right of the search bar are links for 'Você está em: Início - CEIS'. Below the search bar, a sub-header reads 'CADASTRO NACIONAL DE EMPRESAS INIDÔNEAS E SUSPENSAS (CEIS)'. A note below this states: 'O Cadastro Nacional de Empresas Inidôneas e Suspensas (CEIS) é um banco de informações mantido pela Controladoria-Geral da União que tem como objetivo consolidar a relação das empresas e pessoas físicas que sofreram sanções das quais decorre como efeito restrição ao direito de participar em licitações ou de celebrar contratos com a Administração Pública. [Saiba mais](#)'.

The main search interface includes fields for 'Consulta', 'CPF/CNPJ' (optional), 'Nome, Razão Social ou Nome Fantasia' (optional), 'Tipo de Sangue' (Todos), and buttons for 'Consultar' and 'Limpar filtro'. Below these fields, a message indicates 'Quantidade de registros encontrados: 11730' and 'Data: 26/08/2015 22:04:12'. To the right of the search area is a link 'Imprimir'.

Below the search interface is a table titled 'Dados da Pessoa ou Empresa Sanctionada'. The table has columns: 'Nome', 'Tipo', 'Data Final', 'Nome do Órgão', and 'UF'. The data in the table is as follows:

Dados da Pessoa ou Empresa Sanctionada	Dados de Sanção				
CNPJ/CPF	Nome	Tipo	Data Final	Nome do Órgão	UF
11.972.928/0001-54	SS ARQUITETURA E DESIGN LTDA.	Inidoneidade - Lei de Licitações		Prefeitura de Charqueadas (RS)	RS
11.972.928/0001-54	SS ARQUITETURA E DESIGN LTDA.	Suspensão - Lei de Licitações	16/02/2016	UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL	RS
11.972.928/0001-54	SS ARQUITETURA E DESIGN LTDA.	Suspensão - Lei de Licitações	31/03/2016	UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL	RS
		Suspensão - Lei de Licitações	31/03/2016	UNIVERSIDADE FEDERAL DO RIO GRANDE	RS

At the bottom of the page, there is a footer with the URL 'www.transparencia.gov.br'.

Figure 2-1. The National Debarment List

As of 2010, the portal also includes two new sections dedicated to information on a couple of major sporting events hosted by Brazil: the 2014 football World Cup and the 2016 Olympic Games. Given the large investments in public infrastructure and services involved in these events, the government opted to make all budgetary infor-

mation available on each project. Information is organized by host city and by interest area, such as arenas, airports and security.

The CGU has taken care to include feedback channels to involve citizens and allow them to help improve the site's contents. Every section on the portal includes a survey form or contact link, and there are various ways for whistleblowers to report misconduct or crimes (their identities remain protected). Citizens have used these feedback loops to make suggestions about content on numerous occasions. For instance, citizens and reporters suggested in 2008 that the portal should allow downloads of raw data on administrative agreements between governments and contracted private entities.²⁰ Similarly, various civil society groups suggested in 2006 that users should be able to search by nonprofit organizations receiving government funds. Both these suggestions were later implemented.²¹ Today, data on the portal is updated daily or monthly.²²

Impact

Brazil's Transparency Portal has been widely hailed as a successful example of using open data to reduce corruption and control public spending. Over the years, it has received numerous awards and forms of recognition, including the 2007 Brazilian IT and Government award (e-Democracy category). In addition, it was presented as a Best Practice case study at the 2008 Meeting of the UN Convention against Corruption in Bali; the 2009 Transparency Initiative Conference on International Aid in the Netherlands; and the 2009 Third European Meeting on Anti-Corruption in Brussels.²³

The portal's impact can be measured in several ways:

²⁰ "OECD Integrity Review of Brazil: Managing Risks for a Cleaner Public Service."

OECD Publishing. 2012. http://www.keepeek.com/Digital-Asset-Management/oecd/governance/brazil-oecd-integrity-review_9789264119321-en.

²¹ "OECD Integrity Review of Brazil: Managing Risks for a Cleaner Public Service."

OECD Publishing. 2012. http://www.keepeek.com/Digital-Asset-Management/oecd/governance/brazil-oecd-integrity-review_9789264119321-en.

²² Cruz, Marcos and Alexandra Lazarow. "Innovation in government: Brazil." *McKinsey & Company Insights & Publications*. September 2012. http://www.mckinsey.com/insights/public_sector/innovation_in_government_brazil.

²³ <http://www.portaltransparencia.gov.br/noticias/>

Traffic

Immediately after its inception, in 2004, the portal registered an average of 410,000 monthly visits from around 10,000 unique visitors. By 2012, the number of unique monthly users rose to 336,512; today, the site receives over 900,000 unique visitors per month.²⁴²⁵ The dramatic rise in traffic is a clear indication of the site's relevance and importance in Brazilian public life.

Policy Changes: Official Credit Card Usage

Several changes in the way government operates have directly resulted from information contained on the portal. For example, after the CGU began publishing data on government credit card use, the media began publishing articles on questionable transactions, and this led to a series of scandals. In one instance, the reports led to the resignation of Matilde Ribeiro, Minister of Promotion of Racial Equality.²⁶

Perhaps most importantly, the publication of credit card data led almost immediately to a 25 percent reduction in spending by officials on government cards. In addition, the scandals and rising public awareness led to several changes in credit card policy, including legislation banning the use of cards for payment of travel or per diems; and restrictions on cash withdrawals using official cards (except in very particular situations, for example when an official

²⁴ GovLab interview with Otávio Moreira de Castro Neves, Coordinator for Open Government and Transparency, Transparency and Public Oversight Directorate, Corruption Prevention and Transparency Secretariat, Office of the Comptroller General, Brazil. August 6, 2015.

²⁵ "OECD Integrity Review of Brazil: Managing Risks for a Cleaner Public Service." OECD Publishing. 2012. http://www.keepeek.com/Digital-Asset-Management/oecd/governance/brazil-oecd-integrity-review_9789264119321-en.

²⁶ "Credit happy Brazilian minister admits mistake and resigns." *Merco Press*. February 3, 2008. <http://en.mercopress.com/2008/02/03/credit-happy-brazilian-minister-admits-mistake-and-resigns>.

was located deep in the Amazon, where credit cards are not often accepted).^{27,28}

The Transparency Portal has become a tremendous resource for journalists and media, as well as activist groups investigating corruption. These developments further expand the portal's impact. For example, citizen-group São Paulo Perl Mongers created "Where did my money go?"—a website that presents the portal's data on government spending in a simplified, heavily visual format.²⁹ Meanwhile, the group Open Accounts has been providing Brazilian media with training on how to use the portal, helping them publish stories that "often generate public outrage and a swift official response," such as in 2011, when a story on some government officials' improper credit card use led to an internal investigation. A government minister eventually was forced to repay \$30,000 for "inappropriately charging personal expenses, including vacation travel, to the government."³⁰

Local and Cross-Border Dissemination

As with many of the case studies in this report, the success of Brazil's federal portal has a spillover effect on other governments. In Brazil itself, pressure grew for other levels of government to publish budgetary data, and in 2009, a law was passed requiring that local governments have their own versions of the portal. Today, a large part of CGU's work consists of helping local governments (smaller towns and cities in particular) implement this directive.³¹

27 GovLab interview with Otávio Moreira de Castro Neves, Coordinator for Open Government and Transparency, Transparency and Public Oversight Directorate, Corruption Prevention and Transparency Secretariat, Office of the Comptroller General, Brazil. August 6, 2015.

28 "In 2008, corporate cards yielded CPI's." *Acervo*. January 5, 2013. <http://acervo.estadão.com.br/notícias/acervo.em-2008-cartões-corporativos-renderam-cpis,8813,0.htm>.

29 Cruz, Marcos and Alexandra Lazarow. "Innovation in government: Brazil." *McKinsey & Company Insights & Publications*. September 2012. http://www.mckinsey.com/insights/public_sector/innovation_in_government_brazil.

30 Cruz, Marcos and Alexandra Lazarow. "Innovation in government: Brazil." *McKinsey & Company Insights & Publications*. September 2012. http://www.mckinsey.com/insights/public_sector/innovation_in_government_brazil.

31 GovLab interview with Otávio Moreira de Castro Neves, Coordinator for Open Government and Transparency, Transparency and Public Oversight Directorate, Corruption Prevention and Transparency Secretariat, Office of the Comptroller General, Brazil. August 6, 2015.

Brazil's portal has also had reverberations across Latin America. Many countries expressed interest in the portal early on, and since its inception, at least three—Mexico, Chile and El Salvador—have implemented similar projects, often with assistance from Brazil's CGU.³² Mexico's open data portal (www.catalogo.datos.gob.mx) was launched in October 2013. El Salvador created a similar Fiscal Transparency website (www.transparenciafiscal.gob.sv), and Chile has pushed for greater transparency and openness through its own portal (<http://datos.gob.cl/>). Overall, Brazil's model—and in particular its success—has had a broad impact on Latin America, providing powerful evidence of open data's ability to effect real social, economic and political change.

Changing Climate for Corruption

The Transparency Portal's launch and evolution have taken place within a broader trend of rising intolerance for corruption in Brazil. The 2012 convictions in the country's Mensalao scandal, and prosecutors' willingness to target top officials in the Petrobras scandal, both indicate a possible end to a climate of impunity for corrupt officials. This welcome trend cannot, of course, be directly attributed to the Transparency Portal; it is the result of a number of changes in Brazilian society and politics. But the launch and widespread use of the portal can be considered both a result of Brazil's movement toward greater transparency, as well as one of the dynamic forces continuing to nudge the movement forward today.

The portal also enables internal corruption monitoring that may not have been possible, or as swift, as before. In 2008 CGU created a team of analysts, The Public Spending Observatory, to proactively comb through the portal (and other government databases) to identify “suspicious patterns.”³³ By creating a team dedicated to identifying corruption, the CGU has ensured that the information housed

³² GovLab interview with Otávio Moreira de Castro Neves, Coordinator for Open Government and Transparency, Transparency and Public Oversight Directorate, Corruption Prevention and Transparency Secretariat, Office of the Comptroller General, Brazil. August 6, 2015.

³³ Cruz, Marcos and Alexandra Lazarow. “Innovation in government: Brazil.” *McKinsey & Company Insights & Publications*. September 2012. http://www.mckinsey.com/insights/public_sector/innovation_in_government_brazil.

on the portal will be put to use, whether or not the general public maintains their current level of interest in accessing the portal's data.

In March of 2015, Transparency International stated that Brazil “has made good progress in the past three years passing key anti-corruption laws” (while noting that enforcement of these laws have been limited and that corruption is still a problem).³⁴ Or, as CGU Chief Minister Jorge Hage put it in an interview four years earlier, “Here in Brazil, despite much that has been done in recent years, much remains to be done.”³⁵



Figure 2-2. The Brazilian Transparency Portal's Section Dedicated to the 2014 World Cup

Challenges

Despite its generally positive reaction and relative success, the Transparency Portal, like many projects included in this report, has and continues to face some hurdles to further growth. Three challenges, in particular, are worth examining more closely:

³⁴ “The cost of corruption in Brazil.” Transparency International. March 13, 2015. http://www.transparency.org/news/feature/the_cost_of_corruption_in_brazil

³⁵ “Combating corruption in Brazil depends on agility in ritual procedure, says Brazilian General Comptroller Jorge Hage in interview with UNODC.” *United Nations Office on Drugs and Crime*. February 9, 2011. <https://www.unodc.org/lpo-brazil/en/impressa/intervistas/2011/02/09-combate-a-corrupcao-no-brasil-depende-de-maior-agilidade-no-rito-processual-affirma-jorge-hage-em-entrevista-ao-unodc.html>

Interoperability and Data Standards

As with any new platform, the Transparency Portal deals with several “legacy” issues concerning the way information has historically been collected, stored and disseminated in Brazil. One of its most acute challenges has been to assemble data from various government agencies and present that data in a cohesive, unified manner. This applies most acutely to legacy data, but even for more recent information, CGU officials have had to negotiate with officials and agencies over the manner in which their data is collected and shared. Dealing with such a variety of data formats and systems not only slows down the process, it also increases the risk of introducing errors. As Otávio Moreira de Castro Neves, Coordinator for Open Government and Transparency, puts it: “As we put data up online we are constantly under the risk of human error, or under the risk of delays. So we have to keep a tight grip on this.”³⁶

Leodelma de Marilac Felix, Finance and Control Analyst at the Transparency and Public Oversight Directorate, provides some examples of the types of errors CGU must work to prevent. He cites the (hypothetical) example of a bank deciding to change a particular data field (e.g., the way a date is stored) in its data sets, but doing so without informing CGU. According to Felix, it can take CGU up to four months to identify the problem.³⁷

Such issues are not insurmountable, nor are they unique to the Transparency Portal. For example, CGU is now looking into changing its application rules for how information is received; this could help standardize data. Still, CGU is unlikely to be able to ever fully resolve the issue, and as with most data systems and projects, some degree of error-checking and validation will always be required. However, making the data accessible increases the likelihood that data errors will be corrected, as more people will be using the data, and using it more often, than before. “These are not huge challenges. We can overcome them,” says Izabela Corrêa, former Coor-

³⁶ GovLab interview with Otávio Moreira de Castro Neves, Coordinator for Open Government and Transparency, Transparency and Public Oversight Directorate, Corruption Prevention and Transparency Secretariat, Office of the Comptroller General, Brazil, August 6, 2015.

³⁷ GovLab interview with Leodelma de Marilac Felix, Finance and Control Analyst, Transparency and Public Oversight Directorate, Corruption Prevention and Transparency Secretariat, Office of the Comptroller General, Brazil, August 18, 2015.

dinator for the Promotion of Ethics, Transparency and Integrity, at Brazil's Directorate for Corruption Prevention. "But they are always there. Always."³⁸

Privacy

The need to protect privacy has been a constant, and delicate, issue for the portal. Portal officials report that a large number of the complaints they receive pertain to privacy issues. For instance, more than 100 legal actions were placed against the portal for the publication of public servant salaries; the issue made its way to Brazil's supreme court, which eventually ruled in favor of the portal.³⁹ Portal officials also receive a deluge of emails from public servants whose data shows up on Google; even when such officials realize they can't be removed from the portal, they request portal staff to help them remove their names from Google.

One major issue pertained to the publication of data sets related to government real estate. This information was put up because public servant occupation of real estate is subject to numerous rules—related, for instance, to officials' actual titles and posting—and the portal wanted to provide citizens with information that would help them oversee that the rules were being followed. However, the problem was that publishing real estate data inadvertently revealed the addresses of public officials, which not only violated their privacy but also potentially threatened their security. CGU eventually dealt with this issue by publishing two separate data sets—one with names, and the other with addresses.

Scaling

A final challenge relates to the scaling and wider dissemination of the platform. As mentioned previously, a 2009 law required various regional governments in Brazil to implement a version of the portal. There have been many hiccups in the implementation of this law. A

³⁸ GovLab Interview with Izabela Corrêa, former Coordinator for Promotion of Ethics, Transparency and Integrity, at the Directorate for Corruption Prevention, 2007 to 2012. Interview on August 21, 2015.

³⁹ "Brazilian Supreme Court Upholds the Mandatory Disclosure of Salary Information for Public Officials." *Transparency Audit Network*. May 2015. <http://www.transparencyaudit.net/news/2015-05/brazilian-supreme-court-upholds-mandatory-disclosure-salary-information-public>

2011 study by the *Instituto de Estudos Socioeconômicos* (INESC), a nonprofit focused on “strengthening civil society and enhancing social participation in public policymaking,”⁴⁰ looked at whether local governments were actually opening up data in accordance with the law and with open data principles. While the study found that the data provided was mostly complete, it was often not primary (meaning that the data was not directly sourced by the government) and frequently not provided in a timely manner.

The study also found various accessibility problems, including the fact that very few websites actually offered machine-readable data in nonproprietary formats. The overall conclusion of the report was that subnational governments need to work considerably to improve the open data process and make more information available in a citizen-friendly format.⁴¹

“These are not huge challenges. We can overcome them. But they are always there. Always.”

—Izabela Corrêa, Brazil’s Directorate for Corruption Prevention

Looking Forward

Since its launch in 2004, the portal has been regularly updated and improved. In general, new features have been added to the site at least once a year. Since 2010, many of the most important data sets, such as those on income and spending, are updated on a daily basis. Other data—including that on money transfers between different levels of government—is updated weekly.

As new data is added, the portal team has made design improvements to the user interface to make the data easier to interpret. At the moment, CGU officials are working on an entirely new version of the site. This updated version will not only have a fresh design, but it will also have a new structure and be based on a more advanced underlying technical framework that will, among other things, provide for new ways of visualizing data. CGU is also build-

40 <http://www.inesc.org.br/institutional/know-inesc>

41 Beghin, Nathalie, Carmela Zigona, Gisele Craveiro, Marcelo Tavares. “Open Data in Brazil: budget transparency and people’s rights.” Open Data Research Network. 2014. [http://www.opendataresearch.org/sites/default/files/publications/INESC%20-%20ODDC%20Poster_Inesc_Brazil%20\(1\).pdf](http://www.opendataresearch.org/sites/default/files/publications/INESC%20-%20ODDC%20Poster_Inesc_Brazil%20(1).pdf).

ing a new data warehouse, which will be launched sometime in 2016.⁴²

In order to surmount Brazil's low Internet penetration rates, which stand at around 53 percent⁴³ and limit the portal's accessibility, CGU piloted a project to set up information kiosks run by state-owned banks. However, since the costs of this project were deemed too high, CGU is now focusing on overall strategies to bridge the digital divide in Brazil, and is working with various government and non-governmental partners to achieve this goal.

These various changes and improvements are likely to enhance the Transparency Portal's offerings and reach in coming years, building on its initial successes and playing a role in Brazil's ongoing battle against corruption. Of course the portal is but one of many tools in Brazil's civic toolbox. But it is an important example of the role that open data—and, more generally, government transparency—can play in transforming political, economic and social life.

42 GovLab interview with Otávio Moreira de Castro Neves, Coordinator for Open Government and Transparency, Transparency and Public Oversight Directorate, Corruption Prevention and Transparency Secretariat, Office of the Comptroller General, Brazil. August 6, 2015.

43 <http://www.internetlivestats.com/internet-users/brazil/>

CHAPTER 3

Openaid in Sweden

Enhanced Transparency and Accountability in Development Cooperation

By Ali Clare, Stefaan Verhulst and Andrew Young

Summary: Sweden has a long tradition of openness, democracy and public access to information. In 2010, a reform agenda for Swedish development cooperation (“Openaid”) was launched by the government to bring increased transparency to donor funding through opportunities created by technological advances. Part of this reform agenda included an Aid Transparency Guarantee that required public actors to make available all documentation and public information related to international development cooperation. This spurred the development of the *openaid.se* site by the Swedish Ministry of Foreign Affairs and the Swedish International Development Cooperation Agency (Sida) in April 2011. The data hub, built on open government data, visualizes *when*, to *whom* and *why* aid funding was paid out and *what* the results were. The reforms are seen to be an important force for enhanced transparency and accountability in development cooperation at an international level and increased cooperation and involvement of more actors in Swedish development policy.

Dimension of Impact: Improving Government—Tackling Corruption and Transparency

Key Takeaways:

- Open data can be used to increase the transparency and accountability of distinct groups simultaneously. While traditionally viewed as a means for holding government to account, *openaid.se* demonstrates the capability of holding, for example, both donors and recipients accountable for aid expenditures. The *openaid.se* platform also illustrates how the Swedish principle of open access to information can be updated in the age of digitization.
- International standards—such as the International Aid Transparency Initiative (IATI)—can help governments and other bodies releasing data do so in a manner that enables the highest levels of comparability and broad use. Metadata—in this case, project documentation and geocoding—can further improve the usefulness of standardized data sets.
- A lack of clearly defined, high-level political commitments to publishing open data and enabling reuse can create major, but not insurmountable challenges. While Sweden now boasts such high-level commitments, throughout much of *openaid.se*'s development, no such policy existed.

Context and Background

Transparency and access to public information has a long-standing tradition in Sweden. The government was the first country in the world to enact a bill enforcing the principle of public access to information (“offentlighetsprincipen”) in 1776. This made it a requirement for all authorities to publish documents unless ad hoc legislation restricted their access.¹ An unimpeded view by the public and media into governmental activities is still prioritized today in Bill 2009/10:175 on *Public Administration for Democracy, Participa-*

¹ Östling, Alina. “Independent Reporting Mechanism Sweden: Progress Report 2012-13.” Open Government Partnership. 2014. <http://www.opengovpartnership.org/country/sweden/progress-report/report>

*tion and Growth.*² Underlying this legislation is the belief that more accessible information provides a better basis for decisions and limits the scope for corruption and misuse of resources.³ Sweden's commitment to openness and public scrutiny often places it near the top of transparency rankings.⁴

As part of its commitment to public access, Sweden launched an Aid Transparency Guarantee in 2010. Pursuant to the guarantee, all public actors that are allocated funds under international development cooperation are required to publish related information and documentation in an open format online.⁵ This includes an explanation of when, to whom and why money was made available, and what results have been achieved.⁶ In theory, such information enables relevant stakeholders to follow the whole chain of aid from overall decisions to implementation and monitoring.⁷

The manifestation of the Transparency Guarantee is the Openaid website (www.openaid.se) jointly launched in April 2011 by the Swedish Ministry of Foreign Affairs and the Swedish International Development Cooperation Agency (Sida). The *openaid.se* data hub has several policy objectives: to promote active transparency, provide a better knowledge base for planning, steering and making decisions about Swedish aid allocations and aid policy priorities, increase participation in Swedish development cooperation,

² "Open Government Partnership—Sweden's 2012 National Action Plan—Self-Assessment Report." <http://www.opengovpartnership.org/sites/default/files/OGP%20Sweden's%20Self-assessment%20Report%20-%20FINAL.pdf>; "Sweden's OGP Action Plan 2014-2016." Open Government Partnership. <http://www.opengovpartnership.org/es/files/swedens-ogp-action-plan-2014-2016-pdf/download>

³ "Sweden's OGP Action Plan 2014-2016." Open Government Partnership. <http://www.opengovpartnership.org/es/files/swedens-ogp-action-plan-2014-2016-pdf/download>

⁴ Sweden is ranked sixth out of 68 aid donors by Publish What You Fund, third among bilateral donors and in the top-performing group in the 2014 *Aid Transparency Index*. "Sweden and Sida ranks high in aid transparency index." Sida. October 16, 2014. <http://www.sida.se/English/press/current-topics-archive/2014/sweden-and-sida-ranks-high-in-aid-transparency-index/>

⁵ "A transparency guarantee in Swedish development assistance." Regeringskansliet. 2010. <http://www.regeringen.se/contentassets/38eb101044b640b6b25677b714ed302e/a-transparency-guarantee-in-swedish-development-assistance>

⁶ "Sweden's OGP Action Plan 2014-2016." Open Government Partnership. <http://www.opengovpartnership.org/es/files/swedens-ogp-action-plan-2014-2016-pdf/download>

⁷ "Sweden's OGP Action Plan 2014-2016." Open Government Partnership. <http://www.opengovpartnership.org/es/files/swedens-ogp-action-plan-2014-2016-pdf/download>

strengthen the preconditions for true accountability, restrict the space for corruption, duplication and inefficient use of resources and promote innovative thinking across different sectors related to development.⁸

The reform of aid development cooperation was inspired in part by commitments that had been made in the international fora, including IATI and the Paris Declaration, Accra Agenda for Action.⁹ The platform was also a core component of Sweden's Open Government Partnership (OGP) National Action Plan, signed in September 2011. The Action Plan was focused on enabling the government to stay on track with planned proposals, as well as increasing the amount of input from civil society and committing to publish standardized data in the IATI format.¹⁰ Sweden's commitment to reform in this sector also led the government to sign the Busan Partnership Agreement for Effective Development Cooperation, which sets time-bound commitments to fully publish aid information to a common, open standard.¹¹ It also offers a framework for "continued dialogue and efforts to enhance the effectiveness of development cooperation" through information access on aid flows and activities in both donor and partner countries.¹²

Product Description and Inception

First launched in April 2011, *openaid.se* is a Web-based information service about Swedish aid built on open government data.¹³ The site enables the public, aid actors and other stakeholders to follow when, to whom and for what purposes aid funds have been disbursed, and

8 GovLab interview with Hanna Hellquist, former State Secretary for International Development, Sweden, September 30, 2015.

9 GovLab interview with Hanna Hellquist, former State Secretary for International Development, Sweden, September 30, 2015.

10 <http://www.opengovpartnership.org/country/sweden>

11 "The Busan Partnership for Effective Development Cooperation." OECD. July 2012. <http://www.oecd.org/dac/effectiveness/Busan%20partnership.pdf>

12 "The Busan Partnership for Effective Development Co-operation." OECD. <http://www.oecd.org/development/effectiveness/busanpartnership.htm>

13 <http://www.openaid.se/about/>

with what results.¹⁴ The interface of the platform is extremely simple and intuitive, enabling anyone to use it.

The website is built using public data at the activity level of individual aid contributions from the Ministry for Foreign Affairs, the Swedish International Development Cooperation Agency (Sida) and other authorities and ministries that handle aid funds.¹⁵ The data is visualized as treemaps and graphs showing how and where aid is distributed by recipient type.¹⁶ Depending on the choice of filter, users also have the option of viewing a list of all activities, ranging from one to several thousand for each recipient.¹⁷ These activities can then be downloaded in an Excel compatible comma separated values (CSV) file and used to develop new applications.¹⁸

Other functionalities include visibility of the entire aid chain and activity structure displaying the link between, for example, a policy decision and a disbursement.¹⁹ A contact form is also available on the site in Swedish and English, allowing for citizens to directly share feedback and concerns with Sida. There is also an interface for mobile devices with customized interaction and design and a mobile app.²⁰ One of the key features for this mobile app is a whistle-blower function, enabling users to report suspicions of fraud directly.²¹

14 <http://www.openaid.se/about/>

15 <http://www.openaid.se/about/>

16 <http://www.publishwhatyoufund.org/updates/by-country/sweden/making-openaid-se/>

17 Peterson, Karl. "The making of Openaid.se." Publish What You Fund. September 23, 2014. <http://www.publishwhatyoufund.org/updates/by-country/sweden/making-openaid-se/>

18 GovLab Interview with Carl Elmstam, Transparency Manager, Sida, October 8, 2015.

19 Östling, Alina. "Independent Reporting Mechanism Sweden: Progress Report 2012-13." Open Government Partnership. 2014. <http://www.opengovpartnership.org/country/sweden/progress-report/report>

20 Östling, Alina. "Independent Reporting Mechanism Sweden: Progress Report 2012-13." Open Government Partnership. 2014. <http://www.opengovpartnership.org/country/sweden/progress-report/report>

21 Östling, Alina. "Independent Reporting Mechanism Sweden: Progress Report 2012-13." Open Government Partnership. 2014. <http://www.opengovpartnership.org/country/sweden/progress-report/report>

Underlying Data

The data available on the website is published according to the IATI standard on a monthly basis, making it easier to analyze and compare data sets from various sources. This commitment to a common standard of data is part of Sweden's move toward implementing the Busan Outcome Document, which aims to have the full range of information publicly available in one open aid standard.²²



Figure 3-1. Sweden's Openaid.se

Over 80 percent of the data is currently available in a machine-readable format, with fully automated data collection each night.²³ This includes all data provided by the 16 CSOs that have framework agreements with Sida.²⁴ The first version of the site published over 100,000 documents at the activity level that could be sorted by disbursing agency or implementing partner.²⁵ Since the launch, there have been over 15 additional updates, and the government has constantly sought to improve the quality of the data—including a focus

22 Östling, Alina. "Independent Reporting Mechanism Sweden: Progress Report 2012-13." Open Government Partnership. 2014. <http://www.opengovpartnership.org/country/sweden/progress-report/report>

23 GovLab interview with Karl Peterson, Project Manager, Aid Transparency, Sida, September 23, 2015.

24 Peterson, Karl. "The making of Openaid.se." Publish What You Fund. September 23, 2014. <http://www.publishwhatyoufund.org/updates/by-country/sweden/making-openaid-se/>

25 GovLab Interview with Carl Elmstam, Transparency Manager, Sida, October 8, 2015.

on publishing results and evaluations data.²⁶ There has also been a focus on adding new types of data in a form that can be aggregated. By the end of 2015, Sweden aimed to have 95 percent data compatibility with the IATI standard.²⁷ There is also an ongoing integration between the CSO database and *openaid.se* that will eventually enable the government to show more detailed data and results for CSO activities funded by Sida.²⁸

Open Format

Openaid.se is built as an open source WordPress site enabling other aid donors and recipients to make their own aid tracker installations using their own data and themes.²⁹ This has huge potential value in terms of spreading transparency across the aid ecosystem as a whole. Sida has also enhanced the use of the *openaid.se* open format (API) so that consumers are able to reuse the data for third-party online API collections. Because of Sweden's public access principles, this data belongs to the public domain by default. However, some restrictions are applied to sensitive or classified materials that are filtered out or materials where copyright is owned by someone else. In the case of the latter, the data is still available on *openaid.se*; however, it may not be republished without permission from the original copyright holder.³⁰

²⁶ GovLab Interview. Karl Peterson. September 23, 2015; IRM, Sweden: Östling, Alina. "Independent Reporting Mechanism Sweden: Progress Report 2012-13." Open Government Partnership. 2014. <http://www.opengovpartnership.org/country/sweden/progress-report/report>

²⁷ Östling, Alina. "Independent Reporting Mechanism Sweden: Progress Report 2012-13." Open Government Partnership. 2014. <http://www.opengovpartnership.org/country/sweden/progress-report/report>

²⁸ "Open Government Partnership—Sweden's 2012 National Action Plan—Self-Assessment Report." <http://www.opengovpartnership.org/sites/default/files/OGP%20Sweden's%20Self-assessment%20Report%20-%20FINAL.pdf>; "Sweden's OGP Action Plan 2014-2016." Open Government Partnership. <http://www.opengovpartnership.org/es/files/swedens-ogp-action-plan-2014-2016-pdf/download>

²⁹ Peterson, Karl. "The making of Openaid.se." Publish What You Fund. September 23, 2014. <http://www.publishwhatyoufund.org/updates/by-country/sweden/making-openaid-se/>

³⁰ "Open Government Partnership—Sweden's 2012 National Action Plan—Self-Assessment Report." <http://www.opengovpartnership.org/sites/default/files/OGP%20Sweden's%20Self-assessment%20Report%20-%20FINAL.pdf>; "Sweden's OGP Action Plan 2014-2016." Open Government Partnership. <http://www.opengovpartnership.org/es/files/swedens-ogp-action-plan-2014-2016-pdf/download>

Impact

Transparency is crucial to drive improvements in the way development cooperation is delivered internationally, especially for generous aid donors like Sweden who allocate 1 percent of their estimated gross national income (GNI) toward development assistance.³¹ In terms of impact, Sweden is playing a leading role among major donors through its innovative *openaid.se* platform which has received wide acclaim for publishing information that goes well beyond traditional reporting to include project documentation, geocoding and aid results data. These efforts have borne fruit—Sweden is one of the highest-ranked countries in the Transparency International's 2014 Corruption Perception Index³² and is one of the highest-ranked donors in the Publish What You Fund 2014 Aid Transparency Index.³³ Impact can be gauged according to increased donor transparency and accountability with follow-on effects to data quality as well as funding and management efficiencies:

Increased Transparency and Accountability

By opening access to the chain of aid delivery and expenditures, citizens and other key stakeholders have been able to hold the Swedish government to account and aid processes have become more efficient through improved service delivery, reduced opportunities for diversion and therefore corruption.³⁴ The increased transparency has also enabled donor recipients to plan and manage the resources coming into their country more effectively in settings where multiple actors are active, thereby decreasing the incentive to provide inferior services.³⁵ As such, the *openaid.se* platform has received

31 “Sweden’s OGP Action Plan 2014-2016.” Open Government Partnership. <http://www.opengovpartnership.org/es/files/swedens-ogp-action-plan-2014-2016-pdf/download>

32 <http://www.transparency.org/cpi2014>

33 Sweden was ranked “Very Good” with an indicator score of 83.27 percent, a significant increase over its 2013 ranking. <http://ati.publishwhatyoufund.org/donor/sweden/Sweden>

34 Faust, Jörg. “Donor Transparency and Aid Allocation.” German Development Institute Discussion Paper. December 2011. http://edoc.vifapol.de/opus/volltexte/2012/3552/pdf/DP_12.2011.pdf

35 Rådelius, Elias. “White paper, part 4: Keys to success and the road ahead.” Openaid.se Blog. May 28, 2015. <http://www.openaid.se/blog/part-4-keys-to-success-and-the-road-ahead/>

strong political and management support and organization involvement at all levels.³⁶

By opening aid data to public scrutiny, the demand on the government to improve and maintain high-quality data has also increased. According to Carl Elmstam, the transparency manager at Sida, “The process of implementing transparency and the IATI standard drives quality as it has forced us to take a hard look at our own data and learn from it.”³⁷ According to Hanna Hellquist, former State Secretary for International Development, opening the data also enabled the government to fully understand the underlying problems and challenges that were built into Swedish aid reporting and IT systems.³⁸ This understanding, coupled with external pressure, has forced Sida to constantly improve reporting processes and ameliorate any data flaws.³⁹

This sort of transparency has the potential to radically decrease the incentive structure for corruption by enabling external stakeholders to pinpoint problems and support their demands for reform. It has also enabled aid to be targeted more specifically and in some cases, it has led to funds being earmarked for improving recipient-country governance, including projects that combat corruption.⁴⁰

This would not have been possible without active engagement and strategic communication both internally with employees and externally with the public and recipient countries. Internally, the MFA and Sida started to design the site before any formal decisions had been made. To push the decision through, they held regular meetings with different actors within the government to address objec-

36 Rådelius, Elias. “White paper, part 4: Keys to success and the road ahead.” Openaid.se Blog. May 28, 2015. <http://www.openaid.se/blog/part-4-keys-to-success-and-the-road-ahead/>

37 Rådelius, Elias. “White paper, part 4: Keys to success and the road ahead.” Openaid.se Blog. May 28, 2015. <http://www.openaid.se/blog/part-4-keys-to-success-and-the-road-ahead/>

38 GovLab interview with Hanna Hellquist, Former Secretary of State, Sweden, September 30, 2015.

39 Rådelius, Elias. “Implementing aid transparency in Sweden—white paper, part 1: Why do we care about aid transparency and IATI?” Openaid.se Blog. May 25, 2015. <http://www.openaid.se/blog/implementing-aid-transparency-in-sweden-white-paper-part-1-why-do-we-care-about-aid-transparency-and-iati/>

40 GovLab interview with Hanna Hellquist, former State Secretary for International Development, Sweden, September 30, 2015.

tions and link the initiative to broader commitments made in other parts of government, such as the PSI Directive, which calls for public authorities to release their data for reuse at no charge.⁴¹ In doing so, staffs were reminded of the importance of good and reliable data, not only for the organization but also for global development cooperation and partner countries.⁴² Externally, Sida worked to increase general awareness of the values of transparency, open data and data quality through communication and facilitation of reporting and data conversion for the CSOs they support.⁴³ According to Hellquist, the opening of data has facilitated a cultural and attitude shift in the aid administration (MFA and Sida) regarding its relation to the public.⁴⁴

Improved Management Efficiency

The *openaid.se* platform has also shown substantial impact in terms of having more efficient reporting mechanisms, including improved information management systems and reduced costs for duplicate manual reporting of aid information, particularly at the country office level. These improved reporting mechanisms, in part spurred by the commitment to make their data available in the IATI standard, have enabled the government to save around US\$7 million annually. It has also had follow-on benefits in terms of aid effectiveness, as the government has been able to analyze and compare data sets from partner country budget classifications, enabling stakeholders to see where aid is supporting their own priorities and increasing accountability. According to conservative estimates from IATI, approximately US\$18 billion in aid is lost annually through corruption. Cost-benefit analyses show that greater aid transparency has the potential to reduce such corruption and improve predictability by US\$1.6 billion annually.

41 <http://www.government.se/about-the-website/psi-data/>

42 GovLab Interview. Carl Elmstam. September 23, 2015. <http://www.openaid.se/blog/part-4-keys-to-success-and-the-road-ahead/>

43 Rådelius, Elias. “White paper, part 2: How to get started publishing using the IATI standard as a bilateral donor agency” Openaid.se Blog. May 26, 2015. <http://www.openaid.se/blog/part-2-how-to-get-started-publishing-using-the-iatи-standard-as-a-bilateral-donor-agency/>

44 Hanna Hellquist, GovLab interview with author. 30 September 2015.

Enabling Collaboration with Beneficiaries

Making aid information open is just the start. For transparency to be transformational, beneficiaries must actually use the aid data. One impact of the newly freed data on the *openaid.se* platform has been the ability of Sida to tap into the knowledge and creativity of a wide variety of stakeholders in Swedish society to come up with new ideas for tools, activities and functionalities for the site.⁴⁵

It has also facilitated substantial and complex forms of collaboration with CSOs including PWYF, in pursuit of increased aid transparency. This has shifted the power dynamics between NGOs and the government, bringing the practice of the organization nearer to its principles of participation, accountability and empowerment. In this respect, Sida's efforts have also served to improve other donors' and partner countries' development efforts by acting as a model for countries like Denmark, which recently launched a similar platform —Danida Open Aid.⁴⁶ While it is difficult to assess whether the idea was explicitly emulated, the fact that *openaid.se* uses an open source Wordpress site removes the need for other countries to start from scratch when building their own initiatives.

Challenges

Sweden has made huge strides toward more open, inclusive, accountable and responsive development through the *openaid.se* site, driving both political debate and opinion about aid priorities. Development coordination and spending between donors and partner countries have also improved, enabling greater impact with finite resources. Nonetheless, the government still faces numerous challenges in making the data easily comprehensible with the right tools to facilitate understanding, analysis and use by general audiences.

⁴⁵ "Guest blog from OpenAid—Sweden embraces the transparency agenda." *AidInfo*. July, 21, 2010. <http://www.aidinfo.org/guest-blog-from-openaid-sweden-embraces-the-transparency-agenda.html>

⁴⁶ <http://openaid.um.dk/en/>

ces.⁴⁷ If Sweden is to continue along the path to aid transparency on which it has set itself, several areas should be addressed:

Structural Challenges

Openaid.se is a proprietary system with a funder-centric view.⁴⁸ This means that there are often challenges in accessing and aggregating information and activities from recipient systems, given the software was tailor-made for Sweden.⁴⁹ Accordingly, it is often difficult to get a true understanding of the impact of development aid and much of the data on the platform hasn't been reused, for example, for building applications.⁵⁰ To overcome these structural problems, *openaid.se* could create novel ways to combine data with other systems through the integration of new tools. Easily comparable data would lead to better impact analyses to understand the effects of aid, to more targeted and efficient innovations, more openness and accountability.⁵¹

Data Privacy

The active and instant nature of publishing aid data has exposed privacy concerns among Sida staff, despite the fact that the information was publicly available before the development of *openaid.se*.⁵² This has raised the need to think about *how* transparent the site should be and what tools could be implemented to mitigate apprehensions about data privacy, including the use of unofficial notes and com-

47 Rådelius, Elias. "Implementing aid transparency in Sweden—white paper, part 1: Why do we care about aid transparency and IATI?" Openaid.se Blog, May 25, 2015. <http://www.openaid.se/blog/implementing-aid-transparency-in-sweden-white-paper-part-1-why-do-we-care-about-aid-transparency-and-iat/>

48 Bjelkeman-Pettersson, Thomas. "OpenAid.se, Swedish Development Aid Transparency." Open for Change, April 4, 2011. <http://openforchange.info/contant/openaidse-swedish-development-aid-transparency>.

49 GovLab Interview with Carl Elmstam, Transparency Manager, Sida, October 8, 2015.

50 Bjelkeman-Pettersson, Thomas. "OpenAid.se, Swedish Development Aid Transparency." Open for Change, April 4, 2011. <http://openforchange.info/contant/openaidse-swedish-development-aid-transparency>.

51 Dietrich, Daniel. "Open Data in Development Aid." European Public Sector Information Platform Topic Report No. 2012 / 02. March 2012. <http://www.epsiplatform.eu/sites/default/files/Final%20TR%20Open%20Aid%20Data.pdf>

52 GovLab interview with Karl Peterson, Project Manager, Aid Transparency, Sida, September 23, 2015.

ments only available to staff.⁵³ Staff buy-in is essential for the continued growth of the site. As such, Sida should continue to prioritize communication to staff about the values of transparency and accountability while remaining responsive to their concerns for data privacy.

Usability and Citizen Engagement

When Sida first launched *openaid.se* in 2011, user experience wasn't a key priority. As such, the site offers limited opportunities for citizen engagement or interaction. The only feedback channel is an opinion button which people use for reporting bugs.⁵⁴

Usability is also a concern, and project titles often contain cryptic terms that are interpretable only to those who work internally on projects, making them difficult to understand for a wider audience.⁵⁵ As such, there is a need to raise awareness among staff to ensure the accessibility of data with outside users.

Data Licensing

During the initial development and deployment of *openaid.se*, Sweden's data policy environment presented numerous challenges. Until May 2015, there were no specific regulations regarding open data and open data sources.⁵⁶ The PSI Directive helped to fill this jurisdictional vacuum in relation to the reuse of data within Sweden. Prior to the PSI Directive's introduction, however, *openaid.se* had neither high-level policy support nor clearly defined licensing guidelines or standards. As Sweden continues to take steps to enshrine the legal right to access and reuse public data, early challenges related to clear, legal data licensing structures for *openaid.se* data should continue to wane.

⁵³ GovLab interview with Karl Peterson, Project Manager, Aid Transparency, Sida, September 23, 2015.

⁵⁴ Östling, Alina. "Independent Reporting Mechanism Sweden: Progress Report 2012-13." Open Government Partnership. 2014. <http://www.opengovpartnership.org/country/sweden/progress-report/report>

⁵⁵ Rådelius, Elias. "White paper, part 3: Challenges" Opendata Blog. May 27, 2015. <http://www.openaid.se/blog/part-3-challenges/>

⁵⁶ Månsson, Christer, et al. "From Bureaucracy to Innovation: An introduction to How to Work with Open Data," March 2013. <http://luii.lu.se/new-report-on-open-data/>

Looking Forward

In order to address the above challenges and continue to expand upon the success of the *openaid.se* hub in terms of meaningfully increasing transparency, Sweden should consider implementation of more forward-planning data and increase the functionality of the site.

Usability and functionality

While Sweden has taken significant steps to improve the usability and functionality of the *openaid.se* platform since its launch in 2011, much of the information is still too complex for lay citizens to digest in a meaningful way. Furthermore, given the funder-centric nature of the site, the opportunities for citizens to engage and interact are limited. To continue to improve upon the functionality, more intuitive data visualization and storytelling should be employed to present the large amounts of information available in a way that illustrates the impact of aid funding for the public.⁵⁷ Tools developed by other organizations could be integrated into the current system to enable better field monitoring and evaluation. For example, tools like Water for People's Field Level Observations Watch (FLOW)⁵⁸ or distributed publishing tools like Akvo Really Simple Reporting (RSR)⁵⁹ could be extremely effective.⁶⁰

⁵⁷ Heacock, Rebekah and David Sasaki, "ICT4 Transparency in Sub-Saharan Africa. In *Increasing Transparency and Fighting Corruption through ICT Empowering People and Communities*," The Swedish Program for ICT in Developing Regions (SPIDER) ICT4D Series No. 3, 2010. Stockholm: Stockholm University. IRM, Sweden: Progress Report 2012-13. http://spidercenter.org/polopoly_fs/1.163640.1390315885!/menu/standard/file/Spider%20ICT4D%20series%203%20Increasing%20transparency%20and%20fighting%20corruption%20through%20ICT.pdf

⁵⁸ FLOW is a "system to collect, manage, analyze, and display geographically-referenced monitoring and evaluation data," mainly used for water points to date. <https://www.waterforpeople.org/what-we-do>

⁵⁹ Akvo Really Simple Reporting is a "web and Android-based system that makes it easy for development aid teams to bring complex networks of projects online and instantly share progress with everyone involved and interested." <http://akvo.org/products/rsr/>

⁶⁰ Bjelkeman-Pettersson, Thomas. "OpenAid.se, Swedish Development Aid Transparency." Open for Change, April 4, 2011. <http://openforchange.info/contant/openaidse-swedish-development-aid-transparency>.

In terms of the distributed publishing and feedback functions, Sida should ensure they remain responsive to the data privacy concerns of staff and potentially launch functionalities to relieve tensions such as unofficial comments and notes that are only internally available.⁶¹

Forward-Planning Data

The publication of more and more detailed forward-planning data by Sweden would have a huge impact on recipient countries.⁶² This could include links to project documents as well as results and conditions data for all planned activities in IATI files to promote access and use by others.⁶³ This would enable greater local ownership over development results and have follow-on effects for increased accountability and governance. In its Open Government Partnership Action Plan, Sweden posits that this aid transparency will also enable full use of the available financial resources for poverty eradication and facilitate better division of labor.⁶⁴ According to Samantha Custer of AidData, “Equipping CSOs and government ministries to use this in program planning, advocacy and research is also essential to sustaining practice.”⁶⁵

Openaid.se should also improve the publication and use of monitoring and performance data. Currently this data is only available in PDF format. In the future, having this data available in the IATI standard would enable the information to be integrated into an ecosystem of tools for field monitoring and evaluation, leading to a more streamlined flow.⁶⁶

61 GovLab Interview with Carl Elmstam, Transparency Manager, Sida, October 8, 2015.

62 GovLab Interview with Carl Elmstam, Transparency Manager, Sida, October 8, 2015.

63 <http://ati.publishwhatyoufund.org/donor/sweden/>

64 “Sweden’s OGP Action Plan 2014-2016.” Open Government Partnership. <http://www.opengovpartnership.org/es/files/swedens-ogp-action-plan-2014-2016-pdf/download>

65 “Why Transparency Matters Part 5: Where Do We Go Next?” *InterAction*. January 10, 2014. <http://www.interaction.org/blog/why-transparency-matters-part-5-where-do-we-go-next>

66 GovLab interview with Karl Peterson, Project Manager, Aid Transparency, Sida, September 23, 2015.

Collaboration with Other Partners

Greater collaboration with other donors and local partnerships in recipient countries for the development and rollout of the *openaid.se* platform would be hugely beneficial and overcome some of the structural challenges the site currently faces. To that end, Sweden could work with organizations in developing countries to combine with other recipient systems for contextualized applications to enhance transparency and accountability—i.e., applications to increase citizen access to governmental aid budgets and recipients.⁶⁷

The Swedish legal barriers to reuse the data must also be addressed to create greater impacts. Reuse has the potential to enable interested citizens to create innovative new means of combining, displaying and otherwise reusing data, which has a huge potential for rethinking and reconfiguring the way aid is currently implemented.⁶⁸ Efforts at various levels must harmonize their efforts in order to avoid fragmentation.⁶⁹ In order to avoid fragmentation, though, it is necessary for actors at various levels—from grassroots to ministries—to harmonize their efforts while not cutting off creative potential.

Sweden's efforts to open aid data to the public have made major strides toward increasing transparency and accountability in the field. Given the fact that much of this effort occurred during a time when Sweden lacked the type of high-level policy commitments toward enabling the reuse of open data such as those found in many other countries is particularly notable. By focusing on the usability of the aid data hub, deploying more forward-thinking data and better collaborating with partnering stakeholders, *openaid.se*'s impacts should continue to grow and broaden in the coming years.

⁶⁷ Heacock, Rebekah and David Sasaki, "ICT4 Transparency in Sub-Saharan Africa. In *Increasing Transparency and Fighting Corruption through ICT Empowering People and Communities*." The Swedish Program for ICT in Developing Regions (SPIDER) ICT4D Series No. 3, 2010. Stockholm: Stockholm University. IRM, Sweden: Progress Report 2012-13. http://spidercenter.org/polopoly_fs/1.163640.1390315885!/menu/standard/file/Spider%20ICT4D%20series%203%20Increasing%20transparency%20and%20fighting%20corruption%20through%20ICT.pdf

⁶⁸ Dietrich, Daniel. "Open Data in Development Aid." European Public Sector Information Platform Topic Report No. 2012 / 02. March 2012. <http://www.epiplatform.eu/sites/default/files/Final%20TR%20Open%20Aid%20Data.pdf>

⁶⁹ Dietrich, Daniel. "Open Data in Development Aid." European Public Sector Information Platform Topic Report No. 2012 / 02. March 2012. <http://www.epiplatform.eu/sites/default/files/Final%20TR%20Open%20Aid%20Data.pdf>

CHAPTER 4

Open Contracting and Procurement in Slovakia

Establishing Trust in Government
Through Open Data

By Ali Clare, David Sangokoya, Stefaan Verhulst and Andrew Young

Summary: Corruption has been a long-standing problem in Slovakia, especially in the area of public procurement. In January 2011, Slovakia introduced a regime of unprecedented openness, requiring that all documents related to public procurement (including receipts and contracts) be published online, and making the validity of public contracts contingent on their publication. These reforms appear to have had a dramatic effect on corruption and, equally important for the business climate, perceptions of corruption. More generally, they are among the chief pillars of an effort—within Slovakia and the European Union at large—to reform governance and make it more responsive, transparent and efficient.

Dimension of Impact: Improving Government—Tackling Corruption and Transparency

Key Takeaways:

- Slovakia's 2011 legislation to enforce greater transparency in public procurement stands as one of "the most dramatic trans-

parency reforms in the world.”¹ It improves on an earlier Freedom of Information (FOI) law and builds on successful transparency efforts at the local level, in particular in two municipalities. The success of the national law suggests the extent to which national and regional policies can be positively influenced by small-scale, local efforts.

- Among the most important aspects of Slovakia’s law is a move away from *transparency by demand* to *transparency by default*. Earlier FOI policies required citizens to actively request information related to procurement. Under the current law, all information is required to be made open by default, making it far easier for citizens and watchdog groups to identify suspicious patterns or other indications of corruption.
- Over 2 million contracts have now been posted online in Slovakia. Because these contracts are available by default to anyone with an Internet connection (approximately 83 percent of the population), procurement and more generally corruption oversight is subject to a more “diffuse knowledge”² base. Earlier, a select group of highly motivated activists, journalists or politicians were generally involved in identifying instances of corruption.
- For all the citizenry’s newfound ability to identify corruption, Slovakia still lags when it comes to punishing corruption and enforcing accountability. This is a powerful reminder that more information and greater transparency on their own are not sufficient to change societies and political cultures. Greater openness of data and information must be accompanied by institutional reform and political motivation.
- Despite the impact of the legislation on making information on contracts public, this information must still be converted into “real” open data (e.g., made machine readable and accessible) in order to increase access and engage citizens, journalists and civil society. Efforts from civil society groups (particularly Transpar-

1 Sipos, Gabriel. “Once Riddled with Corruption, Slovakia Sets a New Standard for Transparency.” Open Society Foundations. June 2, 2015. <http://www.opensocietyfoundations.org/voices/once-riddled-corruption-slovakia-sets-new-standard-transparency>.

2 Furnas, Alexander. “Transparency Case Study: Public Procurement in the Slovak Republic.” Sunlight Foundation Blog. August 12, 2013. <http://sunlightfoundation.com/blog/2013/08/12/case-study-public-procurement-in-the-slovak-republic/>.

ency International Slovakia and Fair Play Alliance) are still emerging to standardize, scrape and add functionality to the existing government data to make it truly open. This, too, points to the importance of collaboration between the government and civil society.

Context and Background

Slovakia has long suffered from a problem with corruption. For years it was consistently ranked by Transparency International as one of the most corrupt countries in the European Union.³ According to a 2010 survey conducted by the Slovakian Business Alliance (SBA), corruption was identified by businesspeople and entrepreneurs as the number one barrier to doing business in the country.⁴ Corruption in public procurement was one of the areas identified as most problematic. Robert Kičina, executive director of the SBA, put it this way: “Public procurement didn’t have a sound name in Slovakia. Many people and enterprises think it is a very corrupt area. Maybe these suspicions make many entrepreneurs refuse to participate in tenders.”⁵

Early efforts to address corruption included a Freedom of Information Act (FOIA) adopted in 2000. According to the Transparency International Slovakia, this act “became the cornerstone of ... government transparency.”⁶ However, despite its good intentions, the act had several shortcomings. Most prominent among these was the fact that the act sought to create *transparency by demand*. Citizens

3 Sipos, Gabriel. “Once Riddled with Corruption, Slovakia Sets a New Standard for Transparency.” Open Society Foundations. June 2, 2015. <http://www.opensocietyfoundations.org/voices/once-riddled-corruption-slovakia-sets-new-standard-transparency>.

4 Kičina qtd. in Furnas, Alexander. “Transparency Case Study: Public Procurement in the Slovak Republic.” Sunlight Foundation. Sunlight Foundation Blog. August 12, 2013. <http://sunlightfoundation.com/blog/2013/08/12/case-study-public-procurement-in-the-slovak-republic/>.

5 Furnas, Alexander. “Transparency Case Study: Public Procurement in the Slovak Republic.” Sunlight Foundation Blog. August 12, 2013. <http://sunlightfoundation.com/blog/2013/08/12/case-study-public-procurement-in-the-slovak-republic/>.

6 Sipos, Gabriel, Samuel Spac, and Martin Kollarik. “Not in Force Until Published Online: What the Radical Transparency Regime of Public Contracts Achieved in Slovakia.” Transparency International Slovakia. 2015. <http://transparency.sk/wp-content/uploads/2015/05/Open-Contracts.pdf>.

could access information related to government contracts and projects, but that information was not made public by default—citizens had to proactively request it. Among other problems, this concept of transparency by demand made it difficult for public watchdogs and journalists to identify or notice conflicts of interest and other indications of corruption; in general, corruption could only be identified when individuals or groups already had an inkling of its existence and therefore actively request related information. In this respect, the 2011 reforms represented a clear paradigm shift, toward a notion of *transparency by default*.

The following list includes a narrative description of some prominent cases that highlighted the nation's corruption problem, despite the FOIA Act of 2000 and other measures. While many of these were focused on corruption in public procurement, they were in many ways symptomatic of a more general crisis of governance in Slovakia. Indeed, in late 2011 and early 2012, the country was engulfed by large-scale street demonstrations against the government. In an article on these protests from February 2012, *The Financial Times* described Slovaks as “angered” and “discontent with official cronyism and corruption.”⁷

Contract Type: Construction⁸

Scandal: In 2007 the Slovakian Ministry of Construction and Regional Development published a tender request for construction services totaling €119.5 million, by only posting a hard copy on a small notice board in the hallway inside the Ministry building. As this was the inside of the Ministry building, which is not open to the public, only enterprises with pre-existing relationships with the Ministry were able to see the tender request. A firm that was known to have close ties to Ján Slota, the head of the ruling Slovak National Party, ultimately won the contract.

Response: More than a year after the fact, the procedure came to light and was invalidated by the Slovak Office of Public Procurement.

⁷ Cienski, Jan. “Slovaks Protest over Corruption Claims” *Financial Times*. February 10, 2012. <http://www.ft.com/cms/s/0/6fc1858c-48cd-11e1-954a-00144feabdc0.html#axzz3flczWkuI>.

⁸ Furnas, Alexander. “Transparency Case Study: Public Procurement in the Slovak Republic.” Sunlight Foundation Blog. August 12, 2013. <http://sunlightfoundation.com/blog/2013/08/12/case-study-public-procurement-in-the-slovak-republic/>.

Contract Type: Legal and PR Services⁹

Scandal: Also in 2007, the Ministry of Construction and Regional Development used a notice board to obtain legal and PR services. Only one consortium consisting of companies clearly connected to one of the government parties joined the tender. As a result, the state was contracted to pay €120 million over the course of nine years. A single agency logo's price of €85,000 is especially indicative of the level of government over-payment.

Response: After the contract was signed, nearly a year passed before a media outlet discovered and reported the corruption. At that point, Fair Play Alliance worked with media partners to request and analyze not only the contract but individual payments, invoices and other relevant documents. Given the transparency-by-demand ethos of the time, watchdogs had to rely on the courts to honor such FOI requests.

In the end, public pressure stoked by media and NGOs, and an investigative report prepared by Fair Play Alliance for the European Anti-Fraud Office (OLAF) and the European Commission forced the Minister of Development to cancel the contract in 2009.

Up to €12 million was already paid but more than €100 million could be saved thanks to those efforts. The case has also been investigated by the police and will be tried in Slovak court (though no one involved has been punished as of late 2015).

Product Description and Inception

The genesis of the national legislation under examination here came in response to developments at the European and regional level. First, within Slovakia, two municipalities provided important inspiration for national transparency efforts. In 2005, a group of politicians in the southwestern town of Šaľa grew frustrated by what they perceived as a lack of transparency in the mayor's office. In response, they began publishing public contracts related to municipal business on their website. When the group gained power in city

⁹ Furnas, Alexander. "In Slovakia, Government Contracts Are Published Online." Sunlight Foundation. <http://www.opengovguide.com/country-examples/in-slovakia-government-contracts-are-published-online/>.

council elections a year later, they expanded these efforts, and Šaľa became the first municipality in Slovakia to publish all public contracts and receipts online.¹⁰ At around the same time, in the northern city of Martin, Mayor Andrej Hrnčiar, citing Šaľa as inspiration, similarly began publishing contracts and receipts online.

These efforts proved tremendously popular, and both mayors were subsequently re-elected. They provided an indication to the nation at large of what was possible through proactive transparency efforts. In late 2010, when the Slovakian justice minister spoke to Parliament as it debated the proposed new transparency rules, she paid tribute to the groundwork laid by these two municipalities: “I think it is important to remind us that we are drawing inspiration from municipalities, who showed us that having contracts, orders and receipts published online does not cause any problems; on the contrary, it raises the trustworthiness of town leadership and also ensures effectiveness and accountability when dealing with municipality resources.”¹¹

The legislation, known as Act No. 546/2010 Coll. supplementing Act No. 40/1964, came into effect on January 1, 2011, following an update to the EU Procurement Directives that included reforms such as reverse auction mechanisms for procurement and guidelines for centralized contract repositories.¹² Unlike the existing Slovak FOI law, it was focused on increasing *active* government transparency and openness. Under the new law, the government was required to publish almost all contracts, receipts and orders online, regardless of whether a citizen had made an active request for information. Importantly, government contracts were not considered

¹⁰ Mečiar qtd. in Sipos, Gabriel, Samuel Spac, and Martin Kollarik. “Not in Force Until Published Online: What the Radical Transparency Regime of Public Contracts Achieved in Slovakia.” Transparency International Slovakia. 2015. <http://transparency.sk/wp-content/uploads/2015/05/Open-Contracts.pdf>.

¹¹ Transparency International Slovakia: *Draft of anti-corruption reforms for the city of Martin* qtd. in Sipos, Gabriel, Samuel Spac, and Martin Kollarik. “Not in Force Until Published Online: What the Radical Transparency Regime of Public Contracts Achieved in Slovakia.” Transparency International Slovakia. 2015. <http://transparency.sk/wp-content/uploads/2015/05/Open-Contracts.pdf>.

¹² Furnas, Alexander. “In Slovakia, Government Contracts Are Published Online.” Sunlight Foundation. <http://www.opengovguide.com/country-examples/in-slovakia-government-contracts-are-published-online/>.

valid unless they were published within three months of being signed.¹³

“[T]his country needs the public sector and those in power [to] be checked by 5 million citizens of Slovakia.”

—Miroslav Beblavy, Slovak National Council

Inevitably, there was some pushback and opposition to the law. Some of the most vocal complaints came from mayors at the municipal level who were worried about meeting the law's various requirements, especially with their often quite limited technical and financial resources.¹⁴ Mayors were particularly concerned that the law required all receipts and orders in their municipalities to be made available on the Central Registry, a requirement which they saw as especially onerous. In January 2012, this requirement was repealed and now only metadata needs to be posted on the registry.¹⁵

Following the enactment of the law, a wide variety of documents was made available online. These were published both at the national level, at Central Register of Contracts (or CRZ: <https://www.crz.gov.sk/>) and at the municipal level. The documents published included virtually all procurement information, including judicial verdicts, contracts and receipts related to public entities like village schools, prisons, ministries and various other entities. Perhaps surprisingly, the technical and financial issues were minor and the entire process was completed within two months.¹⁶

Nonetheless, concerns do remain about the scope of documents included on the portal. For example, some 20 exemptions exist to the documents that need to be posted, and these include documents related to employment contracts, national security, family life, and unemployment benefits. In addition, in response to concerns raised

13 Sipos, Gabriel, Samuel Spac, and Martin Kollarik. “Not in Force Until Published Online: What the Radical Transparency Regime of Public Contracts Achieved in Slovakia.” Transparency International Slovakia. 2015. <http://transparency.sk/wp-content/uploads/2015/05/Open-Contracts.pdf>.

14 GovLab interview with Gabriel Sipos, Director of Transparency International, August 4, 2015.

15 GovLab interview with Gabriel Sipos, Director of Transparency International, August 4, 2015.

16 GovLab interview with Gabriel Sipos, Director of Transparency International, August 4, 2015.

by state companies about the need to protect commercial interests (for example, pricing in contract documents), the exemptions also cover certain commercial information. According to Charles Kenny, the exemptions to publication have mushroomed since the portal's launch. In addition, a study by Transparency International found that out of 1,100 state-owned enterprise and municipality contracts included in the study, almost one-fifth weren't published in their entirety.¹⁷

In addition, one major limitation of the register is that it only publishes data but does not include the tools necessary to analyze that data. In response to this shortcoming, Slovak, a consortium of civil society groups, launched a new site (<http://www.otvorenezmluvy.sk>), built on the Central Register data that was searchable and offered a range of analytical tools. Contracts are now analyzed and uploaded each night.

The screenshot shows the Central Register (CRZ) website. At the top, there is a navigation bar with links for Home, Contracts, About the site, FAQ, Contact, and User zone. A search bar is prominently displayed, with the placeholder text 'Vyhľadávanie' and a button labeled 'OK'. Below the search bar, a message indicates 'Celkovo zverejnených 879669 zmlv.' (Total number of published contracts: 879,669). The main content area features a search form with fields for 'Dodatak' (Addendum), 'Názov zmluvy' (Contract name), 'IČO dodavateľa' (Supplier IČO), 'Cena v € s DPH od:' (Price from:), 'Zverejnené od:' (Published from:), 'Rezort' (Ministry), 'Objednávateľ' (Requester), 'Číslo zmluvy' (Contract number), 'IČO odbereľaťa' (Recipient IČO), and a 'Vyhľadať' (Search) button. Below the search form, there is a section titled 'Rezorty' (Ministries) containing a list of entities:

- Agrofond – Vzdelávací fond, s.r.o.
- Akadémia umenia v Bratislavě
- Audiuzávalny fond
- Ekonomická univerzita v Bratislave
- Elektronický kontraktovací systém
- Fonds na podporu učebníkmi
- Fond na podporu vedeckej akcie
- Fond výtvarných umělci
- Generálna prokuratúra Slovenskej republiky

Figure 4-1. Screenshot of Central Register

Despite these (and other) shortcomings, Slovakia's law is now recognized as one of the most ambitious and proactive approaches to transparency legislation in existence. Gabriel Šipoš of Transparency International Slovakia has, for example, included it among "the

17 Kenny, Charles. "Learning from Slovakia's Experience of Contract Publication." Center for Global Development. May 21, 2015, <http://www.cgdev.org/blog/learning-slovakias-experience-contract-publication>.

most dramatic transparency reforms in the world.”¹⁸ Despite initial hesitations among some businessmen (concerned about publishing sensitive commercial information) and politicians (who feared a new, digital bureaucracy), support has grown since the law’s inception. As Miroslav Beblavý, a leading politician, has argued: “Unfortunately, this country needs the public sector and those in power [to] be checked by 5 million citizens of Slovakia.”¹⁹

Impact

The results of Slovakia’s transparency legislation have been striking, affecting a wide range of beneficiaries. Impact can be gauged according to four categories: engagement and use, corruption and perceptions of corruption, citizen awareness and oversight, and regional diffusion.

Intended Beneficiaries

Average Citizens

- Legal framework for more efficient and transparent use of taxpayer money
- Annual use of the platform is at 8 percent of the population, following a reported 11 percent in the first year.²⁰

18 Sipos, Gabriel. “Once Riddled with Corruption, Slovakia Sets a New Standard for Transparency.” Open Society Foundations. June 2, 2015. Accessed July 14, 2015. <http://www.opensocietyfoundations.org/voices/once-riddled-corruption-slovakia-sets-new-standard-transparency>.

19 Sipos, Gabriel, Samuel Spac, and Martin Kollarik. “Not in Force Until Published Online: What the Radical Transparency Regime of Public Contracts Achieved in Slovakia.” Transparency International Slovakia. 2015. <http://transparency.sk/wp-content/uploads/2015/05/Open-Contracts.pdf>.

20 To arrive at this number, a professional polling company was employed to undertake personal surveys with at least 1,000 people in 2012 (out of a country of 5 million) and their demographics were recorded (age, education and region). A follow up survey was undertaken early this year which showed that annual use is around 8 percent. GovLab interview with Gabriel Sipos, Director of Transparency International, August 4, tan2015.

- 90,000 “heavy users,” who have checked at least five public documents²¹

Business Community and Entrepreneurs

- Business associations (including the U.S. Chamber of Commerce in Slovakia) were largely supportive of the new legislation.²²
- Legislation is aimed at addressing Slovakia’s difficult business climate: the World Bank’s Ease of Doing Business list places Slovakia in the bottom 100 for four key indicators.²³ Prior to the law’s enactment, corruption was identified by businesspeople as the number one barrier to doing business in the country.²⁴

Media and Watchdog Groups

- Media seen as among biggest supporters and beneficiaries of new law
- 25 percent increase in stories on procurement in mainstream media
- Increase in variety and source of “tips” received by journalists as average citizens can now identify instances of possible corruption
- Overall increase in number of NGOs and scope of work undertaken by watchdog groups as a result of the legislation

21 Sipos, Gabriel, Samuel Spac, and Martin Kollarik. “Not in Force Until Published Online: What the Radical Transparency Regime of Public Contracts Achieved in Slovakia.” Transparency International Slovakia. 2015. <http://transparency.sk/wp-content/uploads/2015/05/Open-Contracts.pdf>.

22 Sipos, Gabriel, Samuel Spac, and Martin Kollarik. “Not in Force Until Published Online: What the Radical Transparency Regime of Public Contracts Achieved in Slovakia.” Transparency International Slovakia. 2015. <http://transparency.sk/wp-content/uploads/2015/05/Open-Contracts.pdf>.

23 “Doing Business in Slovak Republic.” Doing Business - World Bank Group. 2015. Accessed July 14, 2015. <http://www.doingbusiness.org/data/exploreeconomies/slovakia#dealing-with-construction-permits>.

24 Kicina qtd. in Furnas, Alexander. “Transparency Case Study: Public Procurement in the Slovak Republic.” Sunlight Foundation Blog. August 12, 2013. <http://sunlightfoundation.com/blog/2013/08/12/case-study-public-procurement-in-the-slovak-republic/>.

Engagement and Use

Several indicators point to significant use by citizens, journalists and others of both the central register (CRZ) and documents posted at the municipal level. These include:

- **Contracts Published:** Between 2011 and 2014, over 780,000 contracts were published in an open, machine-readable format on the CRZ.²⁵ A further 1.2 million were published by municipal authorities. The largest number of contracts was published by Všeobecná zdravotná poisťovňa, the main state health insurer, followed by RTVS, the national broadcaster, and Národná diaľničná spoločnosť, the state highway operator.²⁶ About one-quarter of the contracts were for amounts higher than 1,000 euros, while 3 percent were for contracts over 100,000 euros.

Citizen Use and Access of Portal: In the years preceding the passage of the new transparency legislation, fewer than 5 percent of citizens used FOI laws to request information from public agencies. According to surveys conducted by Transparency International Slovakia, 11 percent of the population accessed the portal in its first year of existence, followed by an average of 8 percent of the population in the following years.²⁷ Together with [otvorenezmluvy.sk](#), an unofficial open contracts portal run by Transparency International Slovakia and the Fair Play Alliance, the official CRZ portal attracts approximately 54,000 visits a month. Significantly, this figure has increased by a third since 2012, which suggests that citizen awareness and use of the portal data is gaining momentum.

²⁵ GovLab interview with Gabriel Sipos, Director of Transparency International, August 4, 2015.

²⁶ Sipos, Gabriel, Samuel Spac, and Martin Kollarik. “Not in Force Until Published Online: What the Radical Transparency Regime of Public Contracts Achieved in Slovakia.” Transparency International Slovakia. 2015. <http://transparency.sk/wp-content/uploads/2015/05/Open-Contracts.pdf>.

²⁷ Transparency International Slovakia. Face-to-face omnibus poll. February 3, 2015. Raw data.

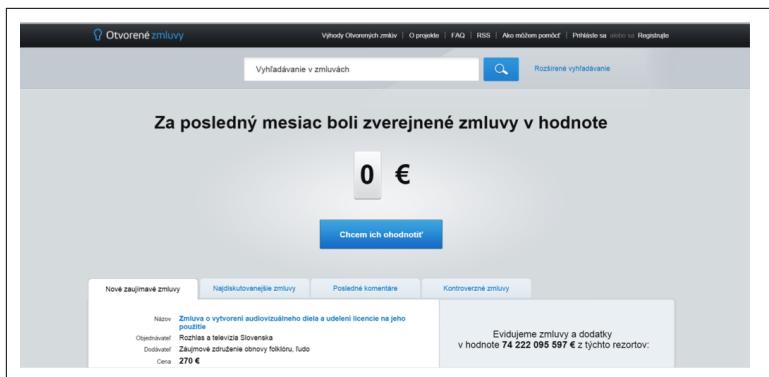


Figure 4-2. *otvorenerezmluvy.sk* main page

A fuller picture of citizen use can be determined from Google Analytics data, provided by Transparency International Slovakia.²⁸ Although the average visitor to the site spends under two minutes on the site (1m44s), 2 percent of sessions (or 170,000 visitors) spend more than 10 minutes. In addition, 20 percent of visitors are returning ones, indicating the possible existence of a committed group of transparency watchdog groups and individuals. However, Maria Zuffova, a researcher at the Slovak Governance Institute argues that the high level of returning visitors each month could actually point to a high number of individuals working in public institutions visiting the site in an official capacity, such as to upload the contracts and receipts each month.²⁹

Corruption and Perceptions of Corruption

Corruption is notoriously difficult to measure. By its very nature, it is hidden, and so challenging to capture. Nonetheless, a few indicators point to the positive impact of the law. These include:

- In the 2014 Corruption Perceptions Index published by Transparency International, Slovakia increased its ranking by six

²⁸ Sipos, Gabriel, Samuel Spac, and Martin Kollarik. "Not in Force Until Published Online: What the Radical Transparency Regime of Public Contracts Achieved in Slovakia." Transparency International Slovakia. 2015. <http://transparency.sk/wp-content/uploads/2015/05/Open-Contracts.pdf>.

²⁹ Peer review input from Maria Zuffova, Slovak Governance Institute, November 29, 2015.

places, to 54. This represented a jump of 12 places since 2011, making Slovakia one of the most improved countries over that period.

- Transparency International's index relies on *perceptions* of corruption. But some more objective indicators also suggest the law's impact. For example, while only 2 percent of tenders were conducted electronically prior to the law, almost half are now done that way. In addition to increasing transparency, this has also led to more competitive bidding—an average of one additional bidder per contract, which in turn decreases the contract price by around 2-3 percent.³⁰
- The competitiveness of bidding on public contracts can be considered a proxy for corruption. Between 2010 and 2014, the average number of bidders on public procurement contracts more than doubled in Slovakia—from 1.6 companies to 3.7 companies.
- The European Research Centre for Anti-Corruption and State Building (ERCAS), citing studies conducted by the Sunlight Foundation, concluded in 2013 that “overall, the Slovakian case supports the idea that transparency may be necessary, but not sufficient, to fight corruption effectively.” The ERCAS identifies areas of progress as a result of the new legislation, but also cites compliance and implementation (subjects we discuss below) as areas of remaining concern.³¹

Citizen Awareness and Oversight

Making available information on procurement is only a first—if important—step. The impact of Slovakia’s reforms will ultimately be determined by the extent to which citizen and civil society groups use that information to enforce accountability on their leaders. In this regard, early indications are encouraging. Media reporting on procurement has increased substantially since the passage of

³⁰ GovLab interview with Gabriel Sipos, Director of Transparency International, August 4, 2015.

³¹ “Transparency in Public Procurement Boosts Anti-Corruption Monitoring in Slovakia.” European Research Centre for Anti-Corruption and State-Building. September 2, 2013. <http://www.againstcorruption.eu/articles/transparency-in-public-procurement-boasts-anti-corruption-monitoring-in-slovakia/>.

reforms (by 25 percent over the last four years, according to one report³²), and Slovakia has also witnessed a noticeable increase in activity by watchdog groups and NGOs that work to limit corruption. As Eva Vozárová from the Fair Play Alliance notes: “Accessing contracts has become a normal source for journalists and civil society groups.”³³

The transition away from a transparency-by-demand approach to transparency by default has had a particularly powerful effect on the way citizens and watchdog groups identify instances of corruption. Previously, corruption had to be proactively identified through FOI requests, often by individuals or groups who had a certain amount of pre-existing knowledge. Under the new system, however, corruption can be more “passively” identified, for example by citizens who happen to notice apparent irregularities while perusing contracts online. This has led to a far greater number of tips, often passed on by ordinary citizens to journalists, who in turn investigate and call for accountability.

Overall, there has been a move toward what one report calls a more “diffuse knowledge”—a greater number of actors, from a greater diversity of backgrounds, is involved in policing corruption. This diffusion of oversight, away from a select group of corruption fighters to the citizenry at large, is perhaps one of the most important impacts of Slovakia’s 2011 transparency reforms.

Regional Diffusion

As with many of the case studies examined in this series, the success of this particular open data effort is also indicated by its regional impact. Slovakia’s transparency reforms have been widely hailed as models for Europe and beyond. While it is difficult to assess whether an idea is being explicitly emulated, or if an idea’s time has

32 Sipos, Gabriel, Samuel Spac, and Martin Kollarik. “Not in Force Until Published Online: What the Radical Transparency Regime of Public Contracts Achieved in Slovakia.” Transparency International Slovakia. 2015. <http://transparency.sk/wp-content/uploads/2015/05/Open-Contracts.pdf>.

33 GovLab Interview with Eva Vozárová, Web & IT Lead, Fair-Play Alliance, June 23, 2015.

34 Furnas, Alexander. “Transparency Case Study: Public Procurement in the Slovak Republic.” Sunlight Foundation Blog. August 12, 2013. <http://sunlightfoundation.com/blog/2013/08/12/case-study-public-procurement-in-the-slovak-republic/>.

simply arrived, many of the specific steps included in those reforms are occurring in neighboring states:

- In August 2011, Slovenia passed legislation making it mandatory to publish certain types of procurement contracts. According to the Slovenian Government Commission for Prevention of Corruption, this legislation was, as in Slovakia, a direct result of political frustration with corruption, particularly at the level of government contracts and tenders. Moreover, such frustration had been legitimized in 2010 by a government analysis of Public Payments Administration data that revealed high levels of corruption.

The outcome of this legislation was a portal called Supervizor, which provides information on the business transactions of public sector bodies, including legislative, judicial and executive entities; community-level agencies; public institutes and more. The project was developed by the Commission for the Prevention of Corruption of the Republic of Slovenia and partners in the Slovenian Ministry of Finance, the Public Payments Administration of the Republic of Slovenia and the Agency of the Republic of Slovenia for Public Legal Records and Related Services.³⁵ The portal at present contains data going back to 2003 (just prior to Slovenia's joining the EU) and indicates various types of information, including that on contracting parties and larger recipients of funds. Data is available in graph or printout form for specified time periods.³⁶ Its content has been described by one reporter as "a breath of fresh air."³⁷

35 "Supervizor." Supervizor - Commission for the Prevention of Corruption. 2011. Accessed July 14, 2015. <https://www.kpk-rs.si/en/project-transparency/supervizor-73>.

36 Del Monte, Davide, Ernesto Belisario, Andrea Menapace, Guido Romeo, and Lorenzo Segato. "Impact of Open Government on Public Sector Modernization Policies." Transparency International Italia. 2014. http://www.eupan.eu/files/repository/20141215142852_RomeDG_-_14_-Impact_of_Open_Government_on_PS_modernization_policies.pdf

37 Coldewey, Devin. "Slovenia Launches Supervizor, An Official Public Web App For Monitoring Public Spending." TechCrunch. August 23, 2011. <http://techcrunch.com/2011/08/23/slovenia-launches-supervizor-an-official-public-web-app-for-monitoring-public-spending/>.



Figure 4-3. Supervizor main page

- In 2015, the Czech Republic passed a similar amendment to its Act on Public Contracts, as well as a Decree on Publication of Notifications, which is believed to be based, at least in part, on the Slovak law. These laws define new guidelines for state contract tender auctions, and all signed contracts for a certain portion of responsible subcontractors must be released to the public. Contract information is now required to be published on the nation's procurement data portal, Věstník Veřejných Zakázek.³⁸ The Government of the Czech Republic has identified “the computerization of invoices in the state sector and the transparent disclosure of expenditure paid from taxpayers’ money, including small-scale contracts” as among its key priorities.³⁹

Challenges

Slovakia has travelled a great distance in its efforts to increase transparency, reduce corruption and improve governance. The reforms it has undertaken are a striking example of how more open and accessible information can have powerful societal, economic and political effects.

³⁸ “Úvodem.” Informační Systému Veřejných Zakázkách. 2013. Accessed July 14, 2015. <http://www.isvz.cz/ISVZ/Podpora/ISVZ.aspx>.

³⁹ “Policy Statement of The Government of The Czech Republic.” Government of The Czech Republic. February 14, 2014. <http://www.vlada.cz/en/media-centrum/dulezite-dokumenty/policy-statement-of-the-government-of-the-czech-republic-116171/>.

Nonetheless, the nation's transition to a more open and transparent society is still underway, the reforms themselves remain a work in progress, and much work remains to be done. If Slovakia is to continue along the path it has set itself, several challenges need to be overcome. The following three are among the most important:

Data Quality

For all the amount of information on public procurement available (over 2 million contracts have now been published), the quality of that information remains a concern. According to a survey, approximately 10 percent of published contracts in Slovakia have at least one piece of key information missing.⁴⁰ Another study found that a quarter of contracts had a “subject” missing, 12 percent were missing price-related information, and in 4 percent of contracts the name of the counterparty was redacted.⁴¹ Civil society groups in Slovakia complain that they spend an inordinate amount of time and resources cleaning data, often posting it on their own sites or unofficial portals; the public’s interests would be better served by directing those efforts at analyzing documents and identifying questionable transactions.

The public’s ability to search and analyze contracts is also limited by an absence of metadata and inter-linking between different databases (for example, databases containing tender notices and invoices and receipts related to that tender). Furthermore, amendments to contracts are often published in isolation, without linkages to the original contract, making it difficult to understand the full context or history of a particular procurement process.⁴²

⁴⁰ Sipos, Gabriel. “Once Riddled with Corruption, Slovakia Sets a New Standard for Transparency.” Open Society Foundations. June 2, 2015. Accessed July 14, 2015. <http://www.opensocietyfoundations.org/voices/once-riddled-corruption-slovakia-sets-new-standard-transparency>.

⁴¹ Sipos, Gabriel, Samuel Spac, and Martin Kollarik. “Not in Force Until Published Online: What the Radical Transparency Regime of Public Contracts Achieved in Slovakia.” Transparency International Slovakia. 2015. <http://transparency.sk/wp-content/uploads/2015/05/Open-Contracts.pdf>.

⁴² Sipos, Gabriel, Samuel Spac, and Martin Kollarik. “Not in Force Until Published Online: What the Radical Transparency Regime of Public Contracts Achieved in Slovakia.” Transparency International Slovakia. 2015. <http://transparency.sk/wp-content/uploads/2015/05/Open-Contracts.pdf>.

Civil society has played a vital role in transforming *public* data on procurement into *open* data that is machine readable and accessible for public consumption. As Eva Vozárová and many journalists involved in open data during the past decade have noted, even when government data is published, it often is not available in usable formats. In other words, the data is still not fully open: It is dispersed across various websites, not available for download, and limited in searchability or standardization (that would enable, for instance, cross-database comparisons).⁴³

The *otvorenezmluvy.sk* portal mentioned above, run by Transparency International and the Fair Play Alliance, plays a vital role in supplementing this government data. Its administrators scrape data from government registers and add functionality, including full-text search, basic analytics and the ability to red-flag possible errors or problematic contracts.⁴⁴ Transparency International Slovakia's Open Procurement Portal similarly builds on the available public procurement data and "visualizes procurement expenditures by procurers, suppliers, sectors and regions, as well as provides downloadable structured procurement data in bulk."⁴⁵ These added functionalities allow for both organizations to conduct broader analyses of government data across key variables.

Cost (and Perceptions of Cost)

When the legislation was being debated, one of the chief concerns was the potential cost of the reforms. Smaller municipalities in particular were afraid of what some termed a new "digital bureaucracy."⁴⁶ The concerns were primarily related to human costs, as the financial costs of building and maintaining the contracts portal (and

⁴³ GovLab Interview with Eva Vozárová, Web & IT Lead, Fair-Play Alliance, June 23, 2015.

⁴⁴ GovLab Interview with Eva Vozárová, Web & IT Lead, Fair-Play Alliance, June 23, 2015.

⁴⁵ Furnas, Alexander. "Transparency Case Study: Public Procurement in the Slovak Republic." Sunlight Foundation Blog. August 12, 2013. <http://sunlightfoundation.com/blog/2013/08/12/case-study-public-procurement-in-the-slovak-republic/>.

⁴⁶ Sipos, Gabriel. "Once Riddled with Corruption, Slovakia Sets a New Standard for Transparency." Open Society Foundations. June 2, 2015. Accessed July 14, 2015. <http://www.opensocietyfoundations.org/voices/once-riddled-corruption-slovakia-sets-new-standard-transparency>.

other associated sites and technology) have been negligible: a mere EUR 20,000 to set it up and an additional EUR 3,000 to maintain it, according to the estimation of one involved technician (another EUR 4,500 were spent to update the portal over the first four years).⁴⁷

In fact, according to a study conducted by Transparency International Slovakia, the human and administrative costs have not proven unduly burdensome for most municipalities.⁴⁸ Admittedly, in some cases, particularly in those regions with outdated software or hardware, it has been something of a challenge to upload and maintain records of all contracts. In addition, certain institutions (e.g., cemeteries, dormitories and organizations involved in the supply of water) have, for various reasons, found it harder to comply with the legislation. The authorities have responded in many cases by increasing the number of exemptions under the law, allowing those for whom compliance has proven particularly burdensome to avoid uploading certain records.

In some cases, exemptions have been and should continue to be granted to certain types of organizations. However, these should be granted cautiously and judiciously, since, as described above, there is concern about an excess of exemptions having been granted since the law's inception.

Enforcement and Accountability (The Human Factor)

Slovakia's law has in many respects been exemplary in the way it emphasizes transparency. Repeatedly, journalists and civil society groups have been able to use the law to shine light on cases of corruption that would probably have remained hidden before the new legislation.

⁴⁷ Sipos, Gabriel, Samuel Spac, and Martin Kollarik. "Not in Force Until Published Online: What the Radical Transparency Regime of Public Contracts Achieved in Slovakia." Transparency International Slovakia. 2015. <http://transparency.sk/wp-content/uploads/2015/05/Open-Contracts.pdf>.

⁴⁸ Sipos, Gabriel, Samuel Spac, and Martin Kollarik. "Not in Force Until Published Online: What the Radical Transparency Regime of Public Contracts Achieved in Slovakia." Transparency International Slovakia. 2015. <http://transparency.sk/wp-content/uploads/2015/05/Open-Contracts.pdf>.

Yet paradoxically, Slovakia is also illustrative of another key principle: that transparency on its own is not enough. Media organizations and watchdog groups have consistently made the point that, even when instances of corruption are brought to public attention, they often go unpunished. The strengthening of transparency has not been accompanied by a similar increase in enforcement or institutional capacity to enforce accountability. As Peter Kunder, from the Slovakian Fair Play Alliance, notes:

One lesson that we learned from publishing data ... is that it is critical and totally important to have the state publish the data but it is only one part of the success. The second part is that other institutions in the society and other aspects in the society need to work—judiciary, police and public pressure—and that is nowadays a bigger problem in Slovakia than the publishing of information.⁴⁹

Looking Forward

In order to address the above challenges and maintain momentum brought about by legislation and access to data toward meaningfully reducing corruption, Slovakia should consider a number of pathways.

Cross-Sectoral Collaboration Around Data Quality

Greater collaboration among civil society groups and government is vital to improvements in data quality and making public data more open. Ongoing quality checks through such a collaboration could help ensure that published data is complete and usable. Additionally, such collaborative portals would allow citizens to participate in scrutinizing the data and reporting errors.

Additionally, the development of automated data quality checks could help mitigate issues surrounding the data. The UK Health & Social Care Information Center, for example, uses an automated process to clean a variety of hospital episode statistics.⁵⁰

⁴⁹ Furnas, Alexander. "Transparency Case Study: Public Procurement in the Slovak Republic." Sunlight Foundation Blog. August 12, 2013. <http://sunlightfoundation.com/blog/2013/08/12/case-study-public-procurement-in-the-slovak-republic/>.

⁵⁰ HES Data Quality Team. "The HES Processing Cycle and Data Quality." Health and Social Care Information Centre, 2013. <http://www.hscic.gov.uk/article/1825/The-processing-cycle-and-HES-data-quality>

Training and Identification of Best Practices

With much of the required technical infrastructure already in place, attention should be paid to the “human factor”—training staff, ensuring they are supplied with the right equipment, and facilitating exchanges of knowledge between different municipalities to establish best practices and other guidelines for using the contracts portal.

Monitoring Compliance

To help ensure ongoing accountability, Slovakia could consider creating an agency within the government focused on monitoring compliance in contracting, perhaps in collaboration with NGOs like the Fair Play Alliance and Transparency International. Such an agency should be empowered to impose punitive (and criminal) sanctions where warranted.

The solution cannot, however, simply lie with new institutions. Existing institutions like the police and judiciary need to be modernized and trained to enforce the letter and spirit of the country’s transparency legislation. All arms of the state need to work together.

Until now, public pressure has often stood in for the state and enforced accountability when formal institutions have refused to take action. The public (including citizens, NGOs, watchdog groups and media organizations) will remain critical in ensuring accountability, and must thus receive support and encouragement. In part, this means maintaining a broader ecosystem of free expression, dissent and the right to demonstrate, a goal toward which Slovakia continues to make progress, though challenges remain—such as the wide use of defamation suits by the powerful to silence dissent and whistle-blowing.⁵¹ Robust protections for such civil liberties are a vital part of ensuring political accountability in any society.

⁵¹ “Nations in Transit 2015—Slovakia.” RefWorld. June 26, 2015. <http://www.refworld.org/topic,50ffbc582,50ffbc5be,55929ef315,0,,SVK.html>.

Slovakia has unquestionably taken significant steps toward combatting corruption through a dual-pronged approach built around legislation and technology. While the actions taken to date have shown great promise and, indeed, appear to be having an impact on corruption in the country, they should be seen as first steps along a much longer path of openness, accountability and citizen involvement.

CHAPTER 5

Indonesia's Kawal Pemilu

Elections: Free, Fair, and Open Data

By Auralice Graft, Stefaan Verhulst and Andrew Young

Summary: Kawal Pemilu (“guard the election,” in Indonesian) was launched in the immediate aftermath of the 2014 presidential elections, as the country was riven by political polarization and the two leading contenders for the presidency traded allegations of vote rigging. Faced with this delicate moment in Indonesia’s ongoing process of democratization, a globally dispersed group of technologists, activists and volunteers came together to create a website that would allow citizens to compare official vote tallies with the original tabulations from polling stations. These tabulations were already made public as part of the Elections General Commission’s (KPU) commitment to openness and transparency. However, Kawal Pemilu’s organizers played a critical role in assembling a team of over 700 volunteers to digitize the often handwritten forms and make the data more legible and accessible. The site was assembled in a mere two days, with a total budget of just \$54. Its overall impact was to enable citizen participation in monitoring the election results, increase public trust in official tallies, and more generally, help ease an important democratic transition.

Dimension of Impact: Improving Government—Tackling Corruption and Transparency

Key Takeaways:

- Successful open data projects rely and often build on a pre-existing government commitment to openness and transpar-

ency. But open data enthusiasts and activists can play a critical role in taking already-released information and making it more accessible.

- When given the opportunity, citizens are willing and able to play a role in digitizing and analyzing data in order to ensure that government is working properly. Kawal Pemilu demonstrates the importance of crowdsourcing, in its various forms.
- Successful open data projects do not require huge budgets or a centralized team. They can be built on a shoestring budget, by dispersed and nonhierarchical teams that may largely consist of volunteers.
- Open data projects that threaten the status quo must guard against hacking and other security attacks. This is an especially important consideration given the rapid and ad hoc nature of many of these projects, which naturally increases their vulnerability.

Context and Background

Corruption in Indonesia

Public demand for transparency, accountability and better government in Indonesia has been growing rapidly since the “Reformasi,” a movement for political reformation that began in the 90s.¹ In spite of this, corruption remains a problem in the country. According to Transparency International’s Corruption Perceptions Index, Indonesia ranked 107 out of 175 countries in 2014. Its Bribe Payers Index ranked Indonesia 25 out of 28 countries in 2011.² According to the World Democracy Audit, Indonesia ranked 88 out of 150 countries in terms of corruption in 2014.³

Corruption has also been evident in the electoral process, and both the 2004 and 2009 elections were characterized by alleged irregularities. Perludem (www.perludem.org), a nongovernmental association for election and democracy in Indonesia, reported instances of

1 <http://www.opengovpartnership.org/country/indonesia>

2 <https://www.transparency.org/country/#IDN>

3 <http://www.worldaudit.org/corruption.htm>

money politics, manipulation of vote counts, intimidation and officials taking advantage of government positions to add candidates or campaigns. As Diah Setiawaty, Perludem program manager for election application programming interface, puts it: “Candidates used to routinely bribe voters. They bribed them with money, with rice and with other food products.”⁴ Observers also note that Indonesia’s electoral process is particularly prone to vote manipulation, due in large part to its size and complexity.⁵ “Our election system is one of the most complex in the world,” says Setiawaty. “It has to accommodate over 100,000,000 registered voters and a complicated geography that includes over 17,000 different islands. It is huge and it all happens in one day.”⁶

Open Data and Indonesian Elections

In 2011, Indonesia joined the Open Government Partnership (OGP), emphasizing a national commitment to budget transparency and more efficient public service delivery. In 2014, Mardianto Jatna, Assistant to the Head of the Presidential Delivery Unit for Development Monitoring and Oversight (UKP4), declared a “Year of Open Data.”⁷ Several initiatives were undertaken as part of this Year of Open Data. One was the launch of the Indonesia Data Portal, which publishes approximately 700 sets of data from 23 government institutions.

Another particularly significant open data effort was the Elections General Commission (KPU)’s move to begin sharing all election-related data on the Internet. The KPU is a quasi-governmental body whose mandate is to ensure transparent and fair elections. In April 2014, amid urgent calls from citizens and political parties on all

⁴ GovLab interview with Diah Setiawaty, Program Manager for Election Application Programming Interface, Perludem, September 10, 2015.

⁵ Brajawidagda, Uuf and Akemi Takeoka Chatfield. “Roles of Social Media in Open Data Environments: A Case Study of the 2014 Indonesian Presidential Election Voting Results.” 25th Australasian Conference on Information Systems. December 8-10, 2014. http://aut.researchgateway.ac.nz/bitstream/handle/10292/8170/acis20140_submission_176.pdf?sequence=1&isAllowed=y

⁶ GovLab interview with Diah Setiawaty, Program Manager for Election Application Programming Interface, Perludem, September 10, 2015.

⁷ “2014, Indonesia’s commitment to Open Data.” *Open Government Indonesia*. March 2, 2014. <http://opengovindonesia.org/uncategorized-en/2014-indonesia-commitment-open-data/>

sides to improve transparency and safeguard the presidential elections due later that year, it announced its decision to share election data on its official website (www.kpu.go.id/).⁸

The political mood during that year was particularly polarized. It was the first time in Indonesia's history that just two candidates were running (typically, three or more run). In addition, the candidates' backgrounds and profiles were dramatically different: Probowo Subianto had a military background and was associated with the former Suharto regime, while Joko Widodo (Jokowi) belonged to a more recent generation and came from a more civilian background. According to Perludem's Diah Setiawaty, the polarization of voters was magnified by their increasing presence on social media. Indonesia had the fourth-largest number of Facebook users in the world in 2014,⁹ and Jakarta was the most active city in the world in terms of Twitter participation in 2012.¹⁰

It was against this background that the KPU made its decision to publish the voter tabulation forms from various levels of the Indonesian voting process. Voting in Indonesia is manual; citizens vote at one of more than about 470,000 polling stations. The results are then tabulated at six different levels: the polling station, the subdistrict, the district, the city regency, the province and the national. Each level uses a different form, and the time between actual voting and national tabulation can be very long. The potential for fraud—for example, by manipulating results along the way—is huge. Initially, in April 2014, the KPU announced that it would open up and make accessible the results tabulation forms for levels 2-6. However, this left out the all-important polling place tabulation form (the so-called C1 form) from the first level (i.e., the actual location of voting). After further pressure from civil society groups and some

⁸ Lukman, Enricko. "Indonesian voters are crowdsourcing ballot counts to protect against election fraud." *Tech in Asia*. July 14, 2014. <https://www.techinasia.com/kawal-suara-indonesia-voters-crowdsourcing-ballot-counts-protect-election-fraud/>

⁹ "Top 10 Countries with most Facebook users 2014." *Addictive Lists*. April 22, 2014. <http://addictivelists.com/top-10-countries-with-most-facebook-users-2014/>

¹⁰ Lipman, Victor. "The World's Most Active Twitter City? You Won't Guess It." *Forbes*. December 30, 2012. <http://www.forbes.com/sites/victorlipman/2012/12/30/the-worlds-most-active-twitter-city-you-wont-guess-it/>

political leaders,¹¹ the KPU announced in July 2014 that it would also make the C1 forms available. These forms were then scanned and published on the KPU's website, offering citizens and the media an unprecedented level of scrutiny over election results.

The KPU's initiatives have been widely praised. Ainun Najib, one of the founders of Kawal Pemilu, says: "I really have to commend the KPU for this move. It was a huge step toward improving transparency in Indonesia's election system."¹² Nonetheless, the KPU's initial efforts were in some respects incomplete, and have since been supplemented by citizen efforts. For example, the scanned C1 forms were filled out by hand and published in a non-machine-readable format (usually JPEG or PDF); often, these forms were hard to decipher, and sometimes published upside down.

To address these (and other) problems, a number of independent, citizen-launched applications and websites emerged around and since the 2014 presidential elections. The first such application appeared almost immediately after the July 11 election, when a Twitter user with 7,000 followers began posting the scanned C1 forms.¹³ Other applications followed, each monitoring different portions of KPU-opened data: Some filled the gap in machine-readability by digitizing the scanned C1 forms and making them more accessible; other applications dealt with district, city/regency and province-level scanned forms, monitoring results at those voting levels; and yet another category of citizen-led applications examined scanned C1 forms for problems or anomalies. Kawal Pemilu, the application under study here, was among the most successful initiatives in this surge of citizen-led voting applications.

¹¹ GovLab interview with Diah Setiawaty, Program Manager for Election Application Programming Interface, Perludem, September 10, 2015. And Lukman, Enricko. "Indonesian voters are crowdsourcing ballot counts to protect against election fraud." *Tech in Asia*. July 14, 2014. <https://www.techinasia.com/kawal-suara-indonesia-voters-crowdsourcing-ballot-counts-protect-election-fraud/>

¹² GovLab interview with Ainun Najib, Co-Founder Kawal Pemilu, August 5, 2015.

¹³ Rakhmani, Inaya. "Technology for Transparency." *New Mandala*. July 21, 2014. <http://asiapacific.anu.edu.au/newmandala/2014/07/21/technology-for-transparency/>

Project Description and Inception

During the summer of 2014, Ainun Najib, who would later go on to co-found Kawal Pemilu, was living and working in Singapore. It was the middle of Ramadan, and he was down with the flu. Still, he was following the elections closely in his native Indonesia, and watched with growing concern as the extreme polarization that had characterized the campaign spilled over into the post-election period, and as the results were contested by both the Jokowi and Probowo camps.

Amid allegations of fraud and vote rigging, Najib began looking at ways he could add more transparency to the electoral process. “We [had to do something] to prevent the nation [from] being ripped apart because of two claims to victory that nobody can verify,” he told *The Financial Times*.¹⁴ Around this time, he messaged a friend, Andrian Kurniady, a Google employee based in Sydney; Najib had met Kurniady in 2007 at a Mathematics Olympiad Champions competition, and though they were not close friends, they were connected on Facebook. The two quickly decided to work together to help safeguard Indonesia’s vote count. Later, they were joined by three other friends, who worked with them to get the application off the ground.

In its early days, Kawal Pemilu tried to fill machine readability gaps in KPU data by digitizing scanned C1 forms and using handwriting recognition to extract the polling data. They quickly ran into obstacles with this approach, however, and soon decided to turn to a form of crowdsourcing. Specifically, they decided to recruit volunteers to manually digitize the almost 500,000 scanned C1 forms on the KPU site. This was the genesis of Kawal Pemilu (“guard the election,” in Indonesian), which was launched on July 12, 2014, with the intent of providing a platform for public participation in guarding the 2014 general election results. As Andrian Kurniady, Najib’s co-founder, put it at the time: “We hope this system can reduce the uncertainty, fear of election fraud, and restore public confidence in one of the

¹⁴ Bland, Ben. “Indonesian techies crowdsource election results.” *Financial Times*. July 18, 2014. <http://www.ft.com/intl/cms/s/0/6c62a8b6-0e33-11e4-85ab-0014feabdc0.html#axzz3x2yImI1u>.

most important points in Indonesian democracy [while it] is still young.”¹⁵

The Kawal Pemilu application consists of two main components. At the back-end, it includes a private, walled-off website, where volunteers and site administrators can enter polling data based on scanned forms ([Figure 5-1](#)). In addition, there is a public-facing site that allows citizens to view the data, broken down by polling station and candidate ([Figure 5-2](#)). Visitors can choose to view results for different levels of the tabulation process. For example, a visitor can choose to examine just C1 results. Results can also be viewed by region.

The underlying data for the site was generated from KPU data through a network of volunteers spread across the globe. Volunteers were recruited through a closed, secret Facebook group, which ensured that only trusted people were included. To start the process of recruiting volunteers, each founder chose 10 trustworthy friends, each of whom was asked to recruit 10 more, and each of those was also asked to recruit another 10 friends—and so on. Over 700 volunteers were recruited in this way in just three days. Volunteer names and identities were initially kept hidden, to prevent any efforts at bribing or intimidation.

Each recruit received a secret link to the nonpublic component of the website, where scanned C1 forms were presented with an accompanying form for the volunteer to fill in with extracted data.¹⁶ The form also allowed for the reporting of errors. The results of this work were posted every 10 minutes to the public-facing website, which was read-only. In addition to allowing citizens to monitor election results in near real-time, this data also allowed them to compare vote tallies listed on the site to official numbers released by polling stations.

¹⁵ Yusuf, Oik. “Siapakah ‘Pandawa’ di Balik Kawal Pemilu?” *Kompas.com*. August 13, 2014. http://teknokompas.com/read/2014/08/13/09050057/Siapakah.Pandawa.di.Balik.Kawal.Pemilu.?utm_source=WP&utm_medium=box&utm_campaign=Khlpw. Translated into English by Google.

¹⁶ GovLab interview with Ainun Najib, Co-Founder Kawal Pemilu, August 5, 2015.

Join us together we Escort 2014 election!
 Visit our [Facebook Page](#) [Guard 2014 election](#) and help us report the problem to the data! [Tweet](#) [33.7k]

Open this page using a browser (browser) modern like [Firefox](#) or [Chrome](#) latest version.
 And do not forget, read the directions for use below. C1 detailed data are available at [here](#).

C1 DAI DB1 DC1 Snippet things. 4 Snippet things. 1 Snippet things. 2 Snippet things. 3 Snippet things. 5
[National](#) »BALI» [GIANYAR](#) »Ubud» [Ubud](#)

TPS	C1 full scan	Page 4	The value of crowdsourcing
# 1	If the snippet on the right one, you can see the other pages of the website Commission Page 1 Page 2 Page 3 Page 4 Page 5		Prabowo: 87 Jokowi: 349 Valid votes: 436 Invalid: 2 Report an error
	If the snippet on the right one, you can see the other names		Prabowo: 98

Figure 5-1. Screen used by volunteers to digitize scanned C1 forms (translated into English by Google)

Bergabung bersama kamu bersama-sama kita Kawal Pemilu 2014! Kunjungi Facebook Page Kawal Pemilu 2014 dan bantu kamu melaporkan masalah pada data! Tweet [33.7k]											
Bukalah halaman ini menggunakan peramban (browser) modern seperti Firefox atau Chrome versi terkini. Dua jangka lama, bisa pertengah pokok di bawahi. Data C1 terperinci tersedia di ini .											
<input checked="" type="checkbox"/> C1 <input type="checkbox"/> DAI <input type="checkbox"/> DB1 <input checked="" type="checkbox"/> DC1 <input checked="" type="checkbox"/> Snippet hal. 4 <input type="checkbox"/> Snippet hal. 1 <input type="checkbox"/> Snippet hal. 2 <input type="checkbox"/> Snippet hal. 3 <input type="checkbox"/> Snippet hal. 5											
Naional											
No	Tempat	Prabowo-Hatta (C1)	Prabowo-Hatta (DC1)	Jokowi-JK (C1)	Jokowi-JK (DC1)	Susra Sab	Tidak Sab	TPS Error	TPS		
	TOTAL	61,677,150 46,99%	62,576,585 46,85%	69,353,222 53,01%	70,997,607 53,13%	131,162,864 1,364,151	1,199 99,76%				
1	ACEH	1,026,671 55,31%	1,452,413 55,31%	833,002 44,79%	911,309 44,79%	1,859,661 1,859,661	55,962 55,962	1,92%	100,00%		
2	SUMATERA UTARA	2,814,953 44,73%	2,811,514 44,73%	3,478,032 55,27%	3,494,835 55,24%	6,292,584 6,292,584	31,041 31,041	0,61%	100,00%		
3	SUMATERA BARAT	1,791,495 76,97%	1,797,505 76,92%	335,979 23,05%	525,320 23,05%	2,327,474 2,327,474	17,689 17,689	0,54%	100,00%		
4	RIAU	1,341,459 50,35%	1,349,338 50,32%	1,333,426 49,05%	1,342,817 49,05%	2,677,694 2,677,694	18,545 18,545	0,75%	100,00%		
5	MIMBARU	689,264 49,27%	671,216 49,25%	695,056 50,25%	697,787 50,25%	1,764,320 1,764,320	13,891 13,891	0,73%	100,00%		
6	SUMATERA SELATAN	2,115,027 51,27%	2,019,376 51,02%	2,019,376 49,73%	2,029,202 49,73%	4,144,938 4,144,938	31,770 31,770	0,53%	100,00%		
7	BENGKULU	433,063 45,27%	432,173 45,27%	523,611 54,73%	521,660 54,73%	956,678 956,678	7,811 7,811	0,12%	100,00%		
8	LAMPUNG	2,026,393 46,65%	2,013,924 46,93%	2,291,229 53,07%	2,298,889 53,07%	4,317,637 4,317,637	29,967 29,967	0,73%	100,00%		
9	KEPULAUAN RANGKAPOL BESITING	200,315 32,73%	200,706 32,74%	411,752 67,27%	412,359 67,26%	613,143 613,143	6,110 6,110	0,23%	100,00%		
10	KEPULAUAN RIAU	330,653 40,35%	332,908 40,37%	488,572 59,64%	491,819 59,63%	815,225 815,225	5,614 5,614	0,64%	100,00%		
11	DKI JAKARTA	2,521,412 46,92%	2,528,054 46,92%	2,052,843 53,08%	2,059,894 53,08%	5,377,828 5,377,828	33,834 33,834	0,68%	100,00%		
12	JAWA BARAT	14,072,239 59,77%	14,051,940 59,61%	9,471,918 40,23%	9,533,750 40,23%	23,541,700 23,541,700	296,185 296,185	1,02%	99,42%		

Figure 5-2. Screenshot of the public-facing Kawal Pemilu website (translated into English by Google)

The entire process of building the site and gathering all the data was performed with remarkable efficiency. Because the co-founders were spread around the world (California, Sydney, Singapore, Indonesia, Netherlands, Germany), they were able to work round-the-clock, taking maximum advantage of different time zones. This enabled them to create the website and counting system in just two days. In addition, all the recruited volunteers and founders worked without pay. The total capital investment for the setup was, as a result, just \$54; these funds were used to purchase the website's domain and

space on a hosting server.¹⁷ Overall, Kawal Pemilu is an impressive example of a “public-service startup”: Despite never benefiting from seed capital or a Silicon Valley-style business ecosystem, it was put together with all the agility and speed that are the hallmarks of far more lavishly funded business ventures.

Impact

Kawal Pemilu was one of many crowdsourced vote count initiatives launched around the time of the 2014 elections, and built on KPU data. Others included Kawal Suara (“guard the votes”),¹⁸ Real Count and a Tumblr site called C1 Yang Aneh. However, Kawal Pemilu stands out as one of the most efficiently built and managed, and one of the most effective initiatives. In an article on “Indonesia’s election tech fighters,” one Indonesian reporter described the site as “the most professionally run” among various efforts.¹⁹ Kawal Pemilu has also been described as a “leading spearhead in monitoring and legitimizing the [2014 electoral] vote count outside the state apparatus.”²⁰

The impact of Kawal Pemilu can be measured in several ways:

Settling the 2014 Election Results

By the fifth day after the 2014 elections (four days after the site’s launch), Kawal Pemilu volunteers had digitized 470,000 or 97 percent of all scanned C1 forms.²¹ In fact, volunteers had competed against each other to scan the highest number of forms, spurring the

17 “The Five Pandawa Behind Kawal Pemilu.” *Global Indonesian Voices*. August 13, 2014. <http://www.globalindonesianvoices.com/14988/the-five-pandawa-behind-kawal-pemilu/>

18 Lukman, Enricko. “Indonesian voters are crowdsourcing ballot counts to protect against election fraud.” *Tech in Asia*. July 14, 2014. <https://www.techinasia.com/kawal-suara-indonesia-voters-crowdsourcing-ballot-counts-protect-election-fraud/>

19 Lukman, Enricko. “5 things you need to know about Indonesia’s election tech fighters.” *Tech in Asia*. July 18, 2014. <https://www.techinasia.com/kawal-pemilu-indonesia-presidential-election-fighters/>

20 <http://theconversation.com/indonesias-democratic-masses-brought-victory-to-jokowi-29569>

21 Lukman, Enricko. “5 things you need to know about Indonesia’s election tech fighters.” *Tech in Asia*. July 18, 2014. <https://www.techinasia.com/kawal-pemilu-indonesia-presidential-election-fighters/>

rapid and efficient process described above.²² Based on these numbers, Kawal Pemilu (and citizens accessing the site) was able to establish that the vote count was in fact very similar to the official KPU result (53.15 percent for Joko Widodo and 46.85 percent for Prabowo Subianto), which was at the time being contested by the Prabowo camp.

A month after the elections, the vote tallies were still being contested, and the results were taken to court. Kawal Pemilu played an important role in the court hearings, and its testimony—along with that provided by KPU officials and other expert witnesses—helped influence the court’s decision to award the elections to Jokowi. In this way, the crowdsourced election results helped settle the elections, legitimize the winner and, more generally, ensure a peaceful transfer of power in Indonesia.

Increased Trust and Greater Public Participation

Beyond its direct impact on the 2014 elections, Kawal Pemilu has also had an overall effect on Indonesia’s electoral process and political climate, helping to increase transparency and public trust. As Uuf Brajawidagda, a lecturer of Indonesian politics at the University of Wollongong in Australia, who studied Kawal Pemilu and similar Indonesian open data initiatives, puts it: “Projects like Kawal Pemilu catalyzed the level of trust among citizens. The KPU move [to open election related data] increased trust, but Kawal Pemilu and others like it transformed the trust into something bigger. Their actions amplified the level of trust.”²³

The greater public trust has, in many ways, translated into a new sense of citizen involvement and stakeholdership in Indonesia’s political process. The sense of involvement is most evident in the more than 700 volunteers who helped crowdsource the KPU data. But observers also note a more general sense of empowerment and

²² After 460,000 forms, recruitment was stopped because Kawal Pemilu founders believed they had achieved their goals. In addition, the site had received considerable media coverage by this time, and the founders decided to stop recruiting because they feared that recruits might not be trustworthy anymore.

GovLab interview with Diah Setiawaty, Program Manager for Election Application Programming Interface, Perludem, September 10, 2015.

²³ GovLab interview with Uuf Brajawidagda, Lecturer of Indonesian Politics, University of Wollongong, Australia, September 3, 2015.

new expectations of transparency. As Perludem program manager for election application programming interface at Perludem Diah Setiawaty, put it: “Kawal Pemilu created a movement—a movement toward increasing open data and transparency in Indonesia. It set the bar really high.”²⁴ As evidence, Setiawaty cited the numerous open data initiatives that have followed Kawal Pemilu (see below for details), and the energy she experiences around these initiatives. “Indonesians are definitely more eager and willing to participate in crowdsourced applications now,” Setiawaty says. “They want to participate in the political process.”

Reduced Polarization

Kawal Pemilu emerged at a time of intense polarization in Indonesia. One of its most important (and perhaps long-term) effects may be to reduce the climate of partisanship and political division. Observers note that, because it presented an independent, verifiable and nonpartisan snapshot of the election results, it helped heal some of the mistrust and mutual suspicion among rival political camps. “I think it closed the gap between the two groups,” said Brajawidagda. “It made people feel better about official results, even if those results contradicted their personal outcome preferences.”²⁵ As evidence, Brajawidagda cites some of the comments he saw posted on social media—comments that point to people’s faith in the final results, regardless of their political leanings. One example he cites is the following comment: “Unbelievable, kawalpemilu (final count) deviates only 0.01 percent. Great. ... This is the real count.” Overall, Brajawidagda feels that, in helping validate the election results, Kawal Pemilu played an important role in fostering a more civil political climate in Indonesia.

Perludem’s Setiawaty agrees. “Politics were still very polarized after Kawal Pemilu,” she says. “But the mood was less negative. People could see things from a more objective perspective.”²⁶

²⁴ GovLab interview with Uuf Brajawidagda, Lecturer of Indonesian Politics, University of Wollongong, Australia, September 3, 2015.

²⁵ GovLab interview with Uuf Brajawidagda, Lecturer of Indonesian Politics, University of Wollongong, Australia, September 3, 2015.

²⁶ GovLab interview with Diah Setiawaty, Program Manager for Election Application Programming Interface, Perludem, September 10, 2015.

Challenges

What happens next for Kawal Pemilu and, more generally, for Indonesia's political climate? Despite the early successes of initiatives like Kawal Pemilu, corruption remains a serious problem in Indonesia, and ample scope exists for improvement in the electoral process. In order to truly change conditions, Kawal Pemilu (and other similar projects) will not only need to scale significantly but also prove that they can grow and last beyond a single election. If that is to happen, a few challenges will need to be overcome.

Legal Changes

In fact, the KPU's move to open up election data in 2014 was just a temporary measure, without any lasting legal foundations. Following the success of Kawal Pemilu and similar applications, there has been a general clamor calling for legislation to require opening of election data for all future elections. The organization Perludem has played a key role in pushing for new legislation and in drafting versions of the proposed law. Setiawaty of Perludem says: "Kawal Pemilu was a definite success, and in addition it has increased the probability of new election laws in Indonesia that emphasize transparency." So the prospects for new, more-encompassing legislation are bright—but until it is passed, the cause of greater transparency in Indonesian politics, and more specifically applications like Kawal Pemilu, will continue to face challenges.

Security

As with any anticorruption project, Kawal Pemilu has come up against some powerful vested interests. Soon after it was launched, some of those vested interests appeared to strike back. Four days into its existence, the site was attacked by what administrators called "hundreds of hackers."²⁷ As a consequence, the site was brought down for a few hours, until the project's administrators returned a "logic bomb" to their attackers. In addition, administrators ran a

²⁷ Lukman, Enricko. "As Indonesia's democracy is on the verge of crisis, hackers and fakers attack crowdsourced vote counts." *Tech in Asia*. July 18, 2014. <https://www.techinasiacom/indonesia-democracy-crisis-hackers-fakers-attack/>

mirrored version of the public site, to limit the potential damage from hacking.

The identity of the hackers remains unknown, but they were mostly believed to be located within Indonesia, and they were widely considered to be responding to the website's growing legitimacy and popularity. In this sense, the hacking attempts were just a particular instance of a more general effort to hijack the project or weaken its credibility.²⁸ For example, in the early days of the Kawal Pemilu, the project's founders went to great lengths to hide their identities (and those of their volunteers), in an effort to thwart attempts at bribery or intimidation. Such threats are only likely to grow as Kawal Pemilu and other popular initiatives seek to take on the establishment and increase transparency in Indonesia.

Reliability of Information

In its early days, Kawal Pemilu experienced some skepticism about the reliability of its data. This skepticism was heightened by the anonymity of its volunteers—ironically, because the anonymity was in fact designed to decrease their susceptibility to untoward influence and was thus a way of ensuring reliability. Ainun Najib says organizers regularly fielded questions from the public about accountability and verification of inputs. He and the rest of the team always took the time to explain their methodology, and to point out that all the data was verifiable within the site itself. Indeed, visitors to the site can click on any piece of information and find the underlying, original scanned form, all the way down to the C1 form from the initial polling station, as proof. In that sense, Kawal Pemilu contains an internal process of legitimization—which helps overcome (if not entirely do away with) public skepticism or doubts.²⁹

Lasting Political Impact

Any issue- or event-focused project faces questions about its survivability and sustainability over time. For Kawal Pemilu, the question is not only whether the project can survive, but what form it can take in the long term. The same factors that permit such initiatives

28 Rakhmani, Inaya. "Technology for Transparency." *New Mandala*. July 21, 2014. <http://asiapacific.anu.edu.au/newmandala/2014/07/21/technology-for-transparency/>

29 GovLab interview with Ainun Najib, Co-Founder Kawal Pemilu, August 5, 2015.

to emerge quickly and rapidly—their decentralized, ad hoc nature—often make it difficult to grow into larger organizations. The founders of Kawal Pemilu are well aware of these challenges. They say they don't have answers yet, but are exploring strategies to ensure the long-term sustainability of Kawal Pemilu.³⁰

Looking Forward

The example offered by Kawal Pemilu has inspired several other open data initiatives that aim to increase transparency and reduce corruption in Indonesia. In some ways, we are witnessing the emergence of a new, citizen-driven open data ecosystem in the country, one whose aims can broadly be defined as improving governance and enhancing democracy.

This ecosystem is likely to grow more vibrant in coming years, and several new applications and websites are already starting to make their presence felt. Here, we highlight three of the most promising.

Kawal APBD

Following his involvement with open data and elections, Kawal Pemilu co-founder Ainun Najib has launched a new site that aims to improve budget transparency and open it up to citizen involvement. His new project, Kawal APBD (www.kawalapbd.org) places different versions of government budgets side by side and allows citizens to find and comment on discrepancies. For example, in one case, the site helped find a discrepancy between the South Jakarta department of education and the education section of the city budget.³¹ In addition to commenting on discrepancies, citizens can also “like” or “dislike” budget allocations, and visualize data, thus helping them better understand the budget process.

The goal of Kawal APBD “is to make more government data accessible and presentable for people,” said Mr. Najib. “We already knew

³⁰ Rakhmani, Inaya. “Technology for Transparency.” *New Mandala*. July 21, 2014. <http://asiapacific.anu.edu.au/newmandala/2014/07/21/technology-for-transparency/>

³¹ “Lihat sendiri perbedaan APDB versi Pemprov dan DPRD di KawalAPBD.org.” *Rappler*. March 10, 2015. <http://www.rappler.com/world/regions/asia-pacific/indonesia/86371-lihat-perbedaan-apbd-versi-pemprov-dan-dprd&prev=search>. Translated by Google.

through Kawal Pemilu that people would participate. We hope more and more government data is going to be exposed to the public.”³²

Mata Massa (“Eye of the Masses”)

Mata Massa was in fact launched before Kawal Pemilu, in the run-up to the 2014 elections. It allows citizens to report campaign-related and other violations through their smartphones to the General Elections Monitoring Body (Bawaslu). For example, citizens have reported instances of vote buying, as well as other administrative violations. Overall, some 1,500 violations have been reported (though reports indicate that little follow-up action has in fact been taken).³³ Originally launched to monitor the trustworthiness of journalists, the app was created and is run by the Alliance of Independent Journalists in Indonesia.³⁴

Kawal Pilkada

Kawal Pilkada sought to replicate Kawal Pemilu’s effort to crowd-source digitizing of KPU C1 forms for regional elections in December 2015.³⁵ The project is built around the same concept as Kawal Pemilu, but under different leadership. There are several challenges here. One is in terms of recruitment. While citizens are eager to participate, crowdsourcing is more challenging than it was in the Kawal Pemilu experience. Diah Setiawaty, whose team provides technical assistance to the Kawal Pilkada team, says there is less intrigue overall in these regional elections, and voters know and care less about leaders in other regions. Kawal Pilkada also faces a time constraint. The KPU only recently announced it would publish related tabulations online, so the Kawal Pilkada team has little time to prepare.

32 <http://blogs.wsj.com/indonesiarealtime/2015/03/11/ouch-jakartans-compare-dueling-spending-plans/>

33 Rakhmani, Inaya. “Technology for Transparency.” *New Mandala*. July 21, 2014. <http://asiapacific.anu.edu.au/newmandala/2014/07/21/technology-for-transparency/>

34 Naing, Saw Yan. “Smartphone-Wielding Indonesians Tasked With Ensuring Fair Elections.” *The Irrawaddy*. April 7, 2014. <http://www.irrawaddy.org/asia/smartphone-wielding-indonesians-tasked-ensuring-fair-elections.html>

35 Ruslan, Heri. “Kawal Pilkada Publikasikan Real Count Pilkada Serentak.” *Republika.co.id*. December 20, 2015. <http://nasional.republika.co.id/berita/nasional/pilkada/15/12/20/nzn40m372-kawal-pilkada-publikasikan-real-count-pilkada-serentak>. Translated by Google.

Another constraint going forward is the frequency of regional elections, which occur every two years. “For each election, there is a lot to monitor at the same time. The political situation in every region is different. We face different challenges in every place,” says Setiawaty.³⁶

Taken together, these and many other existing and emerging applications make up a new landscape for open data, transparency and accountability in Indonesia. They demonstrate that data can be used by ordinary citizens to hold government accountable, and, although many of them remain modest, more proof of concept than full-blown national platforms, they suggest that a citizenry empowered with information can indeed effect genuine political transformation.

³⁶ GovLab interview with Diah Setiawaty, Program Manager for Election Application Programming Interface, Perludem, September 10, 2015.

CHAPTER 6

Denmark's Open Address Data Set

Consolidating and Freeing-up Address Data

By Juliet McMurren, Stefaan Verhulst and Andrew Young

Summary: A broad variety of services depend on accurate, up-to-date address data, including emergency services; the police; postal and transport services; and GPS systems. In 2005, the Building and Dwelling Register of Denmark started to release its address data to the public free of charge. Prior to that date, each municipality charged a separate fee for access, rendering the data practically inaccessible. Other entities, such as the Cadastre, the Land Registry, the Central Business Register and the utilities sector, collected their own address information, leading to significant discrepancies between the content held across different databases. A follow-up study commissioned by the Danish government estimated the direct financial benefits alone for the period 2005-2009 at EUR 62 million, at a cost of only EUR 2 million.

Dimension of Impact: Improving Government—Improving Services

Key Takeaways:

- In highly open, transparent societies, the opening of government data can be successfully justified and driven by expected economic benefits from increased efficiency and innovation within government, without the need for political arguments for transparency.

- Relatedly, efforts to collect and open data can uncover major gaps and deficiencies in official government data sources, creating benefits to internal efficiency and recordkeeping, often with little to do with improving transparency or accountability.
- There is a need to establish robust and carefully planned governance models for data collection systems, and skills for data collection among those charged with collecting it, to ensure the resulting data set is clean, accurate, usable and widely used.

Context and Background

Denmark is a high-income¹ constitutional monarchy in northern Europe with a population of 5.6 million in 2015.² It is ranked 10th on the United Nations Human Development Index.³ Locally, the country is divided into five regions, which are further subdivided into 98 municipalities, reduced from 270 in 2007.

Denmark is considered, both nationally and internationally, among the most highly open and free countries in the world. The country was ranked second on the 2014 Global Open Data Index by the Open Knowledge Foundation,⁴ ninth on the 2015 Open Data Barometer⁵ and fourth on the World Justice Project's Open Government Index.⁶ The Danish media is considered free by Freedom House,⁷ and the country is ranked third in the World Press Freedom Index by Reporters Without Borders.⁸ Consequently, corruption is not widely considered a problem in Denmark.⁹

Denmark is a signatory to the Open Government Partnership. It began the process to join in September 2011, and, as of late 2015, is implementing its first Action Plan and developing its second, which

¹ <http://data.worldbank.org/country/denmark>

² <http://denmark.dk/en/quick-facts/facts/>

³ http://hdr.undp.org/sites/all/themes/hdr_theme/country-notes/DNK.pdf

⁴ <http://index.okfn.org/place/denmark/>

⁵ <http://www.opendatabarometer.org/report/analysis/rankings.html>

⁶ <http://worldjusticeproject.org/open-government-index/open-government-around-world>

⁷ <https://freedomhouse.org/report/freedom-press/freedom-press-2015#.Vf29Nt9Viko>

⁸ <http://index.rsf.org/#/>

⁹ http://transparency.dk/?page_id=1258

has open data as one of its four themes. The Action Plan expresses the government's intention to develop an Open Data Innovation Strategy (ODIS), to draw attention to the potential of public data and make it accessible. ODIS will also operate the public data catalogue and facilitate communication between public sector and private reusers of data.¹⁰

Despite its wealth, Denmark has found itself under increasing pressure to find ways to control government spending. Danish citizens are among the most highly taxed in Europe,¹¹ but there is little public appetite for government program cuts.¹² Nonetheless, according to OECD data from 2011, the Danish government spent over a quarter of its GDP (26.1 percent) on social spending, compared to an average of 21.7 percent across OECD member countries.¹³ The OECD also found that 26.8 percent of the labor force was employed in the government sector (excluding public corporations) in 2012, and compensation for these employees consumed 19 percent of GDP, the highest percentage in the OECD.¹⁴ Faced with an aging population and fewer taxpayers, a real prospect of government spending reaching unsustainable levels, and a population reluctant to embrace budget cuts, the Danish government has had no option but to look for ways to improve government efficiency.¹⁵

10 The Danish Government, "Open Government National Action Plan 2013-2014." October 2013. http://www.opengovpartnership.org/sites/default/files/Denmark_Open%20Government%20Action%20Plan%202013-2014_ENG_1-sided_print.pdf, pp 14-15; "Open Data Innovation Strategy (ODIS)." Digitaliseringsstyrelsen. April 22, 2013. <http://www.digst.dk/Servicemenu/English/Policy-and-Strategy/Open-Data-Innovation-Strategy-ODIS>

11 "Tax reforms in EU Member States 2012: tax policy challenges for economic growth and fiscal stability." Taxation Papers 34 and European Economy 6. http://ec.europa.eu/europe2020/pdf/themes/02_taxonomy.pdf

12 GovLab interview with Thorhildur Jetzek, Department of IT Management, Copenhagen Business School, September 21, 2015.

13 <http://www.oecd.org/social/expenditure.htm>

14 "Human Resources Management Country Profiles: Denmark." OECD. December 6, 2012. <http://www.oecd.org/gov/pem/OECD%20HRM%20Profile%20-%20Denmark.pdf>

15 Jetzek, Thorhildur. "Managing complexity across multiple dimensions of open data: the case of the Danish Basic Data Program." *Government Information Quarterly*. November 24, 2015. <http://www.sciencedirect.com/science/article/pii/S0740624X15300186>

Project Description and Inception

Before 1996, address data was registered and collected individually by the 270 Danish municipalities. Although public data was available, organizations wanting to access the data had to make separate access and pricing agreements with each municipality, rendering the data practically inaccessible. The lack of an accessible, unified public data set resulted in the development of several private databases of varying quality.¹⁶ Furthermore, there was no consensus on the definition of an address, or who (if anyone) held the reference address file, with users such as Post Danmark and the emergency services each holding rival data sets.¹⁷

Appreciating that there would be considerable practical benefit to a single standardized address system, with each address datum attached to a unique geographic point, officials at the Danish National Survey and Cadastre (now the Danish Geodata Agency) began a pilot program coordinating this already-collected data into a single body. Morten Lind, business manager of the Danish Address Program at the Danish Geodata Agency, describes the origins of the program: “We started as a pilot project in 1996, where we said, there’s a potential in having good address data, and in taking addresses and giving them a geographic coordinate so that we know where each address is. If every government agency has its own separate address database, we are not sure that we can actually compare things. So we had this idea of making a common reference data set where we could say: These are the addresses that exist. [And] it was clear that the only party that could create this reference file of addresses was the local authorities.”¹⁸

In order to do this, the Danish National Survey and Cadastre decided to adapt the existing building and dwelling register, with minor adjustments, as the primary address register. The local authorities

¹⁶ Lind, Morten. “The value of Danish address data: social benefits from the 2002 agreement on procuring address data etc. free of charge.” Danish Enterprise and Construction Authority, July 7, 2010. http://www.adresse-info.dk/Portals/2/Benefit/Value_Assessment_Danish_Address_Data_UK_2010-07-07b.pdf

¹⁷ GovLab interview with Morten Lind, Senior Advisor, Danish Geodata Agency, September 15.

¹⁸ GovLab interview with Morten Lind, Senior Advisor, Danish Geodata Agency, September 15.

would collect the data, assign and register addresses and street names, and would have responsibility for and ownership of the data, while the Danish Survey and Cadastre would coordinate the project.

¹⁹ At this point, the submission of data was voluntary, and municipalities could not be compelled to supply address data, but some financial support was offered to help municipalities with compliance.²⁰ Lind and a colleague from the Danish organization of local authorities also met with local authorities throughout the country to promote the project.²¹

The potential interest and utility of the data set quickly became apparent to those working with it. A project manager at the Ministry of Housing, Urban and Rural Affairs commented: “It did not take many years after starting to work with the public data in around 1993 to realize that these data could be used for many interesting things. And we felt it was a shame ... that the data weren’t used more.”²²

By 2001, the data had been collected in a single register owned by the local authorities. Potential users such as the police, emergency services, public health agencies, public transportation companies and agencies like the Danish Ministry of the Environment and county administration were aware of and eager to use the data to provide their services armed with a more accurate understanding of the locations of homes, businesses and other entities, but, as Lind explained, the next hurdle was finding an agreement on how it could be accessed and used. “The police want[ed] to use it, the Ministry of the Environment want[ed] to use it, but ... we had almost 300 local authorities that could decide their own pricing models on a completely individual basis. We couldn’t foresee that an agreement could be reached [on] a price [for] address data. So we were in a

¹⁹ GovLab interview with Morten Lind, Senior Advisor, Danish Geodata Agency, September 15.

²⁰ GovLab interview with Morten Lind, Senior Advisor, Danish Geodata Agency, September 15.

²¹ GovLab interview with Morten Lind, Senior Advisor, Danish Geodata Agency, September 15.

²² Jetzek, Thorhildur. “Managing complexity across multiple dimensions of open data: the case of the Danish Basic Data Program.” *Government Information Quarterly*. November 24, 2015. <http://www.sciencedirect.com/science/article/pii/S0740624X15300186>

deadlock. ... We had a lot of very good data but nobody used it.”²³ Some government agencies tried to access the data via individual agreements with each local authority: “a very cumbersome process even for a large organization.”²⁴

The deadlock was eventually broken through the intervention of the Minister of Finance. Analysis by the Ministry of Finance’s eGovernment Taskforce had identified address data as one of the data sets with the most potential use, leading the Minister of Finance to broker an agreement on access and use with municipalities. This agreement, officially called the Better Access to Public Data, but more widely known as the “free of charge agreement,” would make available data from the Cadastre and municipal property and dwelling registers, which comprised address data and their associated geographic coordinates, free of charge through a government portal, with those accessing it paying only the cost of distribution.²⁵ Although the Danish National Survey and Cadastre had advocated this solution, the outcome was, Lind concedes, surprising, since the government was at that stage committed, by policy and legislation, to selling its data.²⁶

The efforts and influence of Lind were crucial in advancing the cause of opening the data.²⁷ He had succeeded in convincing his superiors at the Danish National Survey and Cadastre of the necessity of opening the data without charge, since an agreement on pricing seemed elusive, and this became the agency’s position. He and his colleagues had involved potential users and stakeholders in pressing for the release of the data through a series of meetings and conferences at which the benefits and opportunities of the data set were explained. “It was very important for us to do that because we

²³ GovLab interview with Morten Lind, Senior Advisor, Danish Geodata Agency, September 15.

²⁴ GovLab interview with Morten Lind, Senior Advisor, Danish Geodata Agency, September 15.

²⁵ Jetzek, Thorhildur. “Managing complexity across multiple dimensions of open data: the case of the Danish Basic Data Program.” *Government Information Quarterly*. November 24, 2015. <http://www.sciencedirect.com/science/article/pii/S0740624X15300186>

²⁶ GovLab interview with Morten Lind, Senior Advisor, Danish Geodata Agency, September 15.

²⁷ GovLab interview with Bo Overgaard, Director, Septima, and Christian Fischer, Senior Advisor, Septima, September 22, 2015; and GovLab Interview with Thorhildur Jetzek, Department of IT Management, Copenhagen Business School, September 21 2015.

had to put ourselves on the agenda for the Minister of Finance, [who] has many important issues to solve, and this one was only a little one.”²⁸ Post Danmark was also, according to Lind, a major driver behind the decision to open, having declared early on that they wanted to rely on public data. Significant users such as Post Danmark and the emergency services have subsequently provided important feedback on how the data set could be improved.²⁹

The agreement came into effect on January 1, 2003, but legal issues delayed its full implementation until 2005, when an amendment to the law governing the Public Data Server removed the legal restrictions on the distribution of address data to third parties, as well as the fee for distribution.³⁰ The release was not conceived of as an open data project, according to Lind: “‘Open data’ ... was not a term we used then. We used the term ‘data free of charge’ ... [but] really it lives up to almost all of the definitions of open data today.”³¹ There would be no restrictions placed on the use or redistribution of the data beyond those required to satisfy the requirements of the law, particularly in regard to personal data or product marketing. Steps should continue to be taken, however, to ensure that such restricted, privacy-eroding activities are not being undertaken by internal or external users of the address data.

The municipalities were compensated EUR 1.3 million for loss of income from sales of data for the three years after the agreement was reached. At that point, the agreement offered them the opportunity to renegotiate for further compensation, but no further negotiations were sought by the municipalities.³² The compensation did not fully recompense the municipalities for the costs of data collection, but, as Lind notes, the data already had to be collected for their own use.

28 GovLab interview with Morten Lind, Senior Advisor, Danish Geodata Agency, September 15.

29 GovLab interview with Morten Lind, Senior Advisor, Danish Geodata Agency, September 15.

30 Morten Lind “The value of Danish address data: social benefits from the 2002 agreement on procuring address data etc. free of charge.” Danish Enterprise and Construction Authority, July 7, 2010. http://www.adresse-info.dk/Portals/2/Benefit_Value_Assessment_Danish_Address_Data_UK_2010-07-07b.pdf

31 GovLab interview with Morten Lind, Senior Advisor, Danish Geodata Agency, September 15.

32 GovLab interview with Morten Lind, Senior Advisor, Danish Geodata Agency, September 15.

“Our position was that the data was there, and of course we needed to have an assurance that the data will be updated, and of course there would be costs associated with this updating ... but this is something that the local authority has to do, agreement or not.”³³ The money offered was, instead, a compensation for no longer being able to sell the data. In addition, Danish Enterprise and Construction Authority (DECA), the agency responsible for road names and addresses in Denmark, calculated that the municipalities would realize savings from no longer having to negotiate data purchase agreements, deliver data or enforce licenses.³⁴

Beyond the local authorities themselves, which were the biggest users of the address data, the primary intended users of the data set were the public sector. Emergency services, public health disease prevention programs, national and county environmental agencies, public transport companies and Post Danmark were all expected to make use of the data set. Journey Planner, a then groundbreaking Web app allowing people to plan journeys using public transportation from one address to another, also relied on the address data set.³⁵ Input from intended users was sought through nationwide conferences and informal meetings between the Danish Survey and Cadastre and potential user groups such as the police.³⁶

Although the developers were aware of the potential benefits of the address data set for the private sector, Lind maintains that this benefit was always secondary to the desire to improve public sector efficiency and services. “We were of course aware that [this data] could have large benefits for the private sector in geomarketing, logistics, transport planning, route planning, and analysis of where to locate a new store, and so forth ... but that was not our selling argument. Our argument was the public sector. We have data that enables the public

³³ GovLab interview with Morten Lind, Senior Advisor, Danish Geodata Agency, September 15.

³⁴ Morten Lind “The value of Danish address data: social benefits from the 2002 agreement on procuring address data etc. free of charge.” Danish Enterprise and Construction Authority, July 7, 2010. http://www.adresse-info.dk/Portals/2/Benefit/Value_Assessment_Danish_Address_Data_UK_2010-07-07b.pdf

³⁵ GovLab interview with Morten Lind, Senior Advisor, Danish Geodata Agency, September 15.

³⁶ GovLab interview with Morten Lind, Senior Advisor, Danish Geodata Agency, September 15.

sector to be more efficient and to provide better services for citizens and for business, and it's stupid that we don't use that data.”³⁷

As far as its developers were aware, the Danish address data set was the first of its kind to be compiled and opened free of charge. Lind reports that they were influenced by a similar data set, Address-Point, compiled in the United Kingdom by the Ordnance Survey. Address-Point took the Royal Mail’s Postcode Address File (PAF), and attached geographic coordinates, creating a unique geolocated identifier for each address. Address-Point was, however, a commercial data set owned by the Ordnance Survey, and was not open.³⁸ Moreover, the PAF which enabled the development of Address-Point was controversially sold at the time of the privatization of the Royal Mail. In March 2014, the chairman of Parliament’s Public Administration Committee argued that, “The sale of the PAF with the Royal Mail was a mistake. Public access to public sector data must never be sold or given away again.”³⁹ Lind recalls, “I think we were totally on our own in deciding that the only possible model could be opening the data free of charge. I’m proud of that situation.”⁴⁰

Impact

The opening of the address data set was, almost universally, enthusiastically received, and the data set widely used. According to DECA, in 2009 address data was delivered to a total of 1,236 public and private parties, including 286 full, nationwide address sets. This response was based on feedback from only 14 of the 22 data distributors, and therefore underrepresents the total number of deliveries. Of these deliveries, about 70 percent were to the private sector, 20 percent to central and regional government, and 10 percent to the

³⁷ GovLab interview with Morten Lind, Senior Advisor, Danish Geodata Agency, September 15.

³⁸ <https://www.ordnancesurvey.co.uk/business-and-government/products/address-point.html>

³⁹ Hogge, Becky. “Open Data: Six Stories About Impact in the UK.” Omidyar Network. November 2015. https://www.omidyar.com/sites/default/files/file_archive/insights/Open%20Data_Six%20Stories%20About%20Impact%20in%20the%20UK/OpenData_CaseStudies_Report_complete_DIGITAL_102715.pdf

⁴⁰ GovLab interview with Morten Lind, Senior Advisor, Danish Geodata Agency, September 15.

municipalities. Of the 286 nationwide address sets delivered, 12 were for IT products with more than 1 million end users.⁴¹

We have data that enables the public sector to be more efficient and to provide better services for citizens and for business, and it's stupid that we don't use that data."

—Morten Lind, Danish Geodata Agency

The exceptions to this warm response were a small number of private vendors with agreements to supply competing address data to municipalities, who were unhappy with the decision to open the public data, since it cut off a source of revenue. The majority, according to Lind, embraced the opportunity to get out of data collection. "The more future-minded of these companies said, we are happy not to use our efforts in collecting these data. We want to use our efforts in making applications [and] value adding onto the data, ... and making something more intelligent."⁴² GPS and navigation companies were likewise happy to abandon data collection and adopt the open address data set, according to Lind.⁴³

Direct Financial Benefits

According to a 2010 study commissioned by the DECA, the direct financial benefits to society of the open address data for the period 2005-2009 totaled EUR 62 million through especially improved government back-end capabilities and more efficient service delivery. The total cost for the program over the same period was EUR 2 million, comprised of EUR 1.3 million in compensation to municipalities and the balance in costs to distribute the data through the Public Data Server (PDS).⁴⁴ The benefits of the agreement in 2010 were

⁴¹ Lind, Morten. "The value of Danish address data: social benefits from the 2002 agreement on procuring address data etc. free of charge." Danish Enterprise and Construction Authority. July 7, 2010. http://www.adresse-info.dk/Portals/2/Benefit/Value_Assessment_Danish_Address_Data_UK_2010-07-07b.pdf

⁴² GovLab interview with Morten Lind, Senior Advisor, Danish Geodata Agency, September 15.

⁴³ GovLab interview with Morten Lind, Senior Advisor, Danish Geodata Agency, September 15.

⁴⁴ Lind, Morten. "The value of Danish address data: social benefits from the 2002 agreement on procuring address data etc. free of charge." Danish Enterprise and Construction Authority. July 7, 2010. http://www.adresse-info.dk/Portals/2/Benefit/Value_Assessment_Danish_Address_Data_UK_2010-07-07b.pdf

expected to be EUR 14 million, while total costs would be EUR 0.2 million for distribution of the data through the PDS. Approximately 30 percent of the benefit was to the public sector, with the remainder to the private sector—including, notably the nongovernment Post Danmark, which saw major efficiency increases based on access to the data.⁴⁵ The study only measured direct financial benefit to those receiving address data via one of the PDS's 22 distributors, and did not attempt to place a value on the supplementary benefits derived further down the distribution chain through secondary and subsequent reuse, except to suggest that they are “very considerable.”⁴⁶

To illustrate these subsequent benefits, Lind points to the ubiquity of GPS technologies, used by nearly half of all Danish households, and reliant on the open address data set.⁴⁷ According to Lind, GPS companies interviewed by the Danish Geodata Agency report that they consider address data in Denmark the best in Europe. “We were able to improve the precision of car navigation systems in Denmark, and this has second-level impact on competitiveness and not using too much petrol.”⁴⁸

Nonfinancial Benefits

The Danish emergency services, who were among the early users of the open address data set, introduced a system built on standard car GPS navigation systems into 1,200 emergency vehicles, in preference to more expensive mobile devices for alarm reception and tracking, with resulting significant reductions in cost, response

45 Lind, Morten. “Addresses and Address Data—Experiences from Denmark.” August 11, 2014. <http://www.slideshare.net/Mortlin/addresses-and-address-data-experiences-from-denmark-37872471>

46 Lind, Morten. “The value of Danish address data: social benefits from the 2002 agreement on procuring address data etc. free of charge.” Danish Enterprise and Construction Authority. July 7, 2010. http://www.adresse-info.dk/Portals/2/Benefit/Value_Assessment_Danish_Address_Data_UK_2010-07-07b.pdf

47 Lind, Morten. “The value of Danish address data: social benefits from the 2002 agreement on procuring address data etc. free of charge.” Danish Enterprise and Construction Authority. July 7, 2010. http://www.adresse-info.dk/Portals/2/Benefit/Value_Assessment_Danish_Address_Data_UK_2010-07-07b.pdf

48 GovLab interview with Morten Lind, Senior Advisor, Danish Geodata Agency, September 15.

times and fleet efficiency.⁴⁹ The opening of the address data also improved response accuracy for the emergency services within Denmark. Lind notes that the increased precision of the geocoded address set improved the capacity of emergency services to find the right address. “The public could be more confident that 1-1-2 [emergency] services would be able to find them, and ... we had a really great impact on that.”⁵⁰

Not all anticipated benefits of the free-of-charge agreement were realized, however, although often these reflect a failure to embrace the data, rather than deficiencies within it. For example, the DECA analysis of the benefits of the address data set reported that adoption of the official address data set by the Central Business Register had been slow, and that in 2010 companies employing approximately 15 percent of the Danish workforce were registered at an address that was not in the official data set. As a result, DECA reports users found that it could take months or years for a new or changed commercial address to appear in their GPS. Given that emergency services as well as clients and suppliers rely on GPS to find commercial locations, the implications of such delays are potentially very costly.

⁵¹

Benefits to Developer Community

The number of applications using national address data went from two or three in 2005 to over 500 by 2009, according to Lind.

“The multiplication of use was astonishing. There has been a lot of development ... in geomarketing and digitization, and each development shows new use cases for the address data. I think that’s the purpose of open data. Originally you collect it with a purpose—hopefully!—but when you put it out into the open, you will experience new usefulness of the data that you haven’t foreseen. Perhaps you will find that there are purposes that you haven’t even thought

49 “112 in Denmark—a short overview of the Danish model.” 112 Roundtable in Bulgaria, March 6, 2009. http://www.eena.org/resource/static/files/2009_03_10_112_in_Denmark.pdf, p 9.

50 GovLab interview with Morten Lind, Senior Advisor, Danish Geodata Agency, September 15.

51 Lind, Morten. “The value of Danish address data: social benefits from the 2002 agreement on procuring address data etc. free of charge.” Danish Enterprise and Construction Authority. July 7, 2010. http://www.adresse-info.dk/Portals/2/Benefit_Value_Assessment_Danish_Address_Data_UK_2010-07-07b.pdf

of, that have an even better business case [with] more benefit than you have estimated from your original [case]. By going into the open data agenda, you have to say to yourself: I don't have the faintest idea what this can be used for! It's not up to me to decide ... or to control."

The Danish Basic Data Program

One of the most significant impacts of the release of the Danish address data was the impetus it gave to open other public data sets. Given the documented benefits of opening Danish address data, the Danish eGovernment strategy from 2011-2015 expanded the program with Good Basic Data for All,⁵² commonly known as the Danish Basic Data Program.

Basic data is essential information used repeatedly by the public sector in case processing, such as personal and company registration numbers, addresses, cadastral data and physical and political mapping data.⁵³ The Basic Data Program consists of a set of seven subagreements between multiple ministries and agencies, involving a total of nine registers, including business registers, the Cadastre, the Building and Dwelling Register, administrative and geographical boundaries, map and elevation data, place name and information register, and the register of property owners. A final subagreement covered the implementation of a common platform to disseminate the data from the other subprograms.⁵⁴

While personal data is protected by Denmark's Act on Processing of Personal Data (2000), all other data within these registers was opened for access and reuse by individuals, the public sector and private companies as of January 1, 2013. Between 2013 and 2015, the agencies involved improved and expanded the nine official registers supplying basic data, with the aim of developing a "high quality common administrative foundation ... efficiently updated at (a) sin-

⁵² "Fact sheet—basic data." Denmark Ministry of Finance. October 8, 2014. http://uk.fm.dk/~media/files/nyheder/press-releases/2012/10/basic-data/fact_sheet_basic_data.ashx?la=da

⁵³ "Fact sheet—basic data." Denmark Ministry of Finance. October 8, 2014. http://uk.fm.dk/~media/files/nyheder/press-releases/2012/10/basic-data/fact_sheet_basic_data.ashx?la=da

⁵⁴ GovLab interview with Thorhildur Jetzek, Department of IT Management, Copenhagen Business School, September 21, 2015.

gle site.” Implementing the Basic Data required the harmonization and conversion to a common format of all existing registers that would supply data; the addition of extra information to the basic data registers to better support public administration, and the development of a common infrastructure to distribute the data, the Data Distributor.⁵⁵

The Basic Data Program did not come about as a result of public pressure, demand from potential users or a desire for transparency. A project manager at the Danish Agency for Digitization interviewed by Thorhildur Jetzek in 2012 noted: “We believed in [transparency] but it was not going to open any doors in Denmark. Because in Denmark we didn’t really have the feeling that our government wasn’t transparent enough. We never really seemed to regard that as a serious issue for Denmark.”⁵⁶

Instead, the Basic Data Program was developed as a result of a business decision to improve public sector efficiency and data quality, influenced by New Public Management reforms over the last two decades in which improving efficiency became the focus of efforts to control spending without increasing taxation.⁵⁷ Jetzek reports that initially there was resistance in some quarters to the idea of fully opening the data, rather than simply finding ways to improve the efficiency of its collection, management and use, while continuing with the prevailing model of offering data for sale, as mandated in Danish law. The decision-makers involved were ultimately persuaded to open the data by the experience of open data projects in the Netherlands, and by the work of Dutch analyst Marc de Vries, who argued that the income from data sales was typically a fraction of the actual costs of collection,⁵⁸ while the economic benefits of opening

55 “Fact sheet—basic data.” Denmark Ministry of Finance. October 8, 2014. http://uk.fm.dk/~media/files/nyheder/press-releases/2012/10/basic-data/fact_sheet_basic_data.ashx?la=da

56 Jetzek, Thorhildur (2015) Managing complexity across multiple dimensions of open data: the case of the Danish Basic Data Program. In press, pp 19-20.

57 GovLab interview with Thorhildur Jetzek, Department of IT Management, Copenhagen Business School, September 21, 2015.

58 See, for example, Sawyer, Geoff and Marc de Vries. “About GMEs and Data: Geese and Golden Eggs.” European Association of Remote Sensing Companies. December 2012. http://earsc.org/file_download/134/Open+Data+study+Final+report.pdf

the data could be up to 17 times that cost.⁵⁹ The Pricing of Public Sector Information Study carried out by de Vries and others in 2011 for the European Commission concluded that only reducing charges to zero or cost-recovery increased data reuse by 1,000 to 10,000 percent, attracted new users such as small and medium enterprises, and increased economic activity, market dynamism, innovation, employment and efficiency, with little impact on cost.⁶⁰

The advent of Denmark's Basic Data Program also coincided with a new government in 2012 that responded to the worldwide economic crisis with a focus on job creation and innovation. Although opening the data was a means to improve efficiency, rather than an end in itself, there was no conflict between efficiency, innovation, and open data ideals. Lind notes, "We could align the purpose of the Basic Data Program to make government more efficient with open data ideas [of] making the basic data open for the private sector and citizens to make smart things."⁶¹

As intended, the Basic Data Program has provided opportunities for the private sector. Septima, a geodata startup, was founded around the same time as the Basic Data Program in 2013. It offers a range of products and services using the data, including data enhancement, such as calculating the distance between schools and houses; Web-based map components and search tools for public data; consulting to organizations like the Geodata Agency; and subcontracting to the large IT company involved in developing the distribution platform for the Basic Data Program.⁶²

Bo Overgaard and Christian Fischer of Septima note that some of the company's work would have been possible without the Basic Data Program, but that the Program opened up possibilities for products and services that had not previously existed.

59 GovLab interview with Thorhildur Jetzek, Department of IT Management, Copenhagen Business School, September 21, 2015.

60 de Vries, Mark, et al. "Pricing of Public Sector Information Study (POPSIS) Summary Report." European Commission Information Society and Media Directorate-General. October 2011. <http://www.epsiplatform.eu/sites/default/files/summary.pdf>

61 GovLab interview with Morten Lind, Danish Geodata Agency, September 15.

62 GovLab interview with Bo Overgaard, Director, Septima, and Christian Fischer, Senior Advisor, Septima, September 22, 2015.

“[The Basic Data release] was part of the basis of us establishing our company, that now we [could] do some interesting things with all the new open data sets.” These new opportunities included the possibility of speculative development, without the need for a customer. “[The Basic Data Program] has been a great help to us. Before, our clients had to have the rights [to the data]. Now we can develop our products without having a customer. We can get the data, [make] some nice products and show them to potential customers. It also [enables us] to do different business models. Before, the license was connected to a user, and if you wanted to make a mobile application, you had to buy a site license, which could be very expensive.”⁶³

Since the establishment of the Basic Data Program, the private sector market for this data has grown. More companies like Septima have been established, and existing companies that had not previously worked with geodata have moved into the sector. According to Overgaard and Fischer, this development has spawned a lot of interesting solutions. It has become standard for insurance companies, for example, to use elevation data to assess risks of flooding where they would previously have had to rely on aggregated risk by region. Municipalities are also using this data to model and plan for flooding risk as sea levels rise.⁶⁴

Jetzek observes that the impact on the private sector is harder to quantify. “There are a few companies in Denmark that you can say were founded on the basis of the Basic Data Program, but I think most of the use is, and will continue to be, companies that are using the data and a thousand other things in their daily processes. For most companies, it’s hard to say exactly what effect it would have if the data were not available, or if [they] were more expensive.”⁶⁵ She reports that the Danish Geodata has been attempting to track the impact of its data releases over time, with a baseline survey conducted in the year the Basic Data Program began, and follow-up surveys planned for future years.⁶⁶

⁶³ GovLab interview with Bo Overgaard, Director, Septima, and Christian Fischer, Senior Advisor, Septima, September 22, 2015.

⁶⁴ GovLab interview with Bo Overgaard, Director, Septima, and Christian Fischer, Senior Advisor, Septima, September 22, 2015.

⁶⁵ GovLab interview with Thorhildur Jetzek, Department of IT Management, Copenhagen Business School, September 21, 2015.

⁶⁶ GovLab interview with Thorhildur Jetzek, Department of IT Management, Copenhagen Business School, September 21, 2015.

The frequency with which data sets are updated varies according to the register. Some elevation data relies on aerial surveys that might be a decade old. Data in the address data set, by contrast, is continuously updated as new addresses are added by local authorities. It takes seconds after addresses are locally created for the information to become available as part of the national data set. “Fifteen seconds after an address is recorded [by] a local authority, it is available in our service, and that’s very important if you are making e-government for GPS. A delay of one to two weeks would be a problem. For e-government, if you go to one part of government and make an address it should be available instantly for the next purpose.”⁶⁷

Challenges

Data Quality

Once the decision to open data for the Danish Data Program was made, the challenges became those of data quality and governance across the range of agencies and registers. Jetzek notes that the nature of the data collection has created significant variability in quality.

“In Denmark, data collection is very distributed, and in most cases the data is just a part of what they do: a necessary part, but not the main part. Making some standards or blueprints for the organizations to follow when they’re going to model or open the data is very important. Some agencies like the Geodata Agency are very professional data custodians: They’ve done it for a long time, and there are standards across Europe, and there’s a lot of knowledge and professionalism. ... Whereas others ... don’t even know what the data model is. The data is just something that someone is keying in somewhere, and then they have this register. It’s very important to have some common understanding and some common skills about how ... we handle all this data ... [but] this skill [isn’t] common in the public sector. We haven’t really focused on data as a key resource.”⁶⁸

⁶⁷ GovLab interview with Morten Lind, Senior Advisor, Danish Geodata Agency, September 15.

⁶⁸ GovLab interview with Thorhildur Jetzek, Department of IT Management, Copenhagen Business School, September 21, 2015.

Overgaard and Fischer observe that accuracy appears to be higher in areas where data collection is directly connected with government revenue by comparison with areas in which data is an incidental by-product of another activity. “If money is directly involved, the quality tends to be good. The building and dwelling register, for example, is involved in taxation, and it’s pretty precise.”⁶⁹ The business registers, by contrast, which are not directly involved in revenue collection, are much more error-prone. Both Jetzek and Overgaard and Fischer cite weaknesses in the register of companies where the nature of the business is described: The company itself is listed, but the nature of its commercial activity is wrong. “In the business register, the businesses are supposed to say what they do, to classify the business, but you can look at the register and see that a pizza shop is classified as a shoemaker because the house used to belong to a shoemaker, and then it changed ... and nobody [corrected] the data.”⁷⁰ Some uses that Septima had hoped to make of the data from the business register, for example, turned out not to be possible because of deficiencies in the quality of the data.⁷¹

Having been open longer, the address data set has had longer to address these quality concerns. The aim of the address data subprogram was “to further improve the quality and accessibility of address data, and enrich ... the infrastructure behind the distribution of address data.”⁷² The Geodata Agency has also collaborated with OpenStreetMap to identify and correct address errors in the data set. One of the greatest boons to accuracy, however, is simply opening the data and having it used, which provides opportunities for scrutiny and correction—what de Vries calls a partial outsourcing of

⁶⁹ GovLab interview with Bo Overgaard, Director, Septima, and Christian Fischer, Senior Advisor, Septima, September 22, 2015.

⁷⁰ GovLab interview with Thorhildur Jetzek, Department of IT Management, Copenhagen Business School, September 21, 2015.

⁷¹ GovLab interview with Bo Overgaard, Director, Septima, and Christian Fischer, Senior Advisor, Septima, September 22, 2015.

⁷² GovLab interview with Morten Lind, Senior Advisor, Danish Geodata Agency, September 15.

quality control⁷³—and raises the quality of the data set.⁷⁴ “Address data is very good in Denmark, and that’s because so many are using them. If no one uses the data, then no one finds the errors,” note Overgaard and Fischer.⁷⁵ Centralization of registration also allows for easier and faster correction, since users only have to report errors and omissions once,⁷⁶ through a portal that connects users with the responsible municipality.⁷⁷ With the Basic Data Program, the address data set also “took an extra step toward the open,” according to Lind, by removing the earlier distribution costs and making the data easier and cheaper to access and use. “Every citizen can now download a data set of every address free of charge, or access our services for a Web application.”⁷⁸

Governance of the Data Project

Project management issues also arose over the governance of the program, which was initially similarly distributed and collaborative. The number of entities involved in the governance of the project created major complexities and management challenges. Delays occurred when the development and implementation of the Data Distributor was much more complicated than originally expected, overran its timeline and began to impede progress elsewhere. “This kind of project needs to be handled like infrastructure,” says Jetzek. “If you want to have an initiative that covers a lot of different data sets, it [is] really important to think about how you govern this. It’s hard to do it top-down; it’s also almost impossible to do it bottom-

73 de Vries, Mark, et al. “Pricing of Public Sector Information Study (POPSIS) Summary Report.” European Commission Information Society and Media Directorate-General. October 2011. <http://www.epsiplatform.eu/sites/default/files/summary.pdf>

74 de Vries, Mark, et al. “Pricing of Public Sector Information Study (POPSIS) Summary Report.” European Commission Information Society and Media Directorate-General. October 2011. <http://www.epsiplatform.eu/sites/default/files/summary.pdf>

75 GovLab interview with Bo Overgaard, Director, Septima, and Christian Fischer, Senior Advisor, Septima, September 22, 2015.

76 Lind, Morten. “The value of Danish address data: social benefits from the 2002 agreement on procuring address data etc. free of charge.” Danish Enterprise and Construction Authority. July 7, 2010. http://www.adresse-info.dk/Portals/2/Benefit_Value_Assessment_Danish_Address_Data_UK_2010-07-07b.pdf

77 GovLab interview with Bo Overgaard, Director, Septima, and Christian Fischer, Senior Advisor, Septima, September 22, 2015.

78 GovLab interview with Morten Lind, Senior Advisor, Danish Geodata Agency, September 15.

up. There has to be a middle way.”⁷⁹ Jetzek notes that, in response to the difficulties encountered, the governance structure of the Basic Data Program was altered to make it “less democratic.”⁸⁰

Looking Forward

Expected Benefits of the Basic Data Program

At launch, the expected benefits of the Basic Data Program were efficiencies to government of about EUR 34.8 million from 2020 through decreased administrative costs, of which two-thirds of the benefit (EUR 23.5 million) would be to municipalities. The release of the data was also expected to generate returns in the private sector of up to EUR 67 million, through the encouragement of new digital services and products, and the removal of bureaucratic barriers and data costs. Finally, individuals would benefit from increasingly improved and integrated public services.⁸¹

Addressing Capacity Issues

Lind identifies secure funding as the main future challenge of the Basic Data Program and projects like it. “One of the issues with open data is that you have to have a stable, sustainable source of funding ... to be able to update, and to ensure the long-term quality of the open data. We [were] lucky that the Danish Basic Data Program came at a time when we needed funding of this kind. If we could have done anything differently or better, we should have focused on getting better [long-term] funding. This reorganization of government process we have just been through shows how easily things can change, and suddenly you can be without any source of funding.”⁸²

⁷⁹ GovLab interview with Thorhildur Jetzek, Department of IT Management, Copenhagen Business School, September 21, 2015.

⁸⁰ GovLab interview with Thorhildur Jetzek, Department of IT Management, Copenhagen Business School, September 21, 2015.

⁸¹ “Fact sheet—basic data.” Denmark Ministry of Finance. October 8, 2014. http://uk.fm.dk/~media/files/nyheder/press-releases/2012/10/basic-data/fact_sheet_basic_data.ashx?la=da

⁸² GovLab interview with Morten Lind, Senior Advisor, Danish Geodata Agency, September 15.

For Jetzek, the major future challenge is developing skills in data collection among public servants. “Getting the data to be really clear and easy to use, so that it’s used widely across the public and private sectors and in business-critical processes, is a huge accomplishment on its own. If you want the data to be really used across the public and private sectors for more than just creating an interesting visualization or one-off something, I think you really need to focus on [data collection].”⁸³

The Danish effort to consolidate and share address data has already had major impacts on the country, particularly in inspiring a much larger effort to open and consolidate all types of public data. To enable ongoing impacts, though, work must be done to engender readiness—in the form of adequate funding and employee training—within the public service, a common theme across global efforts to create impacts through the opening of data.

⁸³ GovLab interview with Thorhildur Jetzek, September 21, 2015.

CHAPTER 7

Opening Canada's T3010 Charity Information Return Data

Accountability of Charities through Open Data

By Juliet McMurren, Stefaan Verhulst and Andrew Young

Summary: Charitable data from the Canada Revenue Agency's information return has been available to the public since 1975. In 2000, the Charities Directorate of the Canada Revenue Agency (CRA) began making information from the T3010 Registered Charity Information Return from organizations with charitable status available online and by CD-ROM on request, and, in 2013, all data sets since 2000 were transcribed and made available online via the government's data portal under a commercial open data license. The resulting data set has been used on its own and in combination with other data by sector groups and academic researchers to explore the state of the nonprofit sector, improve advocacy by creating a common understanding between regulators and charities, and create intelligence products for donors, fundraisers and grant-makers.

Dimension of Impact: Improving Government—Improving Services

Key Takeaways:

- Government is not the only sector about which open data can create broad impacts related to transparency and accountability. There is broad interest in and appetite for giving the public a

greater understanding of the traditionally obscured financial dealings of nonprofits.

- Open data—even in situations that are largely focused on transparency and accountability—could stand to provide benefits to organizations releasing the data. In the nonprofit sector, for example, actors whose financial activities place them in higher standing in the public eye—particularly in comparison to demonstrated bad actors—stand to benefit from increased citizen trust and, likely, more donations and volunteerism as a result.
- Bringing more people to the table in tracking the activities of large organizations—particularly when that tracking can be done through large-scale data analysis—can uncover major instances of corruption or fraud. Not only can openness lead to *more* eyes on relevant data, but it can also lead to the *right* eyes on the data—i.e., making the data accessible to the people with the specific knowledge needed to recognize troubling discrepancies.

Context and Background

Canada is a developed, high-income North American nation with a population of 35.8 million in July 2015.¹ It is a federal parliamentary democracy and a constitutional monarchy, and is officially bilingual (English and French). Canada had the world's 15th highest nominal per capita income in 2014, according to the World Bank² and International Monetary Fund.³ Canada was ranked eighth globally in the 2014 UN Human Development Index.⁴ Its economy is the 11th largest in the world.⁵

Canada is ranked highly on measures of government openness and transparency, civil liberties and press freedom. It is ranked eighth on the 2015 World Press Freedom Index by Reporters without Borders,

1 <http://www.statcan.gc.ca/daily-quotidien/150929/dq150929b-eng.htm>

2 <http://data.worldbank.org/indicator/NY.GDP.PCAP.CD>

3 <http://www.imf.org/external/pubs/ft/weo/2015/01/weodata/weorept>

4 <http://hdr.undp.org/en/content/table-1-human-development-index-and-its-components>

5 <http://databank.worldbank.org/data/download/GDP.pdf>

⁶ and its press was considered free in 2015.⁷ In 2014, it was ranked 10th worldwide in the Corruption Perception Index by Transparency International.⁸

Canada signed the Open Government Partnership in 2011, and, as of late 2015, is implementing its second action plan, which envisages advances in three streams of activity: open data, open information, and open dialogue.⁹ As part of its April 2012 action plan, the Government of Canada committed itself to the development of an Open Government Licence to replace existing licenses on open government data and remove restrictions on the reuse of Government of Canada data. The Open Government Licence has been adopted by the governments of Alberta, British Columbia and Ontario, as well as by the federal government.¹⁰ Canada is also a signatory to the G8 Open Data Charter, which commits countries to a policy of open by default.¹¹ The government of Canada releases data sets on its data portal, open.canada.ca, which had 244,640 data sets available for download in late 2015.¹²

Recent Canadian federal budgets have also contained commitments to advancing open data. The 2014 Canadian federal budget pledged CDN\$3 million over three years for the creation of an Open Data Institute to be based in Waterloo, Ontario. Once established, the Institute will be charged with aggregating large data sets, helping with the development of standards for interoperability and encouraging the development of new data-driven apps. This initiative is being driven by the Canadian Digital Media Network, a federally funded Centre of Excellence for Commercialization and Research responsible for supporting Canadian ICT companies and entrepre-

⁶ <http://index.rsf.org/#/>

⁷ <https://freedomhouse.org/report/freedom-press/freedom-press-2015#.VgvwTfIViko>

⁸ <http://www.transparency.org/cpi2014/results>

⁹ <http://www.opengovpartnership.org/country/canada>

¹⁰ “Canada: Open Government License.” Open Government Partnership. <http://www.opengovpartnership.org/sites/default/files/Inspiring%20Story%20-%20Canada.pdf>

¹¹ “G8 Open Data Charter.” Global Affairs Canada. September 19, 2009. http://www.international.gc.ca/g8/open_data_charter-charte_du_g8_sur_les donnees_ouvertes.aspx?lang=eng&ga=1.182752998.1958574156.144285339

¹² <http://open.canada.ca/data/en/dataset>

neurs, and encouraging their collaboration with research institutes, governments and other organizations.¹³

Canada has the second-largest nonprofit sector per capita in the world after the Netherlands.¹⁴ The 2014 Canadian federal budget states that there are over 86,000 registered charities in Canada.¹⁵ As of 2009, core nonprofits (those other than hospitals, colleges and universities) accounted for 2.5 percent of Canada's annual GDP, and growth in this sector outpaced the Canadian economy for six of the 11 years between 1997 and 2007. The nonprofit sector generated nearly six times more value than the automotive industry, and employed 1.3 million people.¹⁶

Project Description and Inception

Section 149.1(15) of Canada's Income Tax Act obliges the government to make information contained in a public information return available to the public, although the form in which this information is released is left to the discretion of the Minister.¹⁷ Organizations with registered charitable status in Canada are currently required to file an annual T3010, or Registered Charity Information Return, within six months of the end of the fiscal year. The 10-page form, which is similar in content to the US 990 tax return, asks the charity to disclose its domestic and international charitable activities, revenue totals per type of source (e.g. government, foundations, donors), expenditures, charitable gifts, and non-partisan political activities.¹⁸

Since 1967, Canadian charities have been required to make application to a central registry and file annual financial and operational

13 "Chapter 3.2: Fostering Job Creation, Innovation and Trade." Canadian Budget 2014. <http://www.budget.gc.ca/2014/docs/plan/ch3-2-eng.html>

14 "Key Facts About Canada's Charities." *Imagine Canada*. <http://www.imaginecanada.ca/resources-and-tools/research-and-facts/key-facts-about-canada%E2%80%99s-charities>

15 "Chapter 3.4: Supporting Families and Communities." Canadian Budget 2014. <http://www.budget.gc.ca/2014/docs/plan/ch3-4-eng.html>

16 "Satellite Account of Non-profit Institutions and Volunteering." Statistics Canada. 2007. <http://www.statcan.gc.ca/pub/13-015-x/13-015-x2009000-eng.pdf>

17 <http://www.ifrc.org/docs/idrl/635EN.pdf>

18 <http://www.cra-arc.gc.ca/E/pbg/tf/t3010/t3010-15e.pdf>

reports with CRA.¹⁹ The T3010 itself was introduced to replace earlier returns in 1997, and substantially revised in 1999, and again in 2009,²⁰ after consultation with the third sector through the Voluntary Sector Initiative.²¹ Another revision in 2012 increased reporting requirements for charities involved in political activities.²²

Most of the data collected have been made public since 1976, when an amendment to the Income Tax Act required all registered charities to file an annual public information return outlining their income, donations received, gifts made, costs of fundraising, overhead expenses, and salaries.²³ According to David Lasby, Director of Research at Imagine Canada, an umbrella group that supports and advocates for Canada's charities, as the form has been expanded over time, more information has become open, thanks to an implicit "open by default" position. "Back all those many years ago when they wrote the legislation, they built in the public information return component. The initial return was not terribly detailed, but as the return has become more and more detailed, [the additional material] has in the main been considered to be part of the information return and is publicly available. This was open data before open data ever existed."²⁴ The minority of data not made public touches on personal confidentiality: Lasby notes, for example, that the open data set includes the total amount of cross-border funding received by an organization, but that names of individuals and organizations providing that funding are withheld.²⁵

19 "Report on the Law of Charities." Ontario Law Reform Commission. 1996. http://www.mtroyal.ca/cs/groups/public/documents/pdf/npr03_lawcharities.pdf

20 Brouard, François. "T3010 Challenges for Research." Sprott School of Business Discussion Paper. May 16, 2014. <http://http-server.carleton.ca/~fbrouard/documents/PaperT3010ChallengesFB201405.pdf>

21 "Satellite Account of Nonprofit Institutions and Volunteering Final Report." Voluntary Sector Initiative. January 2003. http://www.vsi-isbc.org/eng/knowledge/pdf/satellite_final_report.pdf

22 Broder, Pete. "2012 Federal Budget Features Tighter Reporting on Charities' Political Activities." *LawNow*. May 1, 2012. <http://www.lawnow.org/2012-federal-budget-features-tighter-reporting-on-charities-political-activities/>

23 "Report on the Law of Charities." Ontario Law Reform Commission. 1996. http://www.mtroyal.ca/cs/groups/public/documents/pdf/npr03_lawcharities.pdf

24 GovLab interview with David Lasby, Director of Research, Imagine Canada, October 14, 2015.

25 GovLab interview with David Lasby, Director of Research, Imagine Canada, October 14, 2015.

The information was originally made available upon request on paper,²⁶ and then on data tape during the 1990s.²⁷ In 2000, the Canada Revenue Agency (CRA) began making data from T3010 available to the public online through the Charities Internet Display System (CIDS), and by CD-ROM upon request, in comma-separated value (CSV) format. Those requesting the data were able to reuse it with CRA permission, but the data available online was only updated twice annually, limiting its commercial usefulness.²⁸

In 2013, T3010 data for 2012 was transcribed into machine-readable format and released with a commercially friendly license on the government's open data portal, open.canada.ca. The 2013 data were released the following year, and 2014 data were expected to be released in early 2016. In addition, the released data included digitized records of all gifts from charities and foundations, instead of only the first 10 such gifts on each T3010, as had previously been the case.²⁹ More frequent updates, issued about once a month, are still available on CD-ROM by request from CRA.³⁰

Although Canada's T3010 resembles the US 990 in many respects, the digitization of the Canadian dataset renders it much more commercially useful than the US dataset, which is only available in image PDF format. The T3010 data set contains a digitized, machine-readable record of all gifts from charities and foundations, something no other country provides.³¹

26 GovLab interview with Peter Elson, Community Development Researcher, Mount Royal University and University of Victoria, October 21, 2015.

27 GovLab interview with David Lasby, Director of Research, Imagine Canada, October 14, 2015.

28 GovLab interview with Michael Lenczner, Chief Executive Officer, Ajah, September 23, 2015.

29 GovLab interview with Michael Lenczner, Chief Executive Officer, Ajah, September 23, 2015.

30 GovLab interview with David Lasby, Director of Research, Imagine Canada, October 14, 2015.

31 GovLab interview with Michael Lenczner, Chief Executive Officer, Ajah, September 23, 2015.

Impact

Benefits to Charities

Outside of government, probably the heaviest users of the T3010 data set have been charities themselves. One of the earliest nonprofit users of the data set, Imagine Canada, has been working with the data since the early 1990s.³² In the beginning, the data was used as a population count of charitable organizations, and an indicator of the key areas covered by the T3010, such as the charitable activities, revenues, expenditure, size, location and distribution—enabling a macro-level view of the philanthropic sector that still provides value today. Over time, Imagine Canada's use of the data has become both heavier and more complex. “I use [the T3010 data] for everything from a sampling frame for survey work, to … a data source for answering policy-relevant questions, to being able to tell people what's going on in the charitable sector in a given province. It looms large in every data-driven thing we do. You're essentially talking about the vital statistics of every organization that's in our key stakeholder area. … We use it in our business processes; we use it to understand what's going on in organizations we don't have direct contact with, but are actually responsible for … as a charity that seeks to serve the interests of other charities. Basically, almost every person in the organization on a daily basis touches the data, [although] a lot of the time they don't realize it. … I have the T3010 database open probably nine days out of 10.”³³

A typical research project involving the data set for Imagine Canada was an examination of charity financing and its variability according to the size, subsector and activity type, and region. The project provided a benchmark for the organizations themselves, and allowed them to describe their particular situation to stakeholders and policymakers.

“Policymakers look at the charitable sector and they see the data in the aggregate: where all of the money comes from and where it goes to. Organization leaders see it more from the perspective of where

³² GovLab interview with David Lasby, Director of Research, Imagine Canada, October 14, 2015.

³³ GovLab interview with David Lasby, Director of Research, Imagine Canada, October 14, 2015.

their specific organization gets its funding, and what the norms are. So policymakers look at the sector and say, wow, government funding plays a huge role. Charitable leaders look at the sector and say ... well, most of that funding actually goes to a relatively small number of organizations: most of us are highly dependent on earned income and gifts and donations.”³⁴

Imagine Canada has also combined the T3010 data set with other data, including IRS 990s, foundation annual reports, social media and data mined from websites, to create a by-subscription Canadian directory to foundations and corporations, now called Grant Connect.³⁵

Grant Connect offers searchable information on grant prospects available to Canadian organizations, with detailed descriptions, giving histories and biographies of decision-makers. The data is verified by Imagine Canada with the grant-makers to ensure accuracy. According to Imagine Canada, the service had over 1,000 subscribers in late 2015.³⁶

Philanthropic Foundations of Canada (PFC) is also a heavy user of the data set. A member association of nearly 130 charitable grant-makers, almost of all of whom are charities themselves, PFC has two mandates: to support the growth of effective grant-making philanthropy through services such as professional development, publications and conferences for members and nonmembers; and to advocate for the field, and for the growth of philanthropy in Canada. In support of these aims, PFC carries out research and collects data, both independently and in collaboration with organizations like Imagine Canada, with whom it recently produced a report on the top 150 grant-making bodies in Canada.³⁷ PFC also relies on the basic biographical data from the T3010—dates of establishment and registration, membership of the board of directors, and financials—in its business processes, using it to develop an overview of organization applying for membership, ascertain the eligibility of appli-

³⁴ GovLab interview with David Lasby, Director of Research, Imagine Canada, October 14, 2015.

³⁵ GovLab interview with David Lasby, Director of Research, Imagine Canada, October 14, 2015.

³⁶ <http://www.imaginecanada.ca/grant-connect>

³⁷ GovLab interview with Hilary Pearson, President, Philanthropic Foundations Canada, October 20, 2015.

cants for membership, and verify members' reported assets and grants in fee calculations.³⁸

Lasby notes that one of the benefits of the T3010 data set has been the creation of a common set of facts in discussions between the charitable sector and government. "Because so much of the [T3010] data is public, when we're having policy discussions with government, we are, in the main ... talking about the same data that we can both have access to and both parties can see. ... What this [data set] is tremendously useful for is getting all of the stakeholders on the same page ... without multiple competing realities."³⁹ The data set also allows advocates to demonstrate the impact of the charitable sector to legislators, through the number of employees and revenues.⁴⁰ The data also allows the presence of the sector to be represented spatially: StepUpBC has used the data, for example, to map nonprofit organizations in British Columbia.⁴¹

Both Imagine Canada and CRA have employed the T3010 data to assist the public in finding, researching and giving to charities. CRA encourages potential donors to search for and investigate charities before donating through its Charities Listing search tool,^{42,43} which uses the T3010 data set. Working in partnership with CRA, Imagine Canada has built on the data set to offer Charity Focus, a tool that provides in-depth, year-to-year financial information about charities for potential givers. Charities can also augment the information available on Charity Focus by uploading their annual reports, program information, and mission and financial statements.⁴⁴

38 GovLab interview with Hilary Pearson, President, Philanthropic Foundations Canada, October 20, 2015.

39 GovLab interview with David Lasby, Director of Research, Imagine Canada, October 14, 2015.

40 GovLab interview with Michael Lenczner, Chief Executive Officer, Ajah, September 23, 2015.

41 <http://stepupbc.ca/discover-stepup-bc-sector-information/stepup-open-data#.VgwEqPlViko>

42 GovLab interview with Peter Elson, Community Development Researcher, Mount Royal University and University of Victoria, October 21, 2015.

43 <http://www.cra-arc.gc.ca/chrtsgvng/lstngs/menu-eng.html>

44 <http://www.charityfocus.ca/EN/Pages/Home.aspx>

Enabling New Streams of Research

Despite the relatively small number of Canadian researchers engaged in studying the charitable sector, the academic community has also made use of the data to examine charity financing and other econometric questions.

Peter Elson, a researcher in community development at Mount Royal University and the University of Victoria, has used the data to map the location of charities and compare it to socioeconomic data on prevalence and extent of poverty in those areas. He has also examined the ways in which T3010 data captures policy implementation and charities' engagement in public policy, and the changes in patterns of giving brought about by the introduction of donations of shares. Elson has also used the data set to examine the rate and patterns of voluntary deregistration, largely among faith-based charities as churchgoing declines.⁴⁵ He was also the founding coordinator of the T3010 user group, a group of academic and other users who consult with the CRA on changes and improvements to the form.⁴⁶ François Brouard of Carleton University is the current T3010 user group coordinator and has used the data to conduct research on the fundraising activities of small charities⁴⁷ and on changes in fundraising behavior over time.⁴⁸ Other academics have used the archival data from the T3010 data set to depict the changing demographic of Canadian charities.⁴⁹

⁴⁵ GovLab interview with Peter Elson, Community Development Researcher, Mount Royal University and University of Victoria, October 21, 2015.

⁴⁶ <http://http-server.carleton.ca/~fbrouard/T3010group>

⁴⁷ Brouard, François. "Small Charities: An analysis of change in fundraising activities." Sprott School of Business. <http://www.mtroyal.ca/cs/groups/public/documents/pdf/brouardpro.pdf>

⁴⁸ Neilson, Leighann, François Brouard and Anahit Armenakyan. "Fundraising Methods: Past, Present, Future." SCSE-CSES Research Report. October 2012. <http://www.mtroyal.ca/cs/groups/public/documents/pdf/ins-brouardrpt2.pdf>

⁴⁹ Spyker, Clifford L. "There Is No Accounting for Landscape: Capturing the Demographics of Canadian Registered Charities (2000—2008)." Institute for Nonprofit Studies Working Paper # 3. 2011. <http://www.mtroyal.ca/cs/groups/public/documents/pdf/wkgpa/per3.pdf>

Creating a New Business Asset

The open availability of the data set has encouraged startups interested in developing intelligence products for charities. Ajah, a Montreal-based data company used T3010 data to develop Fundtracker, an online research tool that assists nonprofits to identify, research and approach potential funders. Ajah began scraping the data from the CRA website in 2010. “The CRA didn’t know what to make of it,” Ajah CEO Michael Lenczner reports. “Initially they had a statement on the website saying they didn’t support scraping instead of accessing the data by CD-ROM.”⁵⁰

Emboldened by Treasury Board support for open data projects like theirs, Ajah continued. In 2011, it launched Fundtracker. In addition to the T3010, Fundtracker drew on multiple open sources of government and charitable data—federal and provincial government spending; charity financial and grant records and annual reports; and corporate social responsibility reports⁵¹—to provide a comprehensive and searchable database of potential funders.⁵² By cross-referencing the T3010 with other data, Ajah has been able to amass an unparalleled database on the nonprofit sector,⁵³ which it used to build Fundtracker, but also as a survey frame for other inquiry through its nonprofit consultancy arm, PoweredByData.⁵⁴

Within Canada, the availability of the data set has resulted in more companies competing to help the nonprofit sector use the information than in the US. “In the US, there are two companies,” says Lenczner, “In Canada there are five.”⁵⁵ Experience gained working with the T3010 data set has enabled some of these Canadian companies to enter and compete in markets where the data has not been

50 GovLab interview with Michael Lenczner, Chief Executive Officer, Ajah, September 23, 2015.

51 Lenczner, Michael. “Canada—a World Leader in Non-profit Data.” Government of Canada. July 4, 2014. <http://open.canada.ca/en/blog/ajah-and-fundtracker>

52 <http://ajah.ca/>

53 Lenczner, Michael and Susan Phillips. “From Stories to Evidence: How Mining Data Can Promote Innovation in the Nonprofit Sector.” *Technology Innovation Management Review*. July 2012. <http://timreview.ca/article/575>

54 <http://poweredbydata.org/>

55 GovLab interview with Michael Lenczner, Chief Executive Officer, Ajah, September 23, 2015.

opened, and is not as easy to access. A British Columbia company, Metasoft, has entered the U.S. market and competes with the Foundation Center to provide fundraising intelligence to the nonprofit market there. According to Lenczner, the opportunity to hone their skills, and their software, on a freely accessible and reusable data set gives Canadian companies a competitive advantage. Having developed the software and an understanding of what is possible using this data, they are able to go into markets where the data is not open, and scrape the data from annual reports and other sources to provide the same service. Lenczner notes, “There’s not a lot of ... little plucky startups [in the US] doing this, because there’s a lot of upfront costs. I would not be doing this in the U.S. at all.”⁵⁶

Financial Benefits for Government and Identification of Tax Evasion

Although the intended benefits to government of the T3010 were administrative, the opening of the data has yielded some unexpected financial benefits. A contractor using T3010 data from 2005 to analyze charitable foundations in greater Toronto uncovered evidence of massive tax fraud, in which illegally operating charities in the Toronto area sheltered CAN\$500 million in 2005 alone. A later newspaper investigation suggested that the fraudulent donations amounted to almost CAN\$1 billion per year, with \$3.2 billion illegally sheltered, a sum equivalent to 12 percent of annual charitable giving in Canada. The exposure of the fraud led to systematic closure of charity tax shelters.⁵⁷ Writing about the impact of open data on the tax evasion scandal, nonprofit researcher and open data advocate David Eaves comments, “When data is made readily available in machine readable formats, more eyes can look at it. This means that someone on the ground, in the community ... who knows the sector, is more likely to spot something a public servant in another city might not see because they don’t have the right context or bandwidth.”⁵⁸

⁵⁶ GovLab interview with Michael Lenczner, Chief Executive Officer, Ajah, September 23, 2015.

⁵⁷ Eaves, David. “Case Study: How Open data saved Canada \$3.2 Billion.” Eaves.ca. April 14, 2010. <http://eaves.ca/2010/04/14/case-study-open-data-and-the-public-purse/>

⁵⁸ Eaves, David. “Case Study: How Open data saved Canada \$3.2 Billion.” Eaves.ca. April 14, 2010. <http://eaves.ca/2010/04/14/case-study-open-data-and-the-public-purse/>

Lasby notes that the experience has enabled those working with the data to develop signatures for potential tax evasion, but at the time the tax evasion was missed because those overseeing the data simply were not looking for it. In addition, the community of T3010 data users was too small and technologically unsophisticated at the time to provide additional oversight or analysis. “This data is so variable that it’s hard to spot behavior that looks aberrant. It looked completely aberrant in the rear view mirror, once you know to look for it [but] at the time, I wasn’t constructing a lot of longitudinal data sets. Now it’s completely the norm for me to put together a ten-year slice. [We] didn’t have the capability back then, [or] the statistical or informational horsepower.”⁵⁹ He notes that those working with the data would be more likely to pick up anomalies now, but that the pool of analysts is still small, and not well integrated. Comparing it to the massed debugging capacity of the open source movement, Lasby notes, “What we have right now … is a lot of parallel play. We all build data sets, we all work with it in our isolated rural splendor. We do talk to each other, and share information both formally and informally, but it’s not systematic. We tend not to share data sets.”⁶⁰

More controversially, information from the T3010 was also used by the former Conservative Government to target certain organizations, notably environmental and human rights charities and left-leaning think tanks,⁶¹ for lengthy and costly tax audits of their political activities. Regulations since 2003 prohibit charitable organizations from spending more than 10 percent of all expenses on political activities.^{62, 63} The 2012 revision of the T3010 added a new question and schedule requiring charities to disclose whether they had participated in political activities, including making gifts intended for use for political activities, and whether they had

⁵⁹ GovLab interview with David Lasby, Director of Research, Imagine Canada, October 14, 2015.

⁶⁰ GovLab interview with David Lasby, Director of Research, Imagine Canada, October 14, 2015.

⁶¹ Beeby, Dean. “Academics call for moratorium on CRA political audit of think-tank.” *The Globe and Mail*. September 14, 2014. <http://www.theglobeandmail.com/news/politics/academics-calls-for-moratorium-on-cra-political-audit-of-think-tank/article20593498/>

⁶² <http://cponline.thecanadianpress.com/graphics/2014/cra-audit-charities/index.html>

⁶³ <http://www.cra-arc.gc.ca/chrts-gvng/chrts/plcy/cps/cps-022-eng.html#N1039E>

received funds from outside Canada for political activities. It also asked them to describe their political activities and how these related to their charitable purpose.⁶⁴

Between 2012 and 2016, the CRA, with \$13.4 million in special funding, was charged with reviewing political activities of charitable organizations that reported that they engaged in political activities or recorded expenses for political activities on their T3010.⁶⁵ In 2014 the CRA was auditing the political activities of 52 groups, many of whom had been critical of the Conservative Government's environmental and aid policies.⁶⁶ As of March 31, 2015, 21 audits had been completed, another 28 were in process, and a further 11 were expected to be carried out before the end of the review in 2016. These audits had resulted in six education letters, eight compliance agreements, five notices of intention to revoke charitable status, one voluntary revocation and one annulment.⁶⁷ Revocation of charitable status carried crippling consequences for organizations, stripping them of their power to issue tax receipts, and therefore cutting them off from an important source of funding. Charities complained that those audited were being drained of resources through the demands of compliance with the audit, while the threat of audits had created an "advocacy chill," where charities self-censored in order to avoid being audited.^{68,69} Furthermore, organizations making grants to other charities for political activity were required to report that as

64 http://www.cba.org/CBA/sections_charities/newsletters2013/cra.aspx

65 Blumberg, Mark. "Political Activity Audits and the CRA Program Update 2015." *GlobalPhilanthropy.ca*. April 24, 2015. http://www.globalphilanthropy.ca/blog/cra_program_update_2015_and_political_activity_audits

66 Beeby, Dean. "Timeline | Canada Revenue Agency's political activity audits of charities." *CBC News*. August 5, 2014. <http://www.cbc.ca/news/politics/canada-revenue-agency-s-political-activity-audits-of-charities-1.2728023>

67 http://www.cra-arc.gc.ca/chrts-gvng/chrts/bt/chrtsprgrm_pdt-2015-eng.html

68 "Preventing poverty' not a valid goal for tax purposes, CRA tells Oxfam Canada." *CBC News*. July 25, 2014. <http://www.cbc.ca/news/politics/preventing-poverty-not-a-valid-goal-for-tax-purposes-cra-tells-oxfam-canada-1.2717774>

69 "Study cites 'chill' from tax agency audits of charities' political activities." *CBC News*. July 10, 2014. <http://www.cbc.ca/news/politics/study-cites-chill-from-tax-agency-audits-of-charities-political-activities-1.2703182>

political activity on their own T3010, leading to a pronounced reluctance to make grants for political activity.⁷⁰

Lasby acknowledges that opening the data does make it available to people wanting to use it for political capital.⁷¹ Those responsible for instigating the audit process would, however, have had access to the data whether it was open or not. “I don’t know how you avoid [the political use of the data], and frankly I don’t know that you want to avoid that. You would hope that people are informed ... and that they make fair points, and that there are lots of voices out there. The challenge there is not, I think, about the data being open or closed, because all that it means is that that contradiction gets played out in a somewhat more public way.”⁷² As other commentators have also noted, the audits illustrate that the definitions of political activity may not be clear in practice, or even in law.⁷³

Challenges

Mismatch Between Expectations and Reality

Without exception, those interviewed about their work with the T3010 data set describe a gap between the apparent promise of the data and what it can actually deliver. Because the T3010 is intended primarily for compliance and administration rather than research, it does not provide all of the information that users would like. It can help give a better high-level view of the financial and human resources of the sector, but has little to say about the clients, networks and funders of charities.⁷⁴ Hilary Pearson, President of PFC, notes, in

⁷⁰ GovLab interview with Hilary Pearson, President, Philanthropic Foundations Canada, October 20, 2015.

⁷¹ GovLab interview with David Lasby, Director of Research, Imagine Canada, October 14, 2015.

⁷² GovLab interview with David Lasby, Director of Research, Imagine Canada, October 14, 2015.

⁷³ Tsao, Dora, et al. “Tax Audits of Environmental Groups: The Pressing Need for Law Reform.” Environmental Law Centre, University of Victoria. March 2015. http://www.elc.uvic.ca/wordpress/wp-content/uploads/2015/03/Tax-Audits-of-Environmental-Groups_2015Mar25.pdf

⁷⁴ Lenczner, Michael and Susan Phillips. “From Stories to Evidence: How Mining Data Can Promote Innovation in the Nonprofit Sector.” *Technology Innovation Management Review*. July 2012. <http://timreview.ca/article/575>

addition, that the T3010 shows where grants came from, but not what they were used for.⁷⁵

Significantly, the T3010 also says nothing at all about nonprofits without charitable status, which make up nearly half of the sector. Because nonprofit organizations without charitable status are covered by different legislation and legal requirements, they are not required to file an annual return.⁷⁶ As a result, relatively little is known about Canadian nonprofits, although they make up at least 44 percent of the nonprofit sector, according to the National Survey of Non-profit and Voluntary Organisations (NSNVO).⁷⁷ Brouard notes the absence of a directorate devoted to nonprofits without charitable status: “CRA isn’t interested in nonprofits [without charitable status] because they don’t pay tax.”⁷⁸

Noting that many of those using the T3010 data set tend to fall in love with its possibilities, and then out of love with it as they encounter its limitations, Lasby cautions that it needs to be treated as a foundation, rather than as a complete source. “If you treat it as a found, high-quality data set you’re pretty much inevitably going to be disappointed. The number of research questions you can answer [without gathering more data] is actually pretty limited. People [think], ‘Oh, there’s this tremendously rich source of data that’s going to answer all my questions!’ Well, no: You’ve got a tremendously rich source of data that can provide you with a framework to work from in going out and gathering more data to answer your questions.”⁷⁹

⁷⁵ GovLab interview with Hilary Pearson, President, Philanthropic Foundations Canada, October 20, 2015.

⁷⁶ GovLab interview with David Lasby, Director of Research, Imagine Canada, October 14, 2015.

⁷⁷ “Cornerstones of Community: Highlights of the National Survey of Nonprofit and Voluntary Organisations 2003 Revised.” Statistics Canada. 2005. http://www.imaginecanada.ca/sites/default/files/www/en/library/nsnvo/nsnvo_report_english.pdf

⁷⁸ GovLab interview with François Brouard, Founding Director, Sprott Centre for Social Enterprises, Carleton University, October 15, 2015.

⁷⁹ GovLab interview with David Lasby, Director of Research, Imagine Canada, October 14, 2015.

Data Quality: Transcription and Reporting Errors

One of the common complaints about the T3010 data is that it is currently manually entered from paper returns without double-keying, increasing the likelihood that transcription errors will occur and not be caught.^{80,81,82} Because data is submitted and keyed de novo each year, rather than being prepopulated, consistency of even the most basic data, such as an organization's name, is limited, according to Brouard.⁸³ The 2014 Canadian federal budget committed the government to reducing the administrative burden on charities by introducing electronic filing of T3010s and online applications for registrations for charitable status, at a cost of \$1 million in 2014–15 and \$0.5 million in 2015–16.⁸⁴ Users hope that this will increase the speed and accuracy of the data release.

Lasby believes reporting errors are a far more significant problem than transcription errors.⁸⁵ During the 1990s, it was discovered that some organizations were reporting their financials in thousands of dollars, without indicating the last three zeroes, rather than submitting exact figures. As a result, financial reporting on the T3010 for these organizations was out by three orders of magnitude.⁸⁶ Returns from 30 percent of universities studied were affected, resulting in a significant understatement of the size and wealth of the sector. The mistake was able to be identified by comparing the T3010 data against data published by the Association of Universities and Col-

80 GovLab interview with Hilary Pearson, President, Philanthropic Foundations Canada, October 20, 2015.

81 GovLab interview with David Lasby, Director of Research, Imagine Canada, October 14, 2015.

82 GovLab interview with François Brouard, Founding Director, Sprott Centre for Social Enterprises, Carleton University, October 15, 2015.

83 GovLab interview with François Brouard, Founding Director, Sprott Centre for Social Enterprises, Carleton University, October 15, 2015.

84 "Chapter 3.4: Supporting Families and Communities." Canadian Budget 2014. <http://www.budget.gc.ca/2014/docs/plan/ch3-4-eng.html>

85 GovLab interview with David Lasby, Director of Research, Imagine Canada, October 14, 2015.

86 GovLab interview with François Brouard, Founding Director, Sprott Centre for Social Enterprises, Carleton University, October 15, 2015.

leges of Canada, but similar mistakes in reporting by other charities may well have gone uncorrected.⁸⁷

Furthermore, little is done by CRA to verify the data supplied by charities, making it hard for users to assess data quality.⁸⁸ Brouard questions whether some of the data collected is used, or even monitored, by government. “As a charity, you will be penalized if you don’t file your T3010—but if you file it with garbage, no one seems to really care.”⁸⁹ Lasby points out that the T3010 does not have to be audited before filing. “You’ve got something that looks like a financial statement [based on] generally accepted accounting principles ... but’s that’s not what it is. It’s not a balance sheet, it’s a self-reported compliance document.”⁹⁰

The quality of the data set is also affected by the fact that the data is reported by the executive directors and board members of the organizations themselves, who may not have financial or legal expertise. Lasby reports that there have been considerable efforts to train those responsible for completing T3010s—“you name it, it’s been done”—and the form comes with a general guide and accompanying resources on the CRA website, but that ensuring the accuracy of data collection nonetheless remains a challenge. “The stark reality for the sector is that half of the organizations out there are 100 percent volunteer run. In a lot of cases, you’re asking someone who doesn’t necessarily have financial training to fill out a document that looks a lot like a financial document, and a lot of the time they don’t have legal training ... and they don’t fully understand ... the legal framework around charities.”⁹¹ This is complicated by what Phillips and Lenczner refer to as the lack of a “capacity for numer-

87 “Working Together: A Government of Canada/Voluntary Sector Joint Initiative – Report of the Joint Tables” Voluntary Sector Initiative. July 28, 1999. http://www.vsi-isbc.org/eng/knowledge/working_together/reports_wt_suppd.cfm

88 GovLab interview with Hilary Pearson, President, Philanthropic Foundations Canada, October 20, 2015.

89 GovLab interview with François Brouard, Founding Director, Sprott Centre for Social Enterprises, Carleton University, October 15, 2015.

90 GovLab interview with David Lasby, Director of Research, Imagine Canada, October 14, 2015.

91 GovLab interview with David Lasby, Director of Research, Imagine Canada, October 14, 2015.

acy”⁹² within the sector. Brouard reports that research units at Carleton had been contracted for projects to help improve compliance among smaller charities, whom CRA had recognized had particular difficulty filling in the T3010 correctly, and who lacked the financial resources to retain accountants to help them.⁹³

Definitional Challenges

Definitional problems are also common, creating confusion for charities and data users alike. Users note that terms used in the form—for fundraising methods⁹⁴ and categories of revenue sources,⁹⁵ to give two examples—are often different from those in common or academic use, and are not well understood by even the largest and most sophisticated of charitable organizations completing the document, or indeed by government departments. This leads to organizations with similar funding mixes having radically different T3010 reporting, according to Lasby.⁹⁶

Sometimes the problem is not the definition, but the absence of a definition. Pearson notes that the T3010 does not specify whether assets reported should be book or market value. “For us, that matters a lot. Market value ... can be much higher than book, and if we calculate our membership fee as a percentage of assets, if it’s book value then it’s going to be a lot less. But it’s also misleading in reporting in aggregate ways on the value of foundations. Are we underre-

92 Lenczner, Michael and Susan Phillips. “From Stories to Evidence: How Mining Data Can Promote Innovation in the Nonprofit Sector.” *Technology Innovation Management Review*. July 2012. <http://timreview.ca/article/575>

93 GovLab interview with François Brouard, Founding Director, Sprott Centre for Social Enterprises, Carleton University, October 15, 2015; See also Brouard, Francois, Leigh-ann Neilson and Anahit Armenakyan. “Small Charities: An Analysis of Change in Fundraising Activities.” SCSE-CSES Research Report. October 2012. <http://sprott.carleton.ca/research/wp-content/files/Paper2charitiesfundraisinganalysisFBLNAA20121031.pdf>

94 GovLab interview with François Brouard, Founding Director, Sprott Centre for Social Enterprises, Carleton University, October 15, 2015.

95 GovLab interview with David Lasby, Director of Research, Imagine Canada, October 14, 2015.

96 GovLab interview with David Lasby, Director of Research, Imagine Canada, October 14, 2015.

porting the whole field?”⁹⁷ These technical issues have made Pearson wary about depending on T3010 data in making statements about or assessments of the grant-making sector. “The only number that I can say, yes, I am certain that this is a correct number is the number of registered foundations, public and private, because CRA counts them and reports them. [But] even that misses how many of them are active and how many are just sitting on the books.”⁹⁸

Definitional confusion has implications not just for the quality of data collection, but also for organizational behavior. “The T3010 attempts to ask every charity what percentage of your budget, itemized, goes into administration [and] what percent goes into charitable activity, and an awful lot of charities out there are confused about that. In spite of CRA’s attempts to explain what is administrative and what is charitable, it’s not clear. Just about the only thing that does seem to be clear is that the fee you pay to your auditor for auditing your financial statements is an administrative expense. An awful lot of foundations still think that they have to put all their salaries into the administrative expense box, and therefore they don’t hire people ... because they don’t want to have any overhead.”⁹⁹

Shifting Government Priorities

The T3010 has also changed over time in response to changes in government priorities. Having requested archival data since 2000, Brouard found himself frustrated by the frequent changes in the form to reflect changes in tax policy, which made it difficult, and in some cases impossible, to compare data over time. Some of the changes seem arbitrary and designed to confuse rather than enlighten. Brouard cites a change in the wording of a question that asked one year whether an organization was active, and whether the organization was inactive in the following year’s form. “So a yes one year means a no in another year.”¹⁰⁰ Changes to the form are also

⁹⁷ GovLab interview with Hilary Pearson, President, Philanthropic Foundations Canada, October 20, 2015.

⁹⁸ GovLab interview with Hilary Pearson, President, Philanthropic Foundations Canada, October 20, 2015.

⁹⁹ GovLab interview with Hilary Pearson, President, Philanthropic Foundations Canada, October 20, 2015.

¹⁰⁰ GovLab interview with François Brouard, Founding Director, Sprott Centre for Social Enterprises, Carleton University, October 15, 2015.

accompanied by a temporary drop in data quality, according to Lasby: “Every single time we change the form, data quality suffers, and then as that section of the form becomes more and more familiar … data quality gets better.”¹⁰¹ Brouard suggests that assistance from Statistics Canada could help to improve the design of the T3010, but that communication between Statistics Canada and the Canada Revenue Agency is poor, although efforts have been made to improve them.¹⁰²

Those seeking to use the data must therefore either accept its imperfections or find ways to eliminate erroneous or questionable data. Lasby notes that academic and research use requires substantial time and effort to clean the data and ensure data from the subset of charities under study is coherent, free from obvious error, and meets necessary criteria.¹⁰³ Imagine Canada and Ajah have found ways to circumvent these problems, at least in part, by combining the T3010 data set with data from other sources. According to Lasby, however, there are few tools to assist those attempting to work with the T3010 data, and becoming conversant with it and skilled in its use is a time- and labor-intensive exercise in self-teaching. “The principal [question is], whether you have the systems in place to work with the data? Do you have the technological infrastructure to make use of it? … Only a minority of people have the skills required to work with it directly, and the technology does not yet exist, or is not yet widespread, to allow folks who lack those technical skills to work with it.”¹⁰⁴ Brouard puts it more succinctly: “It’s one thing to have the data; it’s another to be able to work with it easily.”¹⁰⁵

¹⁰¹ GovLab interview with David Lasby, Director of Research, Imagine Canada, October 14, 2015.

¹⁰² GovLab interview with François Brouard, Founding Director, Sprott Centre for Social Enterprises, Carleton University, October 15, 2015.

¹⁰³ GovLab interview with David Lasby, Director of Research, Imagine Canada, October 14, 2015.

¹⁰⁴ GovLab interview with David Lasby, Director of Research, Imagine Canada, October 14, 2015.

¹⁰⁵ GovLab interview with François Brouard, Founding Director, Sprott Centre for Social Enterprises, Carleton University, October 15, 2015.

Looking Forward

Improving the Usability of the Data

Users are equivocal about the usefulness of the current T3010 data release on the government's open data portal. Lasby says that some of the ways the data has been processed for release on the open data portal do not make sense to him, and make the data harder to work with. He prefers to get data updates on CD-ROM, which are more frequent, with disc releases taking place roughly once a month, and containing a couple of years' worth of both newly released data and updates and corrections to previously released data.¹⁰⁶ CRA has expressed interest in creating an official research database, which Elson believes would be a boon for the comparability of research,¹⁰⁷ but Lasby is concerned that accuracy would suffer without the continual updating of the CD-ROM releases.¹⁰⁸

Some parts of the data set remain largely unexplored territory. Lenczner notes that no one has made much use of the newly complete records of charitable gifts, since the entire list of gifts has only been made available since 2013. Overall, he believes that there is still a lot of value to be extracted from the entire data set. "It takes time for innovation to happen. It takes the market knowing what's available and people knowing how to exploit what's available."¹⁰⁹ Lasby believes the use of the data is increasing as its availability has become better known. "One of the changes I've observed over the past couple of years is more and more journalists working with it, as a completely standard thing ... and they work with it as a data set, not [by] going in and looking at a single organization."¹¹⁰

¹⁰⁶ GovLab interview with David Lasby, Director of Research, Imagine Canada, October 14, 2015.

¹⁰⁷ GovLab interview with Peter Elson, Community Development Researcher, Mount Royal University and University of Victoria, October 21, 2015.

¹⁰⁸ GovLab interview with David Lasby, Director of Research, Imagine Canada, October 14, 2015.

¹⁰⁹ GovLab interview with Michael Lenczner, Chief Executive Officer, Ajah, September 23, 2015.

¹¹⁰ GovLab interview with David Lasby, Director of Research, Imagine Canada, October 14, 2015.

Both Pearson and Brouard believe that separating the regulation of charities from tax collection would improve data quality and coverage. Pearson suggests that the solution may ultimately lie in separating the collection of charities data from CRA. “I think the fundamental issue ... is one that the U.K. tackled some time ago: Should the regulation of the sector be handled out of the [tax] department? The U.K. created the Charities Commission for very good reason. An independent organization is able to regulate and collect data on charities outside of the requirements of revenue collection. I think that’s really fundamental, but that’s not an issue for CRA: It’s a political issue, and a policy issue.”

IT Support for Nonprofits

There are several indications that Canada will be furthering its leadership in this space.¹¹¹ The already launched Non-Profit Organization Risk Identification Project Report should increase public knowledge about the significant number of nonprofits that do not have charitable status.¹¹² Additionally, the announced IT Modernization Charities project—which is aimed at modernizing the CRA’s IT systems so that it can provide new electronic service options for charities—contains several measures to improve the quality and availability of that data. Those initiatives, as well as civil initiatives such as the Community Knowledge Exchange and Ajah’s own PoweredbyData, give Canada the opportunity to continue to develop its expertise in this area.¹¹³

Nonprofit Data Collaboratives

As openness in the Canadian nonprofit sector continues to take hold, actors in the field should look to not only move toward compliance and transparency, but consider how data can enable greater collaboration among organizations. As David Eaves argued prior to the release of nonprofit data in 2013:

¹¹¹ Lenczner, Michael. “Canada—a World Leader in Non-profit Data.” Government of Canada. July 4, 2014. <http://open.canada.ca/en/blog/ajah-and-fundtracker>

¹¹² <http://www.cra-arc.gc.ca/tx/nnpfnt/npprft-eng.html>

¹¹³ Lenczner, Michael. “Canada—a World Leader in Non-profit Data.” Government of Canada. July 4, 2014. <http://open.canada.ca/en/blog/ajah-and-fundtracker>

“[T]here is a tremendous opportunity in the nonprofit space around not just using data, but also sharing data, to better understand some of the world’s toughest challenges. I’m also left acknowledging that even where data usage is strong—such as in the environmental community—few stakeholders in these sectors see open data as something relevant or meaningful to their organizations and their strategies. I’m not sure how to change that, but it seems like a huge opportunity.”¹¹⁴

The Community Knowledge Exchange Summit, for instance, demonstrates the interest in and potential impacts of greater coordination between actors in the Canadian nonprofit space. One of the key aspects of the summit was field building: “Creating and holding a space for community organizations, academic and research institutions, governments, funders and grant-makers, and engaged citizens to come together in the spirit of collaboration and openness.”¹¹⁵

It has been evident in many cases examined in this report that data collaboratives bringing together actors within and across sectors can amplify the impacts of open data. Now that the proverbial cat is out of the bag, and open data is the default in the nonprofit sector, organizations can start to consider how they can turn this openness into new ways to solve problems together.

¹¹⁴ Eaves, David. “Nonprofits Should Share Their Data, Too.” *TechPresident*. January 24, 2013. <http://techpresident.com/news/23412/nonprofits-should-share-their-data-too>

¹¹⁵ Rose, Lee. “Reflections on CKX—The Community Knowledge Exchange Summit.” Toronto Foundation. December 2, 2014. <https://torontofoundation.ca/news/reflections-ckx-%E2%80%93-community-knowledge-exchange-summit>

PART III

Case Studies: Empowering Citizens

In a number of cases, open data is empowering citizens to take control of their lives and demand change. The new ways of communicating and accessing information exemplified by open data are enabling more informed decision making and new forms of social mobilization.

The case studies in this section examine:

- An improved capacity to track school performance and demand improvements in Mexico and Tanzania.
- More informed decision-making regarding health care by citizens in Uruguay;
- Crowdsourced confirmation of election results in Indonesia, leading to greater faith and confidence in the political system; and
- Potential citizen-driven improvements to transparency and accountability among public- and private-sector entities in Kenya.

CHAPTER 8

Open Education Information in Tanzania

A Tale of Two Dashboards

*By Juliet McMurren, Stefaan Verhulst, Andrew Young
and David Sangokoya*

Summary: Low national examination pass rates in 2012 caused a public outcry in Tanzania, but the public's understanding of the broader context (and thus ability to demand accountability) was limited by a lack of information about the country's education sector. Two recently established portals are trying to remedy that situation, providing the public with more data on examination pass rates and other information related to schools. The first, the Education Open Data Dashboard (educationdashboard.org), is a project established by the Tanzania Open Data Initiative, a government program supported by the World Bank and the United Kingdom Department for International Development (DFID) to support open data publication, accessibility and use. The second, Shule (shule.info), was spearheaded by Arnold Minde, a programmer, entrepreneur and open data enthusiast who has developed a number of technologies and businesses focused on catalyzing social change in Tanzania. Despite the challenges posed by Tanzania's low Internet penetration rates, these sites are slowly changing the way citizens access information and make decisions. More generally, these projects are encouraging citizens to demand greater accountability from their school system and public officials.

Dimension of Impact: Empowering Citizens—Social Mobilization

Key Takeaways:

- Dashboards like Shule and the Education Open Data Dashboard are cheaply repurposable tools that can be quickly and readily implemented by local programmers. Both dashboards were built and launched in two to three weeks, although the data set used to build Shule had been scraped in increments over several years.
- In countries with low Internet penetration, raising awareness among infomediaries, particularly in the media, about the availability and use of open data can be critical in bridging the access gap and successfully engaging the public. Civil society organizations, both local and international, can play a vital role in spurring the use of open data by raising awareness among infomediaries.
- Shule and the Education Open Data Dashboard illustrate the benefits of coordination and collaboration between government open data initiatives and those developed by civil society and private organizations. Shule was developed by a single dedicated programmer with little outside help; the Education Open Data Dashboard was developed by a local developer in coordination with government data practitioners, NGOs and programmers (including the man behind Shule). Taken together, the information these sites provide about Tanzanian education is richer and more interesting than either site alone.
- The collaboration between open data champions like Arnold Minde and civil society organizations such as Twaweza can radically increase the value and impact of open data initiatives.

Context and Background

Since 2012, education in Tanzania has become the subject of significant public discontent and controversy. That year, six out of every 10 Tanzanian students failed the standardized national secondary-level examination, resulting in significant public discontent and demand

for reforms.¹ The poor results were in many ways the outcome of recent changes to the Tanzanian education system. In 2002, tuition fees for government primary schools were eliminated in an effort to raise school enrolment and literacy rates. The move triggered rapid growth in net primary enrollment, from 66 percent in 2001 to 90 percent in 2004,² but this growth was not matched by a proportional increase in school funding, hence the education sector's problems a few years later.

As the Tanzanian school system strained under the burden of the additional enrollments, examination pass rates among the 30 percent of secondary-aged children enrolled in school³ began to decline. After the particularly bad set of results in 2012, the government introduced changes to the grading system⁴ that appeared to raise the pass rate in 2013 and 2014.⁵ However, the root causes of the nation's education problems—inadequately funded and supplied schools, a shortage of trained teachers,⁶ limited teacher training and professional development, delays in the payment of teachers' salaries,⁷ and stubborn regional, economic and social inequalities—remained unaddressed.

At the same time, information about the state of public education was not easy to come by, making it hard for citizens to understand

1 Kimboy, Frank. "High pass rate greeted as good as well as bad news." *The Citizen*. July 23, 2014. <http://www.thecitizen.co.tz/magazine/political-reforms/High-pass-rate-greeted-as-good/-/1843776/2394162/-/umh9xl/-/index.html>

2 "World Data on Education (Revised Version)." UNESCO. August 2010. http://www.ibe.unesco.org/fileadmin/user_upload/Publications/WDE/2010/pdf-versions/United_Republic_of_Tanzania.pdf

3 "Tanzania: World Bank Development Indicators." Accessed August 28, 2015. <http://data.bank.worldbank.org/data//reports.aspx?source=2&country=TZA&series=&period=>

4 Magolanga, Elisha. "No more Div. Zero as govt guts national grading system." *The Citizen*. November 1, 2013. <http://www.thecitizen.co.tz/News/Govt-in-major-change-of-national-exam-grading/-/1840392/2055404/-/3kae76z/-/index.html>

5 Kimboy, Frank. "High pass rate greeted as good as well as bad news." *The Citizen*. July 23, 2014. <http://www.thecitizen.co.tz/magazine/political-reforms/High-pass-rate-greeted-as-good/-/1843776/2394162/-/umh9xl/-/index.html>

6 The Citizen Reporter. "Teachers shortage hurting Tanzania." *The Citizen*. October 14, 2014. <http://www.thecitizen.co.tz/News/Teachers-shortage-hurting-Tanzania/-/1840340/2485582/-/tpeqyi/-/index.html>

7 Qorro, Edward. "Education Policy: What lies ahead for teachers." *The Citizen*. February 24, 2015. <http://www.thecitizen.co.tz/magazine/success/-/1843788/2633842/-/item/3/-/7gk0yk/-/index.html>

the true state of the education sector and demand accountability from government officials. The Tanzanian media is considered only partly free by Freedom House,⁸ and the country was ranked 75th out of 180 countries in the 2015 World Press Freedom Index.⁹ Although several Access to Information bills have gone before the Tanzanian Parliament, none has yet been enacted, while other legislation, including the country's defamation law, constrains the media's capacity to function critically and independently.

In addition, there exists a noticeable lack of independent voices in the Tanzanian media. While media ownership is transparent, it remains concentrated among a few proprietors. All four radio stations with national reach are regarded as favoring the ruling party.¹⁰ Media outlets favorable to the opposition reportedly have government advertising contracts withheld.¹¹ Consequently, when stories about the state of education do make it to press, they tend to favor the official version of events, and often lack balance or context.

Citizens were for the most part unable to turn to the Internet or open data as substitutes for the information they needed. Open data use in Tanzania is still in its infancy, although recent initiatives such as the Africa Open Data Conference, convened by the president of Tanzania, are encouraging development.¹² The Open Data Barometer places Tanzania in the "capacity constrained" cluster of countries whose open data initiatives are challenged by limits in government, civil society or private sector capacity, Internet penetration, and data collection and management.¹³ Tanzania joined the Open Government Partnership (OGP) Initiative in September 2011. The second phase of its OGP action plan commits the government to establishing an open data portal (<http://www.opendata.go.tz>) that would release key data sets in the education, health and water sectors in

8 <https://freedomhouse.org/country/tanzania#.VaQZFvIViko>

9 <http://index.rsf.org/#!/index-details/TZA>

10 In 2010, the African Media Barometer did report, however, that the state-run Radio Tanzania has demonstrated more balanced views within recent years

11 Tanzania Country Report—Freedom of the Press 2011. Freedom House. Accessed August 28, 2015. https://freedomhouse.org/report/freedom-press/2011/tanzania#.VaARQ_IViKp

12 <http://www.africaopenadata.net>

13 <http://www.opendatabarometer.org/report/analysis/index.html>

machine-readable form.¹⁴ Publicly launched in September 2015, the portal at that date had 81 data sets available for download.¹⁵

“Data is frightening for many people, so raw data is going to appeal to a vanishing few. Open data needs to be open plus curated plus chewed plus digested to appeal to most people, including policy-makers.”

— Aidan Eyakuze, Twaweza

Product Description and Inception

In 2013, the National Examinations Council of Tanzania (NECTA) rolled out a dashboard¹⁶ offering data downloads, searches, and visualizations of primary and secondary examination results by district, with statistics on annual and average pass rates, national ranking and changes in pass rates since 2011.

The launch of its Big Results Now (BRN) program in 2012 committed the Government of Tanzania to publishing key data sets, powered by its parallel open data and OGP agendas, for public consumption. To support this, Tanzania Open Data Initiative launched the prototype Education Open Data Dashboard in 2015 (available at educationashboard.org).¹⁷

The Education Open Data Dashboard offers indicators such as pupil-teacher ratios, regional and district rankings, and improvement rankings over time, all of which are navigated via a clickable map and drop-down menu of schools. Although the site represents a significant advance by bringing together data from several sources, the dashboard’s remit limits its scope to BRN data, which does not include pass rates before 2012, average pass rates over time, or pass rates by gender or region. The Education Open Data Dashboard also offers no analysis of its data visualizations.

14 <http://www.opengovpartnership.org/country/tanzania/action-plan>

15 <http://opendata.go.tz/dataset>

16 Available at <http://www.necta.go.tz/opendata/>, and subsequently updated at <http://www.necta.go.tz/opendata/brn/>

17 <http://theEducationOpenDataDashboard.org/#/>

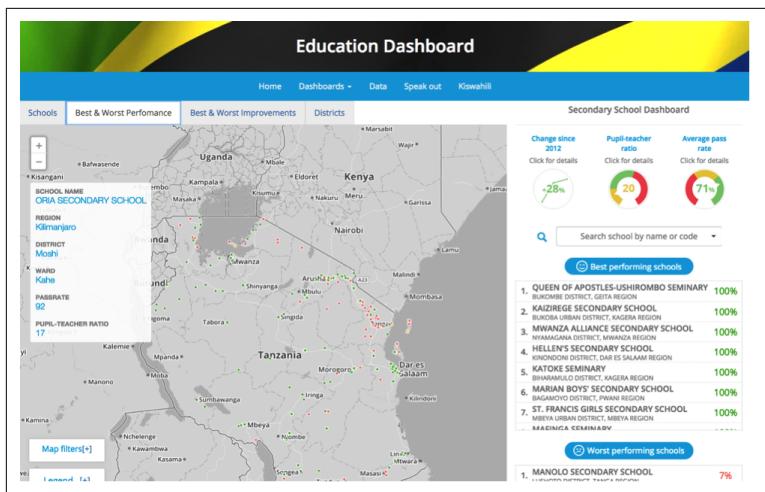


Figure 8-1. Mapping of school performance on the educationashboard.org

Many of these gaps have been filled in recent years by a pioneering project named Shule (shule.info). Shule is the brainchild of the Tanzanian programmer Arnold Minde. It was released shortly after NECTA's own original dashboard, and was conceived when Minde, an entrepreneur and open data enthusiast, became aware that NECTA had been publishing individual exam results online since 2004. Although this data was available in isolated reports and websites, usually intended for individual students, it had never been made fully open in searchable and machine-readable format for citizens at large.¹⁸

Intrigued by the possibilities, Minde began scraping, cleaning and consolidating this data from the examination results as they were released each year. It wasn't until 2012, however, when poor examination pass rates prompted nationwide soul-searching, that Minde began working on the project in earnest. At that point, he realized the potential value of a single, readily usable source of national examination data.¹⁹ Such data needed to be online, he concluded, so that citizens could see that the low pass rate of 2012 was not a new phenomenon, but part of a trend of worsening results over the past

18 <http://www.shule.info/about>

19 <http://www.shule.info/about>

six to seven years. Minde had previously been involved in data visualization through his work for the Tanzanian development policy think tank **REPOA** (formerly Research on Poverty Alleviation); that work convinced him of the power of data visualization to communicate data trends and linkages, and helped shape the development of Shule.²⁰



Figure 8-2. Education results by region on Shule.info

Minde states that his aim in developing Shule was to make information available to everyone with a potential interest in it: parents choosing schools for their children; students looking up examination results; policymakers seeking to track educational trends and progress; and journalists wanting to improve their educational coverage. The original pilot of the site was ready in three weeks, but was revised and improved after Twaweza, a Tanzanian civil society organization promoting effective and transparent governance, became interested in Minde's project. Twaweza's advice prompted Minde to modify the design of his site for increased appeal and usability, and to double the number of indicators.²¹

As of late 2015, the site presents data for Form 4 examination results from 2004 to 2013 at candidate, school, regional and national levels, and offers data visualizations of results by regions and gender, aver-

²⁰ GovLab Interview with Arnold Minde, Founder, Shule, July 9, 2015.

²¹ GovLab Interview with Arnold Minde, Founder, Shule, July 9, 2015.

age performance over time, the number of candidates in each grading division over time, and the impact of the government's controversial revision of the 2012 results. All data used to build the site is available for download. In addition, and in contrast to NECTA's dashboards, Shule offers commentary on its data visualizations, making it easier for users to understand the significance of the data they are accessing.

Impact

Tanzania is a country with low Internet penetration rates (4.9 percent in 2014, according to ITU, the United Nations specialized agency for information and communication technologies²²), and a general lack of familiarity with the concept and potential of open data. As such, the impact of Shule and the Education Open Data Dashboard has been more modest and difficult to assess than it might have been in another country. Nonetheless, early signs of impact by these two education data dashboards are discernible. Impact can be gauged in three ways: engagement and use by both citizens and infomediaries; data quality and diversity; and spillover effects on other open data projects.

Intended Beneficiaries

- | | |
|----------------------------------|---|
| Parents | <ul style="list-style-type: none">• Able to visualize and compare school performance before choosing a school for their children• Able to hold schools to account for performance |
| Students | <ul style="list-style-type: none">• Can access examination results using candidate number and examination center |
| Policymakers and planners | <ul style="list-style-type: none">• Data archives and visualizations allow the tracking of trends in education over time and by region• Data visualization encourages better allocation of resources according to need |

²² ICT Statistics. ITU. Accessed August 28, 2015. <http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx>

Journalists

- Data visualizations and analysis offer context and prompt questions to improve depth of coverage of education stories
- Able to translate and digest data for general public, improving penetration

Engagement and use

Since Shule went live in June 2013, the site has averaged around 1,500 visits per month, according to Arnold Minde.²³ Feedback directly on the site and through Twaweza suggests that visitors fall into two categories. The first consists of data sophisticates, typically programmers or employees of civil society organizations, who are already aware of the potential of open data to inform decision-making, and visit the site to research education in Tanzania and better understand the overall educational context. These visitors may have become aware of the site through Twaweza, REPOA's network of civil society partners, or the emerging open data community in Dar es Salaam.

The second category of site visitors consists of former students making use of the site's archive of examination results to look up their scores.²⁴ These students may not initially be interested in or even aware of open data, but they are nonetheless exposed to Shule's visualizations and other tools when they access the site.

Engaging the ordinary Tanzanian families Minde had originally hoped to reach has been more challenging. Low rates of Internet penetration and a lack of experience using the Internet have suppressed the amount of casual traffic received through search engines. Minde says he fears that average Tanzanians don't have much interest yet in looking at data visualizations, preferring to get their information predigested by the media. "I don't see people asking the real questions," says Minde. "I don't see discussions around the issues, even among people I know."²⁵ Aidan Eyakuze, Executive Director of Twaweza, believes both the public and policymakers are looking for the insight contained in the data, not the data itself. "Data is frightening for many people, so raw data is going to appeal

²³ GovLab Interview with Arnold Minde, Founder, Shule. July 9, 2015.

²⁴ GovLab Interview with Arnold Minde, Founder, Shule. July 9, 2015.

²⁵ GovLab Interview with Arnold Minde, Founder, Shule. July 9, 2015.

to a vanishing few. Open data needs to be open plus curated plus chewed plus digested to appeal to most people, including policy-makers.²⁶ Few in the media, however, have the knowledge and skills to digest Shule’s data offerings, despite initiatives like the Data Bootcamp, which was designed to introduce members of the Tanzanian media to open data.²⁷

The Education Open Data Dashboard’s use is similarly constrained by Tanzania’s low rate of Internet use. Nonetheless, the site’s developers point out that Tanzanians don’t necessarily need Internet access to benefit from the information stored on the site. Members of civil society organizations, for example, including Tanzania’s active parent-teacher organizations, can potentially act as infomediaries, printing out information about school performance to share on a community notice board or at meetings.²⁸ For its part, the government of Tanzania has recognized the potential impact of this tool. Using the open-source code of educationdashboard.org, a second iteration of the Education Open Data Dashboard—elimu.takwimu.org—will soon be integrated directly into the Tanzanian open data portal to help build demand for additional data set availability and use.

Data Quality and Diversity

The combination of the Education Open Data Dashboard and Shule increases the diversity and thus the usefulness of available data on education in Tanzania. Taken together, the information they provide is richer and more interesting than either site would be on its own. The Education Open Data Dashboard offers indicators such as pupil-teacher ratios, regional and district rankings, and improvement rankings over time, all of which are navigated via a clickable map and drop-down menu of schools. Shule captures a much longer span of data, with examination results going back to 2004. In addition to results by gender, Shule offers average performance over time, and looks at the breakdown of candidate numbers per grading

²⁶ GovLab Interview with Aidan Eyakuze, Executive Director, Twaweza. July 14, 2015.

²⁷ Bauer, Michael. “The Data Bootcamp in Tanzania.” *Open Knowledge Blog*. October 25, 2012. <http://blog.okfn.org/2012/10/25/the-data-bootcamp-in-tanzania/>

²⁸ Interview with Samhir Vasdev, ICT Sector Unit, World Bank Group and Verena Luise Knipper, Senior Governance Specialist, World Bank June 30, 2015.

division over time. It also models the effect of the 2012 grading revision to examine how it changed candidate pass rates.

Although based on government data, the data set used to build Shule is not completely identical to that used for the government dashboard because of differences in their methods of data collection. Perhaps as a result, Shule's figures can depart in significant ways from the government version. For example, NECTA has traditionally published an annual list of the 10 government and secondary schools with the highest examination results. In 2012, Minde reports that NECTA's official list contained a number of government schools, but Shule's analysis showed that all 10 of the top-performing schools were private.

For the developers of the Education Open Data Dashboard, one of the more surprising discoveries was the dashboard itself became a potent tool for teaching and understanding data management, publication, cleaning and licensing. Regional officials and head teachers were excited by finding their school or region in the dashboard, and by seeing what the data they submitted was creating, and this excitement encouraged increased understanding of data management. This suggests that the novelty of open data and data visualization, as manifested in a dashboard, can be a valuable entry point in building data management capacity.²⁹

Impact on Other Data Projects

As the developers of the latest version of the Education Open Data Dashboard have indicated, Shule forms part of a nascent data ecosystem of which they were very much aware during the development and refinement of their own site. The existence of such independent projects validated both the demand for the kinds of open data portal they were building and provided evidence that the local technical and other capacity existed to build it.³⁰ Their own dashboard was, in turn, a powerful tool in demonstrating the potential and uses of open data to a nontechnical audience, particularly among policymakers. In addition, the data visualizations and link-

²⁹ Interview with Samhir Vasdev, ICT Sector Unit, World Bank Group and Verena Luise Knippel, Senior Governance Specialist, World Bank June 30, 2015.

³⁰ Interview with Samhir Vasdev, ICT Sector Unit, World Bank Group and Verena Luise Knippel, Senior Governance Specialist, World Bank June 30, 2015.

ages it made possible ignited interest in, and impetus for, the development of dashboards in other sectors, such as water and health, both priorities for the BRN initiative.³¹

Outside Tanzania, Shule has been used as an example of the potential and use of open data in education. Both it and the Education Open Data Dashboard illustrate the power of a deceptively simple tool, one that can be cheaply and easily produced in a matter of a few weeks by local programmers, then refined through user feedback. As one of the developers of the Education Open Data Dashboard put it: “Get a minimum viable product out there ... and provoke a response.”³²

Challenges

Shule and the Education Open Data Dashboard are both young projects, launched into a society and country that is only beginning to grasp the potential of open data. Their most influential years could well lie ahead. However, if they are to grow and disseminate further among the population, they will need to overcome some challenges. This section examines two of the most important challenges they are likely to face.

Lack of Internet Penetration and Use

Perhaps the most important challenge stems from Tanzania’s low Internet penetration and usage rates. The two dashboards take as a given that providing information to target audiences will lead to improved conditions on the ground. It will be difficult to prove this, particularly in rural areas, where Internet penetration rates are estimated to be about a quarter of those in urban areas.³³ This clearly limits the reach of education-related data, and open data more broadly. Furthermore, of the 4.9 percent of Tanzanians who used the Internet in 2014, the great majority do so only by mobile phone;

31 “Social Justice.” Open Data for Tanzania. Accessed August 28, 2015. <http://tanzania.opendataforafrica.org/igcpumb/social-justice>

32 Interview with Samhir Vasdev, ICT Sector Unit, World Bank Group and Verena Luise Knippel, Senior Governance Specialist, World Bank June 30, 2015.

33 “Tanzania: Old Media, New Media.” *AfricaFocus*. April 5, 2011. <http://www.africafocus.org/docs11/tan1104.php>

only 0.2 percent of Tanzanians had a fixed broadband subscription.³⁴ In order to appeal more widely, any open data site clearly needs to consider launching a mobile application to appeal to “the retail user of data sitting in a bus shelter with a mobile phone.”³⁵

The barriers may be even higher, in fact, when it comes to using data and technology as instruments of change. Minde notes that, in general, the Tanzanian public is deeply unfamiliar with the potential of the Internet, and perhaps not yet inclined to trust it. He adds that Tanzanians have yet to embrace or commit to digital solutions for the problems of everyday life, whether complex or mundane. As an example, he cites the difficulty he experienced in convincing bus operators to adopt an earlier application he developed that allowed passengers to purchase tickets by phone. “It will only take one [company], and then people will see the benefit,” he says. “But first you have to convince the one.”³⁶

Meanwhile, as Internet penetration slowly expands, civil society organizations like parent-teacher organizations or NGOs have an important role to play as infomediaries who can share information with parents directly or through low-tech solutions like community notice boards. A recent study notes, for instance, that ultra low-tech solutions like posting printouts of information drawn from open data dashboards on school or community noticeboards can be effective in getting information to the people who can use it.³⁷

Fragmentation

Given low Internet penetration rates, the existence of two separate dashboards for education information could also prove confusing to parents, and limit the effectiveness of both platforms. Indeed, the greatest impact on education in Tanzania could very well come from integrating the two platforms and cooperatively advancing a single project, rather than providing a limited user base with two separate entry points for accessing essentially the same information. Moves

³⁴ ICT Statistics. ITU. Accessed August 28, 2015. <http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx>

³⁵ GovLab Interview with Aidan Eyakuze, Executive Director, Twaweza. July 14, 2015.

³⁶ GovLab Interview with Arnold Minde, Founder, Shule. July 9, 2015.

³⁷ Interview with Samhir Vasdev, ICT Sector Unit, World Bank Group and Verena Luise Knipper, Senior Governance Specialist, World Bank June 30, 2015.

toward greater coordination have been made, however, including notably Minde's involvement in development strategy workshops for the Education Open Data Dashboard,³⁸ where government education specialists worked with Minde and with their own data to better understand opportunities and techniques to prepare and publish open data.

Looking Forward

Both the platforms under study here have begun a process of transformation that is likely to be slow and gradual. Ultimately, if they are successful in overcoming their various challenges and scaling up usage, their impact is likely to be felt far beyond the education sector.

In the near term, the individuals and groups behind the platforms have two main strategies for extending their reach.

Improving User Engagement through Expanded Functionality

One of the key goals in coming years, according to Minde, is to engage more average families and schools as users. Ordinary citizens want to know how schools are performing, he argues; they have a stake in the education data. Schools, especially private ones, also have a stake—they want to understand how their competitors are doing, and the data provided on Shule can help benchmark their own performance.³⁹

In order to expand his user base, Minde plans to include a wider variety of data and new services. For example, he hopes to expand the site in the future to include Form 4 examination results, as well as the Form 2 and 6 results already included.⁴⁰ In addition, he would also like to increase the range of information about schools, for example mapping their locations and showing contact details, performance over time and sample student profiles. He is also considering offering an online application system that would give schools a

³⁸ Interview with Samhir Vasdev, ICT Sector Unit, World Bank Group and Verena Luise Knippel, Senior Governance Specialist, World Bank June 30, 2015.

³⁹ GovLab Interview with Arnold Minde, Founder, Shule. July 9, 2015.

⁴⁰ <http://www.shule.info/about>

better and more efficient way of connecting with potential students.

⁴¹

Education of Infomediaries

As noted, infomediaries and civil society groups have a key role to play in overcoming the challenge posed by low Internet usage. Such groups can help disseminate insights gleaned from open data among citizens who would otherwise not have access to the data.⁴²

Some efforts have already taken place to involve civil society. For instance, in 2012, in an effort to encourage interest and build skills among coders and the media, the World Bank Institute and the Africa Media Initiative combined to offer the Data Bootcamp in Dar es Salaam.⁴³ A similar initiative was offered by Twaweza in 2013, and community groups such as the Open Knowledge Foundation Network TZ have attempted to promote open data meetups in Dar es Salaam. In 2015, the Africa Open Data Conference was organized by a group that included both civil society organizations such as Code for Africa and the Open Knowledge Foundation, and the World Bank. These collaborations, and smaller informal meetups, provide an opportunity to bring together atypical combinations of stakeholders to meet and work with data.⁴⁴

Such efforts are likely to become more frequent in the coming years—and, as they do, the impact of open data on education and many other sectors will begin to be felt more widely. Greater impact will also increase awareness, thus creating a virtuous cycle of change and empowerment. For the moment, significant challenges remain. But the two projects studied here indicate the clear potential of open data to, application by application, sector by sector, chip away at those challenges.

⁴¹ GovLab Interview with Arnold Minde, Founder, Shule. July 9, 2015.

⁴² It is worth noting, however, that even among such groups, awareness of the potential of open data remains at best nascent. Like the public at large, civil society groups also need to be trained to analyze and visualize data.

⁴³ Bauer, Michael. “The Data Bootcamp in Tanzania.” *Open Knowledge Blog*. October 25, 2012. <http://blog.okfn.org/2012/10/25/the-data-bootcamp-in-tanzania/>

⁴⁴ Interview with Samhir Vasdev, ICT Sector Unit, World Bank Group and Verena Luise Knipper, Senior Governance Specialist, World Bank June 30, 2015.

CHAPTER 9

Kenya's Open Duka

Open Data for Transactional Transparency

By Andrew Young and Stefaan Verhulst

Summary: Open Duka is a platform developed by the civil society organization the Open Institute that aims to address issues of opacity in governance in the private and public sectors, promoting corporate accountability and transparency. As a case study, it exemplifies the challenge for open data initiatives to generate sufficient awareness and use necessary to achieve impact. Open Duka provides citizens, journalists and civic activists with a tool that gives insight into the relationships, connections (and, to some extent, the dynamics) of those in and around the public arena. The platform has the ability to create and visualize relationships between different entities ranging from organizations, individuals, tenders and contracts awarded. While a diversity of setbacks have led to limited impacts to date, the platform seeks to act as a core data set to enable these stakeholders to build third-party public transparency and public accountability apps or services, by allowing them to cross reference the Open Duka data against other data sets. Some of the data that can be found on Open Duka includes information published in the *Kenya Gazette*, a weekly publication released by the Kenyan government that publishes notices of new legislation, notices required to be published by law or policy and announcements for general public information.

Dimension of Impact: Empowering Citizens—Informed Decision-making

Key Takeaways:

- Open data initiatives can be resource- and time-intensive. Cleaning data, building partnerships, tech development and other important steps can require more effort than originally planned, leading to long stretches of time before any real impacts are felt.
- Opening data is not always sufficient. Citizen engagement, outreach and training efforts—regardless of the tech capacity or data literacy of the targeted community of users—can be important drivers of use for open data platforms.
- Political and cultural changes can throw open data initiatives into upheaval, even if they do not, on the surface, seem to directly affect the project at hand. Developing an open data *platform, application or service* often requires a supportive open data *ecosystem* to reach its full potential.

Context and Background

Corruption in Kenya

Corruption in Kenya is a major and by many indications worsening problem. In 2014, Transparency International ranked the country 145 (out of 175) in its annual Corruption Perceptions Index, down from 136 two years earlier.¹ Recently, an official audit found that just 1 percent of government spending was properly accounted for.² The average urban Kenyan pays 16 bribes a month.³ When Barack Obama visited the country in mid-2015, he warned of a “cancer of corruption” that was blighting the nation.⁴

1 <https://www.transparency.org/country/#KEN>

2 <http://www.businessinsider.com/afp-corruption-in-kenya-worse-than-ever-says-veteran-campaigner-2015-8?IR=T>

3 http://www.ucl.ac.uk/dpu-projects/drivers_urb_change/urb_governance/pdf_trans_corrupt/TI_Corruption_kenya.pdf, p. 7.

4 Loeb, Saul. “Barack Obama urges Kenya to fight corruption in government.” *Daily Nation*. July 26, 2015. <http://www.nation.co.ke/news/Fight-corruption-from-the-top-Obama/-/1056/2809526/-/2y8ovu/-/index.html>

Open Data and Technology in Kenya

Can open data—and technology more generally—offer a solution? Kenya is generally considered one of the most technologically advanced countries in Africa. As of 2014, Kenya had the fourth-highest Internet penetration rate in Africa with 21.3 million users, trailing only Nigeria, Egypt and South Africa.⁵ Also as of 2014, almost 80 percent of Kenyans who owned cell phones (around three-quarters of the population) used them for mobile payments and banking.⁶

The country had an early taste of open data's potential in combating corruption in 2005. Dr. Bitange Ndemo, a lecturer on entrepreneurship and research methods at the University of Nairobi's Business School, worked with then-President Mwai Kibaki's government to uncover illicit allocation of government funds by mapping distributions of the Constituency Development Funds (CDF). CDFs were supposed to spur grassroots development and distribute resources equitably across the nation. The mapping exercise, built on previously closed data, showed that funds were heavily allocated to rich areas of the country over more deserving areas. Despite its promising results, however, this initial effort to use open data met with political opposition and was soon shut down.⁷ “[The politicians] were very angry with me,” Ndemo recalls, “and within a very short time, we pulled down the website.”⁸

In 2011, Ndemo, by then the Permanent Secretary of Information and Communications in the government, worked with a volunteer task force of technical personnel and World Bank officials to launch a new project—the Kenyan Open Data Initiative. The centerpiece of the initiative was the Kenyan Open Data Portal (<https://open>

5 <http://www.internetworldstats.com/stats1.htm>

6 Gilpin, Lyndsey. “The world’s unlikely leader in mobile payments: Kenya.” *TechRepublic*. March 20, 2014.

<http://www.techrepublic.com/article/the-worlds-unlikely-leader-in-mobile-payments-kenya/>

7 Mutuku, Leonida and Christine M. Mahihu. “Open Data in Developing Countries—Understanding the Impacts of Kenya Open Data Applications and Services.” iHub Research. 2014. <http://www.opendataresearch.org/sites/default/files/publications/ODDC%20Report%20iHub.pdf>

8 GovLab Interview with Dr. Bitange Ndemo, Chairman Open Institute. September 9, 2015.

data.go.ke/), the first comprehensive open data platform in sub-Saharan Africa (and the second in Africa, after Morocco's). As of late 2015, the portal hosts over 400 data sets from across the government, including data from the national census, data pertaining to national and regional expenditure, and information on key public services such as education, health and agriculture.

Determining whether the broad Kenyan Open Data Initiative, and the Data Portal in particular, had any impact is difficult given conflicting signals. On the one hand, the portal's 44 million page views as of August 2015, for example, point to a widely used, impactful platform. Even more important than the number of page views, Linet Kwamboka, the Open Data project coordinator, announced that users have interacted with the data found on the portal—through downloads and embeds—2.6 million times.⁹

That being said, prior to these milestones, critics questioned how well the information available on the platform was targeted to the needs of users. For example, a 2014 survey of relevant users found that approximately 50 percent of desired education data does not exist on the portal.¹⁰ The lack of many types of potentially useful data is not surprising when considering the number of government departments that have not yet opened their data. As of 2015, four years after the portal's official launch, the site has only received data from 26 of over 83 government departments.¹¹

9 Mutegi, Lilian. "Kenya Open Data Portal Hits 44 Million Page Views." All Africa. August 18, 2015. <http://allafrica.com/stories/201508190752.html>.

10 Mutuki, Leonida, Christine M. Mahihu, Dr. Raed Sharif. "Exploratory Study on the Role and Impact of Kenyan Open Data Technology Intermediaries." iHub Research. April 30, 2014. http://www.opendataresearch.org/sites/default/files/publications/ODDC_Event_Handout%20.pdf.

11 Ochieng, Lilian. "Kenya: State Agencies Hold Firmly to Their Data." AllAfrica. August 19, 2015. <http://allafrica.com/stories/201508200439.html>.

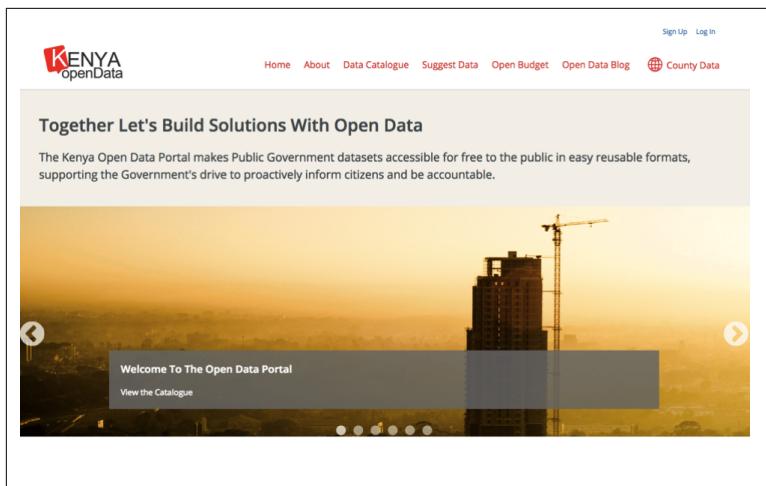


Figure 9-1. Kenya's Open Data Portal

The political and legal climate in Kenya also remains somewhat ambiguous when it comes to open data. In 2013, President Uhuru Kenyatta was elected as the fourth president of Kenya. His Jubilee Party's election manifesto had "pledged to increase transparency in government as well as promote information-sharing within public institutions."¹² The manifesto also highlighted the importance of improved digitization of government, including "cleaning up and managing databases that will be stored in a secure and centralized location and which can be accessed and used by all ministries and branches of government ... to make government more efficient."¹³

All of this led to hopes among open data and government transparency activists. Immediately following that initial optimism, hopes began to deflate. In 2013, the Sunlight Foundation published an article on "why Kenya's open data portal is failing." Many of the reasons cited—like government's reluctance to open all potentially useful data and challenges related to the country's constitutional devolution—created similar issues for Open Duka, as discussed below. It is worth noting, however, that much has changed for the portal since

12 H.E. President Uhuru Kenyatta. "The Harmonised Jubilee Coalition Manifesto." <http://www.scribd.com/doc/123569244/The-Harmonised-Jubilee-Coalition-Manifesto#scribd>.

13 H.E. President Uhuru Kenyatta. "The Harmonised Jubilee Coalition Manifesto." <http://www.scribd.com/doc/123569244/The-Harmonised-Jubilee-Coalition-Manifesto#scribd>.

2013, and even at the time, the piece’s author still expressed reasons for longer-term optimism.¹⁴ The situation surrounding the open data portal isn’t helped by the absence of a legal framework to enforce a constitutionally guaranteed freedom of information. Although the Bill of Rights includes this freedom, in fact there are no clear guidelines to guide how public bodies should make information available, what types of information should be released, and what remedies exist for violations. With that said, the surge in use of the portal—culminating in the milestones in site visits and data downloads witnessed in 2015—is a lesson in the need for both iteration and persistence in order for open data efforts to take hold.

Project Description and Inception

Even as the government has launched its open data initiatives, civil society in Kenya has become increasingly active in the field. According to Jay Bhalla, director of the Open Institute, a civil society organization that works “to open up government to the citizens,” the government’s own open data initiative “acted as a catalyst for organizations [like the Open Institute] to begin working around open data.”¹⁵ He cites the example of Code4Kenya, which works to catalyze demand for open data by embedding fellows with data skills into African newsrooms and civic organizations, as an example, stating that it was at least in part “spun out of the fact that there was the Kenya Open Data Initiative.”

Open Duka (“duka” means “shop” in Swahili) grew out of this civil society interest in open data. Launched in 2014, it contains data pertaining to the relationships among Kenyan institutions, individuals and other entities with an impact on public life. The project is driven by the Open Institute in partnership with the National Council of Law Reporting, with funding from the African Technology & Transparency Initiative.

Open Duka is built on a simple principle: Its founders believe that giving citizens information that enables them to draw connections

¹⁴ Brown, Greg. “Why Kenya’s open data portal is failing—and why it can still succeed.” Sunlight Foundation. September 23, 2013. <https://sunlightfoundation.com/blog/2013/09/23/why-kenyas-open-data-portal-is-failing-and-why-it-can-still-succeed/>.

¹⁵ GovLab Interview with Jay Bhalla, Executive Director of the Open Institute Trust. August 12, 2015.

among individuals and organizations can uncover and possibly remedy many instances of corruption. Al Kags, a founding trustee of the Open Institute, notes that every time a new corruption scandal hits Kenya, citizens begin asking questions about connections. Who is connected to whom? Who owns the companies involved, and how do those individuals relate to government entities or other power players? Prior to Open Duka, he says, there was no way to uncover this type of “relational data,” which can “enable citizens to make connections between people, issues, companies and public organizations.”¹⁶

Open Duka’s Mission Statement:

“Providing citizens, journalists, and civic activists with a practical and easy-to-use tool to understand the ownership structure of the world they live in, demonstrating the practical applications of open information for normal citizens.”

The platform does not itself seek to draw any connection; it just posts data, relying on its users’ investigative skills (and serendipity) to draw out the links that underlie public life in Kenya, and that often drive corruption. As Bhalla puts it: “We’re just getting all the data that is out there ... and we’re putting it up there for people to make their decisions. ... We’re just getting the data, scraping it, putting it up, and seeing what connections come up ... and what stories can then be taught.”¹⁷

Kags cites the following hypothetical example of how Open Duka might be used. Using the platform, a citizen could discover that Individual A is the director of two companies; she could then uncover the fact that Individual A’s brother directs a third company, and also happens to work with a government agency; in addition, the citizen might determine that both brothers went to school with Judge C.¹⁸ In this way, the citizen-user of Open Duka could possibly determine whether Individual A has used his connections (via the

¹⁶ GovLab Interview with Al Kags, Founder Open Institute. March 4, 2014. <https://www.youtube.com/watch?v=GYfovn5TQpI>.

¹⁷ GovLab Interview with Jay Bhalla, Executive Director of the Open Institute Trust. August 12, 2015.

¹⁸ GovLab Interview with Al Kags, Founder Open Institute. March 4, 2014. <https://www.youtube.com/watch?v=GYfovn5TQpI>.

brother or the judge) to improperly benefit from government contracting, or engage in any other form of corruption.¹⁹

At its core, Open Duka is a platform that enables citizens to peer behind the veil of previously unknown (and unknowable) relationships that drive a variety of economic and political activities, including land purchases, tendering contracts and other regulatory decisions. When selecting a person, organization, tender, contract, court case or piece of land on the site, Open Duka provides a simple visualization of how that entity connects with others found on the platform, with additional categorized data about each entity found in the visualization displayed below. These relationships, adds Kags, “have an impact on the integrity of a country.”²⁰



Figure 9-2. Visualization of an Organization's Connections found on Open Duka

Data Sources

As of late 2015, Open Duka includes information on 30,955 people, 3,832 organizations, 1,800 tenders, 226 contracts, 22,011 court cases and 4,418 pieces of land.²¹ The user can search across each of these six areas from a simple search bar on the main page, or enter into

19 GovLab Interview with Al Kags, Founder Open Institute. March 4, 2014. <https://www.youtube.com/watch?v=GYfovn5TQpI>.

20 GovLab Interview with Al Kags, Founder Open Institute. March 4, 2014. <https://www.youtube.com/watch?v=GYfovn5TQpI>.

21 <http://www.openduka.org/>

the platform by clicking on one of the six categories of content, at which point she is presented with a long, searchable, alphabetical list of relevant entities.

The screenshot displays a web-based land information system. At the top, there's a navigation bar with links to Home, About, API, Blog, FAQ, and a prominent 'Donate' button. Below the navigation is a search bar labeled 'Search results [Land]' with a placeholder 'Search by name, company or organisation' and a 'Go' button. The main content area shows a grid of search results. Each result row contains a small thumbnail image, a unique identifier (e.g., 11, 12654, 12657, 2581, 2584, 2587, 3809, 12, 12655, 13, 2582, 2585, 2879, 4375, 12653, 12656, 14, 2583, 2586, 3776, 69), and a link to 'Gazette Notice No. 5926', 'THE REGISTERED LAND ACT (Cap. 300)', or 'Abdella Hafiani Choz'. Below the grid is a navigation bar with links to page 1, 2, 3, 4, 5, 6, 'NEXT', and 'LAST'.

Figure 9-3. Land Use Information Hub on Open Duka

Increased transparency around land records is seen as a particularly important piece of Open Duka's value proposition. The goal is to allow individuals to draw connections between land parcels (where illicit funds are often parked) and entities or individuals who might not be direct owners but nonetheless have connections to the owners of the land. In addition, Open Duka's organizers have also taken steps to include information not just on governmental stakeholders, but also private corporations. This data is often harder to access, but it is essential to drawing the necessary connections. Thus, Open Duka comprises data related, for example, to contract information, tender information and company directors (including whether the directors are involved in any court cases and whether they own land). Kags notes that Open Duka's focus expands beyond government because "if you don't know what a company, civil society organization, school, etc. is doing, you don't know what they're doing wrong and you have no way to add value."²²

²² GovLab Interview with Al Kags, Founder Open Institute. March 4, 2014. <https://www.youtube.com/watch?v=GJfovn5TQpI>.

While some of this information originates from publicly available government databases, Open Duka's organizers also employ a number of additional measures to supplement the information and make their databases more comprehensive. In particular, Open Duka scrapes a considerable amount of information from the media, particularly from the *Kenyan Gazette* (which contains near-comprehensive information on land transfers and purchases). Prior to Open Duka's launch, the *Gazette* had partnered with Google to digitize all its editions; however, while the information was accessible in digital format, it was not categorized or searchable in any meaningful way. So Open Duka's organizers approached the government and asked them for access to the digital archives of the *Gazette*, which was then imported into the platform.

Despite Open Duka's ingenuity in sourcing data from nonobvious sources, the project has in general struggled to cope with the insufficiency of available data. This lack of data is not unique to Kenya, however. Improvements to transparency regarding the beneficial ownership of legal entities—i.e., business owners, shareholders and senior managing officials—was a high-level principle arising from the 2014 G20 Summit in Australia, in the interest of “preventing the misuse of these entities for illicit purposes such as corruption, tax evasion and money laundering.”²³ Despite official governmental enthusiasm for releasing data, its organizers say, the reality is often quite different. Bhalla cites the case of the Ministry of Lands, where much information still remains inaccessible due to stonewalling or vested interests, even though it is more or less completely digitized. The same is true of many other departments. As he puts it: “When you’re doing anything around government transparency and accountability, the doors often get shut to people trying to get the data.”²⁴

²³ “G20 High-Level Principles on Beneficial Ownership Transparency.” G20.org. December 2014. https://g20.org/wp-content/uploads/2014/12/g20_high-level_principles_beneficial_ownership_transparency.pdf

²⁴ GovLab Interview with Jay Bhalla, Executive Director of the Open Institute Trust. August 12, 2015.

Impact

Intended Beneficiaries

Average Citizens

- Provides citizens with the ability to recognize and avoid potentially corrupt dealings in areas like contracts, land purchases.

Media and Research Community

- Uncovers connections between stakeholders and power players in the country to help in the identification of corruption or conflicts of interest.
- Enables large-scale analyses of data found on the platform to uncover high-level insights.

Developers and Civic Activists and Researchers

- An API enables interested parties to incorporate Open Duka's data into other applications focused on transparency and accountability in the public and private sectors.

Despite considerable promise and excitement about the potential of Open Duka, its impact has in fact been quite limited to date. No systematic study has been done of the project's impact, and the success stories that do exist have been largely anecdotal (Bhalla, for instance, refers to the “the odd success story” where “some person managed to figure out that someone was trying to con him with selling a piece of land, for example, and he managed to use Open Duka as the platform to find out that this guy is the fraudster”²⁵). Bhalla also notes that the media has benefitted from the platform, but establishing direct ties between news pieces on corruption and journalists' use of Open Duka is a challenge (and, as of 2015, the site does not offer any illustrative examples of such uses). Ndemo believes that more effort needs to be expended to improve the media's data literacy and capacity to create impacts as infomediaries for the public. Rather than focusing on providing sensationalistic headlines, there should be a greater focus in the media on, “[doing] the analysis that can help people understand to make decisions. ... The media does not just have to look at headlines, they must synthesize the data and

²⁵ GovLab Interview with Jay Bhalla, Executive Director of the Open Institute Trust. August 12, 2015.

make it much simpler, and to do that, we need to do a lot of capacity-building on data science, data visualization.”²⁶

Open Duka’s limited success can in large part be attributed to the many challenges the project has encountered since its launch, notably the difficulty of sourcing data and the obstacles raised by a recent move to devolve governmental authority (explained further below).

Seeding the Notion of Openness

As it stands, much of Open Duka’s impact has been less related to its specific purpose, and more general—i.e., spreading the idea of openness and accountability across sectors. For instance, ONE and the Accountability Lab, civil society organizations focused on combatting poverty and corruption, present Honesty Oscars that honor “the creative work of activists and organizations that fight global corruption.” In 2015, Open Duka won the “Best Special Effects Oscar” based on how it “provides citizens, journalists and civic activists with a practical and easy-to-use tool to help them understand the ownership structure of the world they live in.” The awards citation went on to say that Open Duka “helps demonstrate the practical application of open information for citizens around the world.”²⁷

Challenges

As noted, Open Duka has faced considerable hurdles in its efforts to use data to draw connections and reduce corruption. This section discusses some of the most important challenges.

Government Devolution

In 2010, the Kenyan government devolved a number of powers (including key financial authority) to its 47 counties. While this devolution was championed as a way to empower local communities and spur grassroots development, it also created a number of difficulties for open data advocates. Bhalla, for instance, notes the com-

²⁶ GovLab Interview with Dr. Bitange Ndemo, Chairman Open Institute. September 9, 2015.

²⁷ Walker, Emily. “And the Honesty Oscars go to...” One.org. February 23, 2015. <http://www.one.org/us/2015/02/23/and-the-honesty-oscars-go-to/>.

plexity of now having to convince government representatives about the benefits of transparency not just in one jurisdiction (i.e., at the federal level), but in 47. And while the Open Institute is trying to work with each of these counties, he also notes that many of the regional governors (and other officials) are struggling to cope with the responsibilities of their new authority. Perhaps not surprisingly, when faced with essentially building a local government from scratch, open data simply isn't a high priority. However, as described below, though the government devolution process created significant hurdles for Open Duka and open data projects more generally, it is also providing new opportunities for building counties' data capacity to internal and public benefit.

Data Quality and Availability

As noted, the broader political and legal climate around open data in Kenya is not fully encouraging or supportive. Many government departments have yet to fully release their data, and even when they do, the quality of the data can be subpar. When browsing Open Duka's database, the data quality challenge can be evident—with some entries having two-letter or confusing titles scraped from the *Kenyan Gazette*.

In addition, Ndemo points out that many government officers resist releasing data in a timely manner, thus reducing the timeliness and relevance of the information. He cites the case of government's national statistic officers, who "tend to think that they need to collect the data, [and] curate it before they can give it." He goes on: "The statisticians have not come to terms that there are other people who can handle data. They just think this is our data, and these are computer geeks [in the open data community] who are interfering with the processes."²⁸ Such "data hoarding" by government officers makes it very difficult for a project like Open Duka to achieve the scale it requires.

Public Trust and Expectations

A final challenge is posed by a general lack of public interest in open data, and even a sense of disenchantment. Over the years, Kenya has

²⁸ GovLab Interview with Dr. Bitange Ndemo, Chairman Open Institute. September 9, 2015.

seen no shortage of attempts to improve public life and the quality of governance. There have also been several anti-corruption efforts before Open Duka. Many of these efforts have been sponsored by foreign donor entities, and have been accompanied by considerable hype. Few of them have made much of a dent.

According to Bhalla, this creates a sense of disenchantment and distrust among the population. “It’s gotten to a stage where even if there is something useful that works, there’s a lot of distrust,” he says. So Open Duka’s task—and challenge—is not only to prove that it is useful on its own terms, but also to distance itself from previous, failed attempts to improve public life.

More generally, there is concern that the lack of public confidence in open data could lower the pressure on governments to release data, thus creating a vicious cycle and undermining the sector as a whole. Ndemo speaks about the dangers of public disenchantment and a lack of demand for open data. “We must create interest,” he says. “We must create the demand for data. If we can’t create that demand, even if government were to release it, it would make no sense at all.”²⁹ He goes on to speak about the need to increase public awareness and educate the public about the potential of open data.

Looking Forward

Overall, Open Duka’s organizers say that the project has been far more resource-intensive and time-consuming than they had originally anticipated. Among other things, organizers underestimated the difficulty of sourcing data, of government resistance, and of convincing a somewhat disenchanted population to use it. While the project was initiated in 2013, Bhalla notes that it was not until early 2015 that, on the technical side, “we got the platform to where we wanted it to be.”³⁰

Today, the project’s organizers are focused on the future, with a number of streams of activity in the works that are focused on turning optimism and initial excitement into the type of real-world

²⁹ GovLab Interview with Dr. Bitange Ndemo, Chairman Open Institute. September 9, 2015.

³⁰ GovLab Interview with Jay Bhalla, Executive Director of the Open Institute Trust. August 12, 2015.

impacts that have eluded the platform to date. The task at hand for the team behind the project is to learn from the challenges mentioned above and iterate quickly and strategically—despite continued outside challenges and resource strains. Thankfully, the Open Duka team recognizes this need, noting on the site that, “Open Duka is a work in progress (and a labour of love) for us and we shall continue to add new features and data.”³¹

One encouraging development for the Kenyan open data ecosystem more broadly is the planned launch of Version 2.0 of the Kenyan Open Data Portal, which, among other things, is designed to help users “find the location of donor- and government-funded projects in their community as well as the stage of execution.”³²

Open County Program

The central expansion of Open Duka currently under development is the Open County Program. Under this program, designed to ease some of the challenges that have arisen from devolution, the team is working with new county governments to build their data capacity and push forward citizen engagement opportunities. Overall, the team is seeking to create new areas for coordination between county governments, NGOs, community media and community stakeholders.³³

This effort will not only give counties the capacity to govern better, but will also, in theory, directly benefit Open Duka. By building technical capacity for counties, Open Duka will gain access to constantly updated and uploaded supply-side data of sound quality. On the demand side, engaging with outside stakeholders will help the Open Institute gain a diversity of feedback to guide future development. This marriage of supply and demand represents a promising next step. But, Bhalla notes: “[It is] very resource-intensive.”³⁴

³¹ <http://www.openduka.org/>

³² Mutegi, Lilian. “Kenya Open Data Portal Hits 44 Million Page Views.” All Africa. August 18, 2015. <http://allafrika.com/stories/201508190752.html>.

³³ GovLab Interview with Jay Bhalla, Executive Director of the Open Institute Trust. August 12, 2015.

³⁴ GovLab Interview with Jay Bhalla, Executive Director of the Open Institute Trust. August 12, 2015.

There are currently a few counties in which Open Duka is piloting this new project. So far results have been encouraging. Bhalla notes that through the Open County Project, “Data gets released and now we’re able to embed that data in Open Duka at a very granular level. And at a level that now impacts people on that ground.”³⁵ Moreover, given Open Duka’s relatively simple underlying code, counties will be able to host individual iterations of the platform, meaning that county-level data will feed into the central Open Duka platform, but could also provide the backbone of more granular, regional versions of the tool. Bhalla believes that once Open Duka reaches this level of granularity, “the chance of impact will be much, much higher.”³⁶

Tracking Metrics of Success

While there has been little evidence of impact to date around Open Duka, the development and implementation of concrete metrics is in the works. Bhalla notes that, “we don’t look at page views as a metric.”³⁷ Rather, the Open Institute is focusing on three types of metrics built around three different user groups:

1. By tracking application programming interface (API) calls, the Open Institute will be able to ascertain whether *developers and civic activists* are actually making use of the data on the site to build applications, or if *researchers* or other parties are accessing the data in bulk to perform targeted analyses.³⁸
2. Focusing on the number of embeds of visualizations created through Open Duka, on the other hand, will help to assess the site’s usefulness to the *media*. Since it is unlikely that each and every instance of a journalist using Open Duka to find information for a story will be credited, tracking the number of direct

³⁵ GovLab Interview with Jay Bhalla, Executive Director of the Open Institute Trust. August 12, 2015.

³⁶ GovLab Interview with Jay Bhalla, Executive Director of the Open Institute Trust. August 12, 2015.

³⁷ GovLab Interview with Jay Bhalla, Executive Director of the Open Institute Trust. August 12, 2015.

³⁸ GovLab Interview with Jay Bhalla, Executive Director of the Open Institute Trust. August 12, 2015.

- embeds of visualizations created on the site will help to gain a better understanding of its use in the media.³⁹
3. Finally, the Open Duka team will begin tracking the number of external requests for additional data to be added to the platform. This can provide an indication of *citizen use and engagement*. Such requests, says Bhalla, will show “people have used the platform and they’re not finding what they’re looking for.”⁴⁰

Making the Platform More Accessible

While Kenya is generally considered a leader in Internet penetration in Africa, many citizens may nonetheless remain excluded from platforms like Open Duka. In particular, citizens in rural areas, and those from otherwise underprivileged backgrounds, generally remain on the wrong side of the digital divide.

Recognizing that the usefulness of its tool is directly related to the extent it is used, the Open Institute is working on a number of measures to improve access. For example, an SMS-based solution is in the works, and Bhalla speaks of the potential of a version of Open Duka that would allow, for instance, a citizen considering a real estate purchase to quickly send a text message to the platform to gain access to all legal proceedings against the seller.⁴¹ Ndemo, too, speaks of the potential of a mobile-based solution. “You cannot underestimate the ubiquity of the mobile,” he says. “I think everything is going to be on the mobile platform. ... We need to focus, that we leverage on the mobile platform on many services to the citizens.”⁴²

39 GovLab Interview with Jay Bhalla, Executive Director of the Open Institute Trust. August 12, 2015.

40 GovLab Interview with Jay Bhalla, Executive Director of the Open Institute Trust. August 12, 2015.

41 GovLab Interview with Jay Bhalla, Executive Director of the Open Institute Trust. August 12, 2015.

42 GovLab Interview with Dr. Bitange Ndemo, Chairman Open Institute. September 9, 2015.

Expansion to Other Countries

Finally, although there is much work to be done in order to make Open Duka as useful as possible in Kenya, the Open Institute still has an eye on broader regional expansion. Kags notes that a bigger goal for the project is to bring Open Duka to Uganda, Tanzania, Nigeria and beyond. This goal is not only based on the belief that a good idea should spread to other contexts, but also on a recognition that a Kenyan might look to buy land in Tanzania, or a Nigerian might be interested in opening a business in Uganda. Developing a cross-border means for citizens to assess the relational factors and potential areas of corruption before making major decisions could result in major benefits across the region.⁴³ Of course, optimizing the platform for Kenya and increasing the on-the-ground impacts of Open Duka there will be essential for the idea to take hold in other countries.

Open Duka represents a first step toward addressing a real and complex problem in Kenya. While the platform has had relatively limited impact to date, in many ways this is the result of unexpected political and administrative hurdles (combined, perhaps, with an underestimation of the time and resources required to build such an ambitious platform). Ultimately, the founders of Open Duka are coming up against the reality that creating an open data *platform* requires a supportive open data *ecosystem*. Ultimately, this may prove to be the largest (or at least the very first) impact and contribution of Open Duka: In pushing against the limits and challenges of open data in Kenya, it is laying the foundations for future success, both for itself and for others.

⁴³ GovLab Interview with Al Kags, Founder Open Institute. March 4, 2014. <https://www.youtube.com/watch?v=GYfovN5TQpI>.

CHAPTER 10

Mexico's Mejora Tu Escuela

Empowering Citizens to Make Data-Driven Decisions About Education

By Andrew Young and Stefaan Verhulst

Summary: Founded by the Mexican Institute for Competitiveness (IMCO), with support from Omidyar Network and others, Mejora Tu Escuela (<http://mejoratuescuela.org>) is an online platform that provides citizens with information about school performance. It helps parents choose the best option for their children, empowers them to demand higher-quality education, and gives them tools to get involved in their children's schooling. It also provides school administrators, policymakers and NGOs with data to identify areas requiring improvement and hotbeds of corruption, in the process raising the overall quality of education in Mexico.

Dimension of Impact: Empowering Citizens—Informed Decision-making

Key Takeaways:

- Making previously inaccessible information available to the public is a major public service, but a necessary next step involves giving citizens a means to meaningfully act on that information.
- Platforms to reduce corruption run up against powerful vested interests. Open data advocates need to develop strategies not only to source and disseminate information, but also to push

back against these interests. The challenges are as much social and political as they are technical.

- In cases where governments are reluctant to release potentially damaging information to the public, civil society can play an important role in finding creative ways to obtain, clean and publish that information.
- In countries with major societal problems like rampant corruption, it is easy to lose sight of the potentially impactful uses of open data to improve the everyday lives of citizens.

Context and Background

Education in Mexico

For years, education in Mexico has been substandard and generally lagged behind standards in other, comparable countries. This poor performance cannot be attributed to a lack of budgetary resources—Mexico spends a larger portion of its budget on education than any other member of the Organization of Economic Cooperation and Development (OECD). It is important to note, though, that, at least in part, this is due to the higher number of Mexican children who pass through the public school system than other OECD countries.¹ But while that money has led to nearly universal basic education, outcomes remain stubbornly poor. Mexico's graduation rates are near the bottom of OECD countries, with fewer than half of students earning a high school diploma.² In addition, Mexican students perform poorly in international comparative tests of math, science and reading skills,³ and, according to a recent study, 80 percent of

1 http://www.sep.gob.mx/wb/sep1/sep1_Estadisticas

2 Villagran, Lauren. "Test the teacher? Educators balk at Mexico's reforms." *The Christian Science Monitor*. March 1, 2013. <http://www.csmonitor.com/World/Americas/2013/0301/Test-the-teacher-Educators-balk-at-Mexico-s-reforms>

3 Villagran, Lauren. "Test the teacher? Educators balk at Mexico's reforms." *The Christian Science Monitor*. March 1, 2013. <http://www.csmonitor.com/World/Americas/2013/0301/Test-the-teacher-Educators-balk-at-Mexico-s-reforms>

teachers failed an evaluation exam designed to test their competencies.⁴

While Mexico's education system has been generously funded, it has suffered over the years from endemic corruption. The scale of corruption is difficult to calculate, but it is by all accounts massive. In 2014, an "abuse meter"—a giant electronic signboard erected in the capital by activists that tries to update in real time the amount of waste—estimated that at least \$2.8 billion are lost to corruption every year in the education sector. The year the billboard was erected, it showed more than \$33 million lost in just the first week of the school year.⁵

The corruption takes many forms: ghost teachers on payrolls, pre-selling of exam results, officials with lavish lifestyles, teachers demanding bribes to give good grades and candidate teachers bribing to pass training tests. The rot is widespread and, increasingly, widely recognized. One activist recently referred to corruption in the education system as "the robbery of the century."⁶ A 2009 report from Transparency International found that the average household paid an additional \$30 per year for their children's education, despite that education being "constitutionally free."⁷ In 2010, Transparencia Mexicana's National Index on Corruption and Governance found that parents paid a bribe 3.5 percent of the time when seeking an enrollment form for a public school.⁸

Information and Open Data in Mexico

In Mexico, as in much of the world, information has increasingly been seen as a tool to combat corruption. As Gavin Starks argues in

4 Villagran, Lauren. "Test the teacher? Educators balk at Mexico's reforms." *The Christian Science Monitor*. March 1, 2013. <http://www.csmonitor.com/World/Americas/2013/0301/Test-the-teacher-Educators-balk-at-Mexico-s-reforms>

5 Cave, Damien. "Billboard Drives Home Extent of Corruption as Schools Suffer." *The New York Times*. September 1, 2014. <http://www.nytimes.com/2014/09/02/world/americas/billboard-drives-home-extent-of-corruption-as-schools-suffer.html>.

6 Cave, Damien. "Billboard Drives Home Extent of Corruption as Schools Suffer." *The New York Times*. September 1, 2014. <http://www.nytimes.com/2014/09/02/world/americas/billboard-drives-home-extent-of-corruption-as-schools-suffer.html>.

7 Transparency International, 'Corruption in the Education Sector', Working Paper no. 04/2009 (Berlin: TI, 2009), p. 6.

8 <http://gateway.transparency.org/tools/detail/331>

The Guardian: “Allowing citizens to freely access data related to the institutions that govern them is essential to a well-functioning democratic society. It is the first step toward holding leaders to account for failures and wrongdoing.”⁹

Mexico has tried to apply these principles to its education sector. In 2008, the Mexican Congress passed a law requiring all states to provide the federal government with information about the condition of schools, payrolls and other expenditures. At the time, this push for greater transparency was seen as a vital component of the effort to combat corruption in education. However, the law has proven largely toothless: Four out of 32 states did not hand over payroll databases for the last trimester of 2013, and eight states handed over empty or incomplete databases.¹⁰

It is in this context that Mexico’s open data initiatives must be considered—both as part of a strategy to combat corruption in education, and also to fight corruption in general. Currently, the country is ranked 24th in the Open Data Barometer, having moved up one spot since 2013.¹¹ The government maintains a federal open data portal, which contains information from federal, state and municipal governments, as autonomous government entities (which are not subject to administrative and technical control from the central government), like the Instituto Federal de Telecomunicaciones. The open data portal is part of a broader National Digital Strategy championed by Ania Calderón, the director general of open data for the Mexican government, who is pushing to transition the way open data is treated from “a passive resource to actionable intelligence that can give us concrete results.”¹²

9 Starks, Gavin. “Why open data should be central to Fifa reform.” *The Guardian*. June 12, 2015. <http://www.theguardian.com/news/datablog/2015/jun/12/open-data-central-fifa-reform>.

10 Zabludovsky, Karla. “Data Dive Finds 100-Year-Old Teachers and Phantom Schools in Mexico.” *Newsweek*. May 30, 2014. <http://www.newsweek.com/data-dive-finds-hundred-year-old-teachers-and-phantom-schools-mexico-252871>.

11 <http://barometer.opendataresearch.org/report/analysis/rankings.html>

12 Interview with Ania Calderón, Director General of Open Data, Mexican Government. March 9, 2014. <https://www.youtube.com/watch?v=j9L5Wo0LEpY>.

The portal has achieved some early wins, such as the release of 100 government data sets in 42 days soon after its launch.¹³ However, while such accomplishments are laudable, and may even have made some inroads, corruption remains a formidable challenge, and it is important to recognize the limitations of Mexico's open data efforts thus far. Oscar Montiel, director of community engagement at Codeando México, an organization that seeks to "bring together the best talent and organizations to create a new generation of civic technology,"¹⁴ notes that while "on a policy level everything sounds fantastic," the government's open data plans are at times poorly executed.¹⁵ He also points out that while there does exist a federal open data portal, questions remain (as they do in all countries) over "who decides what should be open and when." Finally, he adds that because Mexico's open data ecosystem is largely built around (often multi-year) contracts and licenses, changes can take months or even years, and there is a general lack of agility when it comes to releasing open government data.¹⁶

Others are more positive. Rafael García Aceves of Transparencia Mexicana describes a "really vibrant community"¹⁷ of open data users in the country. He argues that it is up to the open data community "to identify and demand which data is needed and to work more on the analytical models," and to come up with new "solutions" that can help "link data with problems." In other words, as the government continues to make data available, the task facing Mexican open data community, both inside and outside government, is to find new, innovative ways of using that data, and to connect the data to real-world problems.

13 Bustos Arellano, Paulina, Ricardo Daniel Alanis Tamez, William Gerry and Codeando Mexico. "Open data in Mexico: how Data Squads jump-started national publishing" Open Data Institute. <https://theodi.org/case-studies/mexico-case-study-using-data-squads-to-jump-start-government-open-data-publishing>

14 Translated by Google. <http://codeandomexico.org/>

15 GovLab Interview with Oscar Montiel, Codeando México. August 19, 2015.

16 GovLab Interview with Oscar Montiel, Codeando México. August 19, 2015.

17 GovLab Interview with Rafael García Aceves, Transparencia Mexicana. September 1, 2015.

Product Description and Inception

It is within this context of formidable challenges and limited successes—as well as a deeply entrenched and widely prevalent culture of corruption—that the success, however provisional, of Mejora Tu Escuela must be considered. Launched in 2013, Mejora Tu Escuela (“Improve Your School,” in Spanish) is a public, independent, non-profit initiative with a team comprising members of the Mexican Institute for Competitiveness (IMCO) and with support from Omidyar Network.¹⁸ The stated goal of the platform is to “promote citizen participation to improve education in Mexico.” The project is premised on the belief that “education in [Mexico] will improve only with the active commitment of all members of the educational community, including parents.”¹⁹

In an interview, Alexandra Zapata Hojel, the coordinator of educational projects at IMCO, explained that “Mejora Tu Escuela was born as an effort to really get parents involved in their child’s education.”²⁰ At IMCO, there was a growing frustration that parent’s lack of ability to engage with the education system in Mexico created an enabling environment for special interests to influence the delivery of education.²¹ The parental disengagement is evident in what Zapata Hojel calls “a singular dissonance” in Mexican education: While the country performs quite poorly on various global measurements of education quality, citizen satisfaction in fact remains quite high. A survey conducted in 2013, for example, found that 78 percent of Mexican parents were satisfied or very satisfied with their children’s education (meanwhile, the country scored last in Programme for International Student Assessment (PISA) tests among OECD countries).²² Unlike other examples in this series of case studies, therefore, the effort to use open data to improve Mexican education did not originate in or draw strength from public discontent—in fact, it began as an attempt to address public complacency.

18 Disclosure: Omidyar Network also funded this case study.

19 <http://www.mejoratuescuela.org/quienes-somos>

20 GovLab Interview with Alexandra Zapata Hojel, Coordinator of Education Projects, IMCO. August 7, 2015.

21 GovLab Interview with Alexandra Zapata Hojel, Coordinator of Education Projects, IMCO. August 7, 2015.

22 <http://datos.gob.mx/historias/educacion/mejora-tu-escuela.html>

For Zapata Hojel, parental complacency is indicative of a wider problem: a lack of information. Historically, citizens have found it difficult to access or comprehend data that would point to the poor state of Mexican education. In addition, Mexico faces some challenges specific to a rapidly developing nation where levels of educational attainment and social mobility have increased dramatically across recent generations. As Zapata Hojel explained: “[For many] families where the parents graduated from 6th grade, if they see their children in middle school they automatically think that things are better than when they were growing up. So the first challenge we had to face was to kind of shake up the system and share with all of these parents the fact that education in Mexico was of an extremely poor quality and that we were on the bottom of the totem pole in terms of results on a worldwide scale. To do that, we had to access information on all of the schools in the country.”²³

Project Description

Currently, the Mejora Tu Escuela project comprises two key parts. The first is the public-facing platform, targeted at parents and other citizens with a role in the education community. The second is focused on the public-policy side and includes reports on problems with transparency, corruption, teacher payrolls and the like. This aspect of the project also enables media to uncover problems—especially problems related to corruption—in the education sector. Although both aspects of the project are important (and in many ways mutually reinforcing) the case study here is primarily concerned with the public-facing side.

The platform side of Mejora Tu Escuela has four components:

- Get to Know Your School—the central information provision aspect of the project focuses on giving parents and other stakeholders access to a diversity of information on schools.
- Compare Your School—allows parents to compare their child’s school to others in the same area, complete with an assigned

²³ GovLab Interview with Alexandra Zapata Hojel, Coordinator of Education Projects, IMCO. August 7, 2015.

school grade. The hope is that this piece of the platform will “get competition going between public schools.”²⁴

- Grade Your School—asks parents and other members of the school community to rate different aspects of a school on a scale of one-to-ten, and leave comments. This information is then fed into the Get to Know Your School stream of information.
- Improve Your School—in its current form, Improve Your School provides “concrete tools and suggestions to parents on how to participate in addressing problems in their child’s school.”²⁵ In the future, this section will also be able to suggest concrete actions to parents or teachers based on the hyper-local needs identified (through crowdsourced reporting and/or existing data) for a given school.²⁶



Figure 10-1. Mejora Tu Escuela

²⁴ GovLab Interview with Alexandra Zapata Hojel, Coordinator of Education Projects, IMCO. August 7, 2015.

²⁵ Hasan, Munyema. “Mexico: Encouraging Parents’ Engagement in Public Education Reform.” Open Government Partnership. October 2013. http://www.opengovpartnership.org/sites/default/files/Mexico_0.pdf

²⁶ GovLab Interview with Alexandra Zapata Hojel, Coordinator of Education Projects, IMCO. August 7, 2015.

Get to Know Your School and Compare Your School

Currently, Mejora Tu Escuela is populated by around 25 databases, with new databases being added as they are made available by the Ministry of Education. Pablo Clark, a content developer at IMCO, notes that likely the most important database is the 2013 census, which includes information about school infrastructure, the number of students in schools, the number of teachers, the names of principals, school locations and contact details (including phone number and email address, if they have one).²⁷ Additional information is drawn from standardized test results, teacher assessments and other data focused on the educational proficiency of schools.

School Grades

In addition to providing—often previously inaccessible—data to parents on Mexican schools, Mejora Tu Escuela, as part of the Compare Your School stream, provides graded assessments of schools to enable parents to make better decisions about their children’s schooling and to push back against underachieving schools.

Grades for schools are calculated through student result data, with a focus on results in mathematics and Spanish.²⁸ Schools are then ranked according to four levels of education quality: “Excellent,” “Good,” “Panzazo”—an expression that roughly translates to “getting through by the skin of one’s teeth”—and “Failed.” While representing broad and perhaps somewhat crude measurements of school quality, these grades do allow parents to make informed decisions about school selection—and, equally important, empower citizens to demand results from school officials.²⁹

Impact

The impact of Mejora Tu Escuela is evident in the amount of traffic the site receives—between 40,000-45,000 unique visits a day. As Zapata Hojel notes, these numbers are all-the-more remarkable given that IMCO has not invested heavily on dedicated marketing

²⁷ GovLab Interview with Pablo Clark, Analyst, IMCO. August 17, 2015.

²⁸ <http://www.mejoratuescuela.org/metodologia>

²⁹ <http://www.mejoratuescuela.org/metodologia>

efforts at this point, instead relying on community engagement and word-of-mouth.³⁰

“We remain optimistic that things will slowly start to change. Our hope is that more monitoring of these information sources will begin to change how money is spent in this sector in Mexico.”

—Alexandra Zapata Hojel, IMCO

Who are these visitors and the site’s users? As with the other case studies in this report, Mejora Tu Escuela’s impact has been diverse and diffuse, and can be measured by considering various demographic groups and dimensions of impact. We consider impact by taking into account two key beneficiaries (parents and officials) and discussing the site’s overall effect on two dimensions of impact (accountability and corruption).

Impact on Beneficiaries

Parents have been the chief beneficiaries of the site. Prior to the site’s launch, it was virtually impossible for parents to access information about their children’s school. Zapata Hojel notes that this inaccessibility extended not just to test results or information about school facilities; even basic information like school addresses and phone numbers was difficult to access.³¹

The lack of information led to a feeling of disempowerment and disengagement on the part of parents. When launching the site, IMCO heard many stories about parents feeling intimidated by teachers, unwilling or unable to challenge them and demand accountability. Teachers were regarded as “untouchable people that can never be fired, that have so much power, and that can take your son or daughter out of school if they don’t like you.”³²

Slowly, but undeniably, the information available on the site is changing the balance of power. IMCO has collected a number of anecdotes demonstrating the impact on parents. They include the

³⁰ GovLab Interview with Alexandra Zapata Hojel, Coordinator of Education Projects, IMCO. August 7, 2015.

³¹ GovLab Interview with Alexandra Zapata Hojel, Coordinator of Education Projects, IMCO. August 7, 2015.

³² GovLab Interview with Alexandra Zapata Hojel, Coordinator of Education Projects, IMCO. August 7, 2015.

story of a woman who was able to combat teacher absenteeism at her child's school using the platform. Teacher absenteeism is a common problem in the Mexican education system, but teachers are rarely fired despite a law that dictates that teachers must be dismissed after three consecutive days of unjustified absences. In this case, the mother was able to use information from Mejora Tu Escuela, and sue the state for failing to fire the teachers in question.³³ Now with the help of IMCO her story is being shared with other parent-teacher groups across the country.

While parents have been the major beneficiaries of the site, the project's sponsors have been surprised by the unintended impact on **policymakers and school officials**. For example, IMCO members and others involved with the site report that they get as many phone calls and emails from principals and local government officials as they do from parents: "We get governors, and people working for the governor's office that call us up and say, 'We're using your data because your data is more reliable than that of the Ministry of Education.'"³⁴ Clark notes that the number of emails and phone calls IMCO receives from teachers and school principals show "they needed this data. It shows that they have a great interest in having an active participation in their school."³⁵ In this way, Mejora Tu Escuela is not only empowering parents, but also gradually changing the attitude of government officials toward education and schools.

Increased Accountability

The impact of Mejora Tu Escuela can also be measured by considering its important role in increasing accountability. This is partly a result of greater parent involvement and empowerment; today, parents are demanding more from schools and school officials. The media, too, has played a role: Using information from the site, jour-

³³ GovLab Interview with Alexandra Zapata Hojel, Coordinator of Education Projects, IMCO. August 7, 2015.

³⁴ GovLab Interview with Alexandra Zapata Hojel, Coordinator of Education Projects, IMCO. August 7, 2015.

³⁵ GovLab Interview with Pablo Clark, Analyst, IMCO. August 17, 2015.

nalists have been able to unearth corruption and shine a spotlight on abuse.³⁶

Government administrators and school officials have also played an important role in enforcing accountability. Like teachers, well-intentioned administrators who have long struggled to reform the education system have themselves been empowered by the site; they, too, have been using information to better monitor schools, teachers and the education system in general. IMCO staff note, for example, that it is only through the site that certain school officials find out they have been enrolled in certain federally funded programs. One official reportedly said: “I’ve never seen a cent of this money.” Through the information available on the site, school officials are able to identify corruption in the system, tracking down missing funds and pinpointing the places in the bureaucracy where money is being stolen.³⁷

Despite such anecdotal evidence, the broader impact on accountability in Mexican education remains to be seen, however. García Aceves of Transparencia Mexicana notes that transparency does not always translate automatically to impact. “Mexico is just starting to realize that this link between transparency and accountability is not something that is linear or direct,” he says. “If you have more transparency, that doesn’t mean that you have more accountability. It could actually be the opposite.”³⁸ Felipe Estefan of Omidyar Network agrees while arguing that, although the link between transparency accountability is not always direct, transparency is a “necessary ingredient for increased accountability,” the ultimate goal of Mejora Tu Escuela.³⁹

36 “IMCO denuncia que SEP manipuló anomalías sobre maestros.” *Televisa*. May 15, 2014. <http://noticieros.televisa.com/mexico/1405/imco-denuncia-sep-desaprecio-anomalias-maestros/>.

37 GovLab Interview with Alexandra Zapata Hojel, Coordinator of Education Projects, IMCO. August 7, 2015.

38 GovLab Interview with Rafael García-Aceves, Open Data Project Coordinator, Transparencia Mexicana. September 1, 2015.

39 GovLab Interview with Felipe Estefan, Associate, Investments, Omidyar Network. January 6, 2016.

Corruption Report

Finally, the website's impact is evident in the way it has highlighted corruption within the education system and pushed the topic to the forefront of public discussion. In 2014, IMCO used Ministry of Education data available on the site to publish a report on the state of corruption in Mexico's education system.⁴⁰ The report's findings were backed by the National Audit Agency and stirred national outrage.

Among the report's findings:

- 1,442 teachers on government payroll were between the ages of 100 and 105. Of these, 1,441 were registered in the state of Hidalgo, and all but one was born on December 12, 1912 (i.e., 12/12/12).
- 70 teachers earned more than President Enrique Peña Nieto (who made around \$15,000 per month). Of these, 19 worked at schools that had failed or barely passed the national standardized test.⁴¹
- The report also found that the average monthly national salary for Mexican teachers was \$1,954, and that the teacher with the highest salary lived in Oaxaca state and earned \$46,849 a month. In the state of Guerrero, there was a school with a single student, but six employees whose payroll added up to \$6,644 per month.⁴²
- The report uncovered 536 telesecundarias, or secondary and high school programs available in rural areas via satellite, that apparently operated without electricity. The report also showed that there existed a large number of "phantom schools" that

⁴⁰ http://imco.org.mx/wp-content/uploads/2014/05/20140514_MapaMagisterioMexico_Presentacion.pdf

⁴¹ Zabludovsky, Karla. "Data Dive Finds 100-Year-Old Teachers and Phantom Schools in Mexico." *Newsweek*. May 30, 2014. <http://www.newsweek.com/data-dive-finds-hundred-year-old-teachers-and-phantom-schools-mexico-252871>.

⁴² Zabludovsky, Karla. "Data Dive Finds 100-Year-Old Teachers and Phantom Schools in Mexico." *Newsweek*. May 30, 2014. <http://www.newsweek.com/data-dive-finds-hundred-year-old-teachers-and-phantom-schools-mexico-252871>.

were unregistered (and thus possibly nonexistent), yet that had several teachers on payroll.⁴³

Prior to the report's public release, 10 states initiated audits, and teacher payrolls are now being funded by the federal government instead of state governments. Zapata Hojel hopes to see more "very clear policy results." She adds: "We remain optimistic that things will slowly start to change. Our hope is that more monitoring of these information sources will begin to change how money is spent in this sector in Mexico."⁴⁴

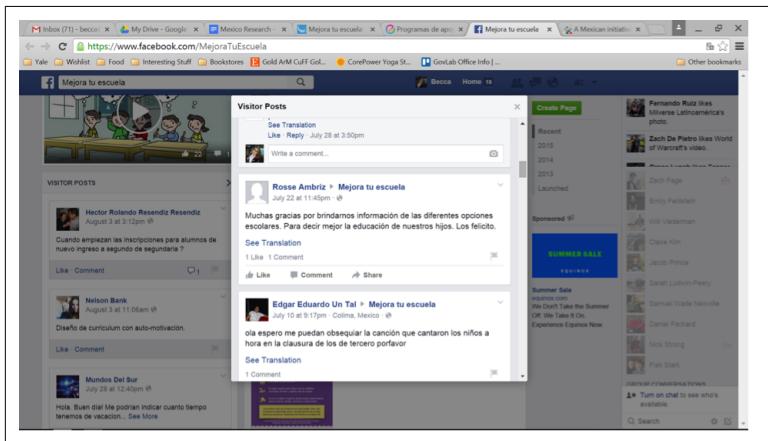


Figure 10-2. A parent thanks Mejora Tu Escuela for granting access to school information

Challenges

In seeking to change Mexican education, Mejora Tu Escuela is working against powerful forces and vested interests. Inevitably, it has faced, and continues to face, a series of challenges to its operations and growth plans. As we have seen so often in this report, open data programs and activists frequently encounter these types of chal-

⁴³ Zabludovsky, Karla. "Data Dive Finds 100-Year-Old Teachers and Phantom Schools in Mexico." *Newsweek*. May 30, 2014. <http://www.newsweek.com/data-dive-finds-hundred-year-old-teachers-and-phantom-schools-mexico-252871>.

⁴⁴ GovLab Interview with Alexandra Zapata Hojel, Coordinator of Education Projects, IMCO. August 7, 2015.

lenges—they are part of the process of social change, and of changing the way information is consumed and disseminated.

Standardized Test Cancellation

Revelations of corruption can have unintended effects, especially when they come up against such vested interests. In 2014, IMCO's corruption report uncovered the fact that 19 of the 70 teachers who make more than the Mexican president work at schools that "failed or barely passed the national standardized test."⁴⁵ That year, under pressure from the teachers' union and others embarrassed by the results, as well as public pressure led by NGOs like Mexicanos Primero, the national standardized testing program was canceled. "As soon as we put standardized test results on the website, and as soon as parents were able to, for the first time ever, compare results amongst schools at the national level including private and public schools, four months later the federal government announced that the standardized tests were being cancelled," laments Zapata Hojel.⁴⁶

The cancellation of the tests led to a major gap in usable data to assess the effectiveness of teachers and schools. Thankfully, the testing has recently been reimplemented (albeit in a new form), and once the data is made public, IMCO will once again import the new data into Mejora Tu Escuela. Still, the cancellation has caused lasting difficulties. In addition to the challenge of a yearlong gap in data, IMCO will have to "find some way to make [the new results] comparable with the results from the last test, which had a different system."⁴⁷

Data Accessibility

While national census data in many ways forms the backbone of Mejora Tu Escuela, it took considerable effort to gain access to that data. For example, only 16 percent of the nonpersonally identifiable information from the census was originally released. The govern-

⁴⁵ Zabludovsky, Karla. "Data Dive Finds 100-Year-Old Teachers and Phantom Schools in Mexico." *Newsweek*. May 30, 2014. <http://www.newsweek.com/data-dive-finds-hundred-year-old-teachers-and-phantom-schools-mexico-252871>.

⁴⁶ GovLab Interview with Alexandra Zapata Hojel, Coordinator of Education Projects, IMCO. August 7, 2015. August 7, 2015.

⁴⁷ GovLab Interview with Pablo Clark, Analyst, IMCO. August 17, 2015.

ment's initial reluctance to release more information was, in part, due to apprehensions about the extent of corruption that would be revealed. Ultimately, much of the remaining data was leaked to IMCO (and other Mexican NGOs), and this data was posted online.

Zapata Hojel notes that much of the data on the website cannot, therefore, be properly called "open data."⁴⁸ The information is, rather, a combination of open data, leaked data and information gleaned through scraping and citizen input. This holds potentially valuable lessons: Sometimes data activists must be creative in the ways they source information, combining a variety of inputs to create a genuinely useful tool.

Data Quality

In addition to problems with accessibility, IMCO has also faced considerable problems with data quality. Oscar Montiel of Codeando México, believes that the problem stems in part from the lower levels of quality required by government itself to do its job. He says: "Their standards are so low, even for themselves, for their own work that when you try to push a bit higher then they don't respond that well."⁴⁹

Whatever the source of the problem, IMCO has from the onset been faced with making accessible error-ridden and outdated data, often stored in inconsistent formats. The IMCO experience points to the significant difference between data being public and data being truly open—i.e., available in standardized, machine-readable technical formats. In many ways, in fact, IMCO has played a major role in cleaning the data, for example through citizen inputs that have corrected errors and updated information. Zapata Hojel notes that, going forward, much of the success of Mejora Tu Escuela will rely on whether or not IMCO and others will be able to push the government "to keep the platform up to date, to keep the platform current, to be able to provide parents and school communities with information that is valuable."⁵⁰

⁴⁸ GovLab Interview with Alexandra Zapata Hojel, Coordinator of Education Projects, IMCO. August 7, 2015.

⁴⁹ GovLab Interview with Oscar Montiel, Codeando México. August 19, 2015.

⁵⁰ GovLab Interview with Alexandra Zapata Hojel, Coordinator of Education Projects, IMCO. August 7, 2015.

Looking Forward

Mejora Tu Escuela's early years have provided valuable proof of concept, and its organizers are pushing ahead with new features and steps to grow the service. The coming months and years will witness a number of enhancements. Some of the additions or steps under consideration are described here.

Connecting Needs with Actions

Under its "Improve Your School" initiative, Mejora Tu Escuela is working to create concrete opportunities that would allow parents (and others) to take actions to address problems they identify. For example, one of the most common forms of cheating in Mexico involves teachers simply dictating answers to their students. Parents who identify this problem now have no way to solve it. Mejora Tu Escuela is looking to include a mechanism that would suggest to such parents that they donate one day per year of their own time to supervise a test.⁵¹ In this way, the gap between information and action can be narrowed.

Publishing a School Budget

Although the site includes a decent amount of information, it still does not clearly link school performance and infrastructure to budgets. This means, for example, that the federal government has no real idea how much it costs to educate a child across the country in various regions and localities. Zapata Hojel states that, in the future, Mejora Tu Escuela will work to add more specific and granular expenditure information—which will be a challenge considering that three levels of government are involved in public school budgeting—so that parents can not only see how schools are performing, but how they are performing relative to funds received. This, Zapata Hojel believes, will take the notion of accountability to a whole new level.⁵²

⁵¹ GovLab Interview with Alexandra Zapata Hojel, Coordinator of Education Projects, IMCO. August 7, 2015.

⁵² GovLab Interview with Alexandra Zapata Hojel, Coordinator of Education Projects, IMCO. August 7, 2015.

Developing a Report Pipeline

As it stands, when a parent has a problem related to their child's school, the pathway to addressing that problem is intimidatingly opaque since different types of problems require different routes toward addressing them. "If it's a problem related to bullying, it's one route they need to take, and if the problem is in school infrastructure or sexual abuse, the trail they have to take is completely different."⁵³ Zapata Hojel believes that the pathways are intentionally obscured by the Ministry of Education, leading to not only parents, but also principals being unable to determine a school's supervisor, for example.

Upon gaining access to databases containing information on who key decision-makers are for different schools, IMCO is developing a new "Report Pipeline" to enable parents to report problems and gain step-by-step instructions for how to get that information to the relevant person. Zapata Hojel sees this as a way to "put pressure on some of these government systems and also ... really be able to give parents help in resolving this."⁵⁴

Translating Information into Action

For the people and organizations behind Mejora Tu Escuela, the larger ambition is not simply to provide an information portal, but to create a platform for action. García Aceves notes that, as it stands, "people may be more informed but probably they do not have the tools for activating the institutional mechanisms that they need to solve a problem."⁵⁵ While Mejora Tu Escuela provides tools that could be valuable for certain users, García-Aceves argues that every organization (including his own) should work toward developing tools with a clear eye toward impact. Estefan of Omidyar Network states that for Mejora Tu Escuela to achieve the greatest impact possible in the future, it should find ways to draw pathways between complaints (and complainers) and policy responses (and policymak-

⁵³ GovLab Interview with Alexandra Zapata Hojel, Coordinator of Education Projects, IMCO. August 7, 2015.

⁵⁴ GovLab Interview with Alexandra Zapata Hojel, Coordinator of Education Projects, IMCO. August 7, 2015.

⁵⁵ GovLab Interview with Rafael García-Aceves, Open Data Project Coordinator, Transparencia Mexicana. September 1, 2015.

ers). Considering the difficulty of simply identifying relevant institutional decision-makers in the education system, Zapata Hojel and the team at IMCO will likely face significant challenges in finding ways to incentivize those decision-makers to actually respond to and address citizen complaints.⁵⁶

⁵⁶ GovLab Interview with Felipe Estefan, Associate, Investments, Omidyar Network.
August 20, 2015.

CHAPTER 11

Uruguay's A Tu Servicio

Empowering Citizens to Make Data-Driven Decisions on Health Care

By David Sangokoya, Ali Clare, Stefaan Verhulst and Andrew Young

Summary: Every February, Uruguayan citizens are given the opportunity to choose whether to change or stay with their existing health care provider. In the country's mixed public-private health care system, several factors come into play when making this decision: the location of the health provider, number of doctors and pediatricians available, hours open, etc. How can Uruguayan citizens make data-driven, informed decisions on choosing health care providers? Datos Abiertos, Transparencia y Acceso a la Inform (DATA) Uruguay, an open data civil society organization, partnered with the Uruguayan Ministry of Health to create A Tu Servicio (<http://atuservicio.uy>), a website providing easily digestible, searchable and visualized infographics based on open government health data. The Web platform allows users to select their location and then to compare local health care providers based on a wide range of parameters and indicators, such as facility type, medical specialty, care goals, wait times and patient rights. A Tu Servicio has introduced a new paradigm of patient choice into Uruguay's health care sector, enabling citizens not only to navigate through a range of options but also generating a healthy and informed debate on how more generally to improve the country's health care sector.

Dimension of Impact: Empowering Citizens—Informed Decision-making

Key Takeaways:

- Open data initiatives, including but not limited to those focused in health care, stand to benefit greatly from collaboration between intermediaries—including civil society, the media and, in this case, health care data providers—and government.
- While increasing access to information is an important first step in efforts to unlock the value of open data, simply making information available is not sufficient: information needs to be standardized, machine-readable and visualizable to be truly useful to citizens.
- Successful open data projects have viral potential and can quickly spread to other sectors and applications, or be replicated in other countries.

Context and Background

With an Internet penetration rate of 60.5 percent,¹ Uruguay is considered an “emerging and advancing” country, according to the Open Data Barometer.² Uruguay has been a member of the Open Government Partnership (OGP) since 2011, and is ranked 12th on the Open Knowledge Foundation’s Global Open Data Index, making it one of the more advanced open data countries in Latin America.³ National data from a variety of units and agencies is made available for reuse through the national open data portal (<http://www.agesic.gub.uy/>), led by AGESIC, the national agency for eGovernment. While recognizing the “great potential” for Uruguay to innovate around open data, the Sunlight Foundation noted recently that many challenges lie ahead for the country as it seeks to transform open data efforts into a mainstream, sustainable government practice.⁴ Despite these challenges, the case study below illustrates the emerging impact of Uruguay’s opening up of government data.

1 <http://www.internetworldstats.com/sa/uy.htm> (2014).

2 <http://www.opendatabarometer.org/report/about/index.html>.

3 <http://global.census.okfn.org/>.

4 Keserű, Júlia and James Kin-sing Chan. “The Social Impact of Open Data.” Sunlight Foundation, 2015. <http://www.opendataresearch.org/dl/symposium2015/odrs2015-paper20.pdf>

Health care in Uruguay is delivered through a mixed public-private framework, with private insurers and hospitals working alongside a public health care infrastructure. Citizens can choose their health care service providers. Once an individual has been with a service provider for three years, they are eligible to change their provider at the beginning of the following February. Fabrizio Scrollini, chairman of DATA Uruguay, noted that, as a result, the month of February is typically marked by public debate and discussion over the factors that influence citizens to choose (or leave) a provider. He adds that February is also typically marked by heavy advertising on the part of providers, many of whom “encourage citizens to join them and leave others, and even potentially pay” users to switch providers.⁵

At the time A Tu Servicio (“At Your Service”) was launched in February 2015, the Ministry of Health was starting to experiment with open data, collecting raw data on service quality from every service provider and publishing these and other health data annually. Data management was generally poor, however, with data frequently out of date, and data sharing hampered by incompatible information management systems. Health care data wasn’t available in a format that would allow citizens to make informed decisions. Additionally, according to Daniel Carranza, cofounder of DATA Uruguay, the pressures from competing health providers led citizens to “rely on marketing and advertising campaigns based on opinion rather than actual data.”⁶

⁵ GovLab interview with Fabrizio Scrollini, Chairman, Data Uruguay, July 14, 2015.

⁶ GovLab interview with Daniel Carranza, Co-Founder, Data Uruguay, June 8, 2015.

Product Description and Inception



Figure 11-1. A Tu Servicio home page

First launched as a pilot program in February 2015 by DATA Uruguay and the Ministry of Health, A Tu Servicio is a Web application based on open government data that provides easily digestible, searchable and visualized infographics of key performance indicators such as facility type, available medical specialties, procedures performed, wait times, goals of care, patient rights and human resources. This platform allows users to compare these indicators across all health care providers in the country, enabling users to make well-informed decisions about their health care providers through machine-readable, interoperable and accessible data that was previously unavailable. For example, users can search across public health care providers (such as Medica Uruguaya and CAM-CEL) and private providers (such as Blue Cross Blue Shield) in Montevideo and compare data on the cost of birth control, average wait times and patient satisfaction. Such comparisons can allow a citizen to decide whether to stick with his or her current provider or switch to a new one. Enabling this type of comparison was not always easy, however, since, as Scrollini notes, “private providers were the worst in terms of compliance when sending the data to the Ministry.”⁷

⁷ GovLab interview with Fabrizio Scrollini, Chairman, Data Uruguay, July 14, 2015.

The screenshot shows a comparison table for three health service providers. The columns represent the providers: MEDICA URUGUAYA, CAMCEL, and BLUE CROSS & BLUE SHIELD (Seguro Privado). The rows show various metrics such as structure, number of employees, website, and communication methods.

CRITERIO	MEDICA URUGUAYA	CAMCEL	BLUE CROSS & BLUE SHIELD (Seguro Privado)
Estructura	MEDICA URUGUAYA	CAMCEL	Blue Cross & Blue Shield (Seguro Privado)
Nombre abreviado	MEDICA URUGUAYA	CAMCEL	Cooperativa Asistencial Medica De Cerro Largo
Nombre Completo	Médica Uruguaya Corporación De Asistencia Médica		
Web	www.medicauruguayo.com.uy	www.camcel.com.uy	www.bcbu.com.uy
Total de afiliados	29853	43613	13056
Mapa Asistencia			
Porcentaje de afiliados referenciados a un médico de referencia	49.1 %	91.9 %	7.7 %
Porcentaje de niños de un año correctamente controlados	93.0 %	100.0 %	62.5 %
Porcentaje de embarazadas controladas en el primer trimestre	98.5 %	99.9 %	100.0 %
Existencia de Espacio Adolescente en la institución	*	*	*
Comunicación en soporte electrónico (algunas web)	Conocer servicios, teléfonos, horarios de atención, carteras de servicios, formas de acceso a los servicios, normas de protección de datos personales y derechos de los usuarios	Conocer servicios, teléfonos, horarios de atención, carteras de servicios, formas de acceso a los servicios, normas de protección de datos personales y derechos de los usuarios	Conocer servicios, teléfonos, horarios de atención, carteras de servicios, formas de acceso a los servicios y normas de protección de datos personales y derechos de los usuarios

Figure 11-2. A Tu Servicio health service comparisons

The project emerged from a public consultation between the government and DATA Uruguay, a volunteer-based civil society organization committed to promoting open data, transparency, and access to information (<http://datauy.org/>). In 2013, DATA Uruguay had begun working with 180 Ciencia, a journalism portal, to create the application “Temporada de pases”⁸ (“Transfer season”) as part of a public service project to inform and empower citizens considering changing health care providers. The application relied on existing data published by the Ministry of Health; however, that data was generally of poor quality, often available only in closed formats, and included little metadata. The need for more open health care data quickly became apparent.⁹

Spurred by the evident shortcomings of the “Temporada de pases” application, the Ministry of Health initiated a discussion with Data Uruguay about ways to make more data available to citizens. This discussion, held under the auspices of the ongoing Open Government Partnership (OGP) roundtable discussions about the Alliance for Open Government, soon resulted in a modest government grant (about \$13,000) to create a pilot application that later became A Tu Servicio.¹⁰ DATA Uruguay also won additional funding from Avina

8 <http://mutualistas.datauy.org/>

9 GovLab interview with Fabrizio Scrollini, Chairman, Data Uruguay, July 14, 2015.

10 GovLab interview with Daniel Carranza, Co-Founder, Data Uruguay, June 8, 2015.

Americas, an organization promoting sustainable development,¹¹ and the Iniciativa Latinoamericana por los Datos Abiertos (“Latin American Open Data Initiative,” ILDA), a network of organizations seeking to promote research on and use of open data in the region.¹²

The key role played by DATA Uruguay is one of the most interesting and important aspects of the A Tu Servicio story. It shows the vital function of intermediaries and civil society in promoting open data, facilitating discussions with the state, and nudging government agencies to release more and higher quality data.

At the same time, the role of the state itself cannot be underestimated. A Tu Servicio was supported from the outset by the Ministry of Health, whose encouragement and input (including in product design) was deemed critical to the success of the project.¹³ Even after the website was launched, DATA Uruguay and the state acted together in collecting user feedback and instituting improvements. This included working directly with health institutions to defuse their resistance to opening up data for comparison with other providers.¹⁴ Ultimately, the success and impact of A Tu Servicio (see next section for a discussion of impact) is testament to the importance of inter-sectoral partnership, and particularly collaboration between the state and civil society.

Impact

In this section, we consider some indicators of impact to determine the success of A Tu Servicio. Since the project—initiated in 2015 and scheduled to run until 2020—is still in its infancy, any measurements of impact should be considered highly preliminary.

In measuring impact within and across these targeted beneficiaries, we consider three categories:

¹¹ <http://www.avinaamericas.org/>

¹² <http://idatosabiertos.org/at-your-service-open-data-for-health-service-delivery/>

¹³ GovLab Interview with Fabrizio Scrollini, Chairman, Data Uruguay, July 14, 2015.

¹⁴ GovLab interview with Diego Soria, Ministry of Health Uruguay, August 18, 2015.

Intended Beneficiaries

Average Citizens	<ul style="list-style-type: none">Enabling the people of Uruguay to make better-informed health decisions as a result of actionable information.Equipping citizens with data-driven evidence and tools to make better decisions on health care choice.Catalyzing citizens to act as agents of monitoring and evaluation around the health services they receive.
Health Providers	<ul style="list-style-type: none">Making clear to citizens which health options are best suited to their needs.Improving the quality and responsiveness of service based on data-driven demand from citizens.
Government Agencies	<ul style="list-style-type: none">Improving the public health system through greater efficiency, transparency and accountability.
Media	<ul style="list-style-type: none">Encouraging better data journalism efforts and data driven arguments for public debate on health care.
Civil Society and Unions	<ul style="list-style-type: none">Enabling better informed argumentation and advocacy around the status of the health care system.

Use and Awareness:

In 2014, before the launch of the application, health care provider data made available by the Ministry of Public Health received fewer than 500 downloads. This low uptake was probably due to limited awareness of the availability of data as well as poor data quality.

Following its launch, the impact of A Tu Servicio was almost immediately apparent. In the first month alone, the site received approximately 35,000 visits, a number equivalent to 1 percent of the total population of Uruguay.¹⁵ The average time spent on the platform was five minutes, and visitors accessed on average five pages per visit.¹⁶

¹⁵ GovLab Interview with Fabrizio Scrollini, Chairman, Data Uruguay, July 14, 2015.

¹⁶ GovLab Interview with Fabrizio Scrollini, Chairman, Data Uruguay, July 14, 2015.

The launch of A Tu Servicio also increased awareness among Uruguayans regarding the availability of data, and its potential to improve their health care. The application's launch received widespread media coverage, thanks to a press conference held by DATA Uruguay and the Ministry of Health in which the virtues of the initiative were emphasized. Social media was also employed strategically to disseminate information about the initiative and maintain interest after the launch. Additionally, several unions, including a doctors' union, endorsed the website, stating the need for more health care information to be made available and accessible.¹⁷

One aspect in particular of A Tu Servicio received considerable press coverage.¹⁸ Shortly after the launch, journalists and citizens began noticing and drawing attention to long wait times at public hospitals, almost all of which were in contravention of maximum wait times established by the government.¹⁹ The resulting public and media furor resulted in several hospitals changing their practices, and more generally, led to a vigorous public debate about wait times and quality of care in Uruguay.²⁰

A Tu Servicio has also played an important role in stimulating and facilitating more informed debate within Uruguay's Parliament about the future of the nation's health care system. For example, on August 11, 2015, a nationalist deputy leader of Montevideo, Martín Lema, spoke out against proposed reforms to the National Resources Fund (FNR), which protects citizens against extraordinary health care expenses in Uruguay.²¹ In making his case, Lema utilized data from the A Tu Servicio platform to refute the government's

17 GovLab Interview with Fabrizio Scrollini, Chairman, Data Uruguay, July 14, 2015.

18 "No vas a comparar." *Montevideo Portal*. February 2, 2015. <http://www.montevideo.com.uy/auc.aspx?259222>.

19 See, for example, "¿Cuánto es el tiempo de espera para especialistas en ASSE y mutualistas?" *El Observador*. February 2, 2015; "Conozca el tiempo de espera de las mutualistas para ver un medico." *Subrayado*. January 31, 2014. <http://subrayado.com.uy/Site/noticia/30503/conozca-el-tiempo-de-espera-de-las-mutualistas-para-ver-un-medico>

20 Keserú, Júlia. "How open data is changing the way Uruguayans choose their health care" Sunlight Foundation. May 13, 2015. <https://sunlightfoundation.com/blog/2015/05/13/how-open-data-is-changing-the-way-uruguayans-choose-their-health-care/>

21 "Ministro de Salud Pública, Fondo Nacional de Recursos." Uruguay Parliament Transcript. August 11, 2015. <http://www.parlamento.gub.uy/indexdb/Distribuidos>ListarDistribuido.asp?URL=/distribuidos/contenido/camara/D20150811-0218-0212.htm&TIPO=CON>

claims that proposed health care reforms would benefit vulnerable populations. He stated that the data showed anyone making this claim was “lying or misinformed.”²² Lema also publicly criticized the proposed reforms via social media²³ and public interviews, where he continued to cite data from the A Tu Servicio platform to back his case.

Data Quality:

As noted earlier, health provider data previously made available by the government was of poor quality, limiting its usability and citizen interest. A Tu Servicio spurred vast improvements in data quality through increased public scrutiny and demand for machine-readable, interoperable and accessible information.

The tool’s user-friendly visualizations helped citizens understand the data in new ways and led to greater public scrutiny and discovery of erroneous data. For the first time, citizens were able to spot errors in the data provided by health providers and, through feedback loops built into the application, request corrections. Instances where health providers had not provided sufficient information were labeled as “not available”; in many cases, citizens could identify such holes in the data and make public requests to the Ministry and private providers for the data to be updated. In addition, the Ministry of Health was able to identify instances where provider data was possibly misrepresented as well as instances where provider data did not agree with crowdsourced user information.²⁴ Providers made aware of errors have also been able to submit revised and corrected data. After the platform’s launch, Fabrizio Scrollini recalls, “Many [providers] were willing to update their data and standardize it according to our preferences.”²⁵ Overall, the launch of the application led to a new commitment to data quality on the part of providers, citizens and politicians.

²² “Lema: Basso “está desinformado o miente” sobre medicamentos de alto costo.” *Espec-tador*. August 13, 2015. <http://www.espectador.com/salud/321257/lema-basso-esta-desinformado-o-miente-sobre-medicamentos-de-alto-costo>

²³ <https://twitter.com/MartinLema404/status/629645255286091776/photo/1>

²⁴ GovLab interview with Daniel Carranza, Co-Founder, Data Uruguay, June 8, 2015.

²⁵ GovLab Interview with Fabrizio Scrollini, Chairman, Data Uruguay, October 1, 2015.

Impact on Other Open Data Projects:

Success often breeds success, and the A Tu Servicio project is a good example of how an effective project in one country and sector can spur positive developments in open data elsewhere. The impact of A Tu Servicio on open data progress can be seen in two ways: in other regional projects encouraged by Uruguay's example; and in other open data projects within Uruguay itself.

Regional and global impact: The most direct evidence of A Tu Servicio's regional spillover is evident in the Mexican state of Sonora, where DATA Uruguay and Codeando Mexico, another Mexican NGO, have been collaborating with a civil society group on the development of a health care data reporting website aiming to promote a standard for health service delivery.²⁶ This website, La Rebellion de los Enfermos,²⁷ gives citizens the ability to report on hospital incidents and provides them with essential information for filing formal complaints. The focus is on creating data standards and exploring further opportunities to open health care information to the public. DATA Uruguay also presented the platform at the Pan American Health Organization, which has used A Tu Servicio as an example of good practice for other health systems in the region. Following international recognition this project has received (for example, as a finalist in the Open Data Institute's Open Data Awards),²⁸ DATA Uruguay has received interest in creating similar platforms in Europe and Africa.²⁹

National impact: Within Uruguay, too, A Tu Servicio's success has opened up new possibilities for open data and citizen empowerment through access to information. In the wake of the application's release, other government ministries and departments have either initiated new open data projects or shown interest in doing so. The impact of A Tu Servicio on other parts of the government is evident at AGESIC, the national eGovernment agency, which is using A Tu Servicio as a "best case" to present to various ministries and other

²⁶ GovLab interview with Fabrizio Scrollini, Chairman, Data Uruguay, July 14, 2015.

²⁷ <http://www.rebeliondeenfermos.org/>

²⁸ <http://awards.theodi.org/2015-finalists/>

²⁹ GovLab interview with Fabrizio Scrollini, Chairman, Data Uruguay, July 14, 2015.

organizations to promote open government and open data.³⁰ Additionally, the Ministry of Health has been ramping up its own efforts and commitment to open data, establishing new standards for openness and quality, and publishing new information about how data is collected, stored and made available for reuse. In addition, several new open projects are scheduled to be released between now and 2010.³¹ A Tu Servicio's national impact is also evidenced by the fact that it survived a change of administration—something that is, as Scrollini puts it, “very unlikely in some polities (including Uruguay).”³²

In the months following the release of A Tu Servicio, the open data movement has continued to pick up steam in Uruguay. At the Open Data Day in Latin America and the Caribbean, a group of developers analyzed open geographical data in order to analyze and visualize the number of streets in Montevideo named after notable women.³³ Upon coming to the unfortunate conclusion that only 100 out of 5,000 streets in the city are named for women, the group developed an interactive website, Calles de Mujeres, providing citizens with more information on the important contributions these women made to Uruguayan history.³⁴ Additionally, the portal located at Datos.gub.uy continues to collect applications and platforms created through the country's open data, including apps focused on transportation, culture and local history.

Challenges

A Tu Servicio has been remarkably successful, especially given the limited time since its launch. A number of challenges, however, are still acting as a barrier to greater impact. Two, in particular, are worth examining in some detail.

³⁰ GovLab interview with Diego Soria, Ministry of Health Uruguay, August 18, 2015.

³¹ GovLab interview with Diego Soria, Ministry of Health Uruguay, August 18, 2015.

³² GovLab interview with Fabrizio Scrollini Chairman, Data Uruguay, October 1, 2015.

³³ Open Data Day was an event for journalists, researchers, entrepreneurs, students and technologists to learn about open data tools, data visualization, storytelling, and to discuss the open data movement in Latin America and the Caribbean. <http://blog.okfn.org/category/events/open-data-day/>

³⁴ Garcia, Yas. “Open Data Day 2015 in Latin América and The Caribbean | Open Knowledge Argentina.” Open Knowledge Argentina. April 15, 2015. <http://ar.okfn.org/2015/04/15/open-data-day-2015-in-latin-america-and-the-caribbean/>.

Timing constraints

A Tu Servicio's period of peak usefulness is very short, limited to the month of February (and the weeks immediately preceding), when citizens can change their service providers. If the service is to scale up and attract more users, it needs to overcome two hurdles:

- a. Find a way of maximizing exposure and reach during its limited period of peak usefulness. In particular, site administrators need to ensure that the application contains all necessary data during this time and that the application is widely disseminated through media and other channels.
- b. Find ways of engaging users outside the period of peak usefulness. While citizens may only be able to change providers at a certain time of the year, health concerns and the need for health-related information are not time-bound. Ultimately, A Tu Servicio may be able to establish itself as a more general-purpose health-information portal, one that could be useful throughout the year.

Outreach and communication

Despite a widely publicized press conference and several mentions in the national media, the outreach campaign for A Tu Servicio was in fact rather hasty and somewhat ad hoc. Future efforts should not only be more sustained, but also focused on reaching under-served and under-privileged populations. These populations may not only lack access to health care; they are also least likely to be Internet savvy, and thus A Tu Servicio must confront something of a double challenge in reaching out to them.

In this regard, A Tu Servicio may find it useful to engage and collaborate with intermediaries, both in the media and especially in civil society. A Tu Servicio's success is itself testament to the power of inter-sectoral collaboration and the powerful part that can be played by civil society. Several other examples included in this report also offer testament to the ability of intermediaries and civil society groups to spread awareness of new applications and websites, particularly among traditionally underserved groups. A Tu Servicio could apply some of those lessons, spreading its usefulness and expanding the potential of open data beyond the 1 percent of Uruguayans who currently use the application.

Looking Forward

Since the launch of A Tu Servicio, DATA Uruguay and the Ministry of Health have worked on a number of initiatives and features to improve the platform. These initiatives focus on increasing usability, particularly through the development of a mobile application; creating better feedback mechanisms for incorporating and crowdsourcing user data; and providing a kind of “TripAdvisor” for users to describe and rate their experiences at specific hospitals. Overall, A Tu Servicio is seen as the first step in a broader scheme of projects within the Ministry of Health to disseminate open government health data. For example, the government is also considering disseminating information related to vaccinations more widely.³⁵

As we see the broad-based adoption of “wearables”, allowing for devices such as smartphones to collect personal data and “quantify oneself,” the possibility of adding a “small data” element to the application may be considered. Small data refers to the practice of collecting individual-level information and then providing that data back to those individuals in a more readily understandable way to improve personal health decision-making. For instance, the application could collect information on an individual’s heart health or glucose levels and then offer that information in a visualized, timecharted way, allowing individuals to track their health and keep chronic conditions under control. This feature is but one of the potential avenues for achieving the broader goal of moving A Tu Servicio towards more of a generalized health information portal. Though, as Scrollini notes, “Uruguay has a very stringent personal data protection law. As a result, the cost of developing [such a small data effort] could be high for a lightweight initiative.”³⁶

On initial planning, A Tu Servicio is scheduled to run until 2020. Given its impact and general popularity, however, there is little doubt that this specific platform, in its current form, represents just the start of a much longer process of transformation and opening up in Uruguay’s health care sector.

³⁵ GovLab interview with Daniel Carranza, Co-Founder, Data Uruguay, June 8, 2015.

³⁶ GovLab interview with Fabrizio Scrollini, Chairman, Data Uruguay, October 1, 2015.

PART IV

Case Studies: Creating Opportunity

According to our research, open data is creating new opportunities for citizens and organizations in many contexts. The ability to leverage open data as an asset is promoting economic growth, job creation and an improved capacity for innovation across regions and sectors.

The case studies in this section examine:

- The creation of new industries built on open weather data released by the United States National Oceanic and Atmospheric Administration (NOAA).
- The generation of billions of dollars of economic activity as a result of the Global Positioning System (GPS) being opened to the global public in the 1980s, and the United Kingdom's Ordnance Survey geospatial offerings.
- A more level playing field for small businesses in New York City seeking market research data.

CHAPTER 12

Great Britain's Ordnance Survey

A Clash of Business Models

By Christina Rogawski, Stefaan Verhulst and Andrew Young

Summary: Ordnance Survey (OS), Britain's mapping agency, is considered to be one of the best sources of geospatial data in Great Britain. Its data supports essentially any industry or activity that uses a map: urban planning, real estate development, environmental science, utilities, retail, and much more. Creating data that is not only profoundly useful but technically impressive results in high demand for OS data. However, despite the launch of its OS OpenData platform in 2010, and despite initiatives to spur innovation with the private sector, much of OS data is part of a fee-based licensing model. OS is required to be self-financing, a goal it has historically achieved by charging fees for its data. In recent years this model has come into conflict with the U.K. government's commitment to open data, the broader open data community, and even elements of the private sector. Today, OS uses a mixed-cost model, with some data open and some data paid, but tensions between these two aspects (open and closed data) persist.

Dimension of Impact: Creating Opportunity—Economic Growth

Key Takeaways:

- The impacts of a given open data set can span a vast array of sectors and users—in the case of OS' data, everything from medicine to mapmaking.

- The relationship between public data providers and private industry actors can create new value and opportunities from open data, but also risk creating conflict and stifling innovation if barriers such as fees are placed upon data sets. Such a system, however, can allow public data providers to grow revenue from the private sector, rather than from taxpayers or other government agencies.
- Some governments and institutions are still making efforts to determine the best open data policies that balance financing needs with open data principles.
- Whether government entities should be investing in research and product development directly, rather than focusing efforts on supporting private sector actors to innovate, remains a difficult debate.
- Collaborating with users through directly crowdsourcing ideas, formal stakeholder engagement or analyzing user behavior can help increase understanding of how to best make use of open data and can ensure that data owners build partnerships, products, and/or release data in ways that can create maximum value for end users.

Context and Background

OS Background

Ordnance Survey traces its roots to 1745, when the British government ordered its defense ministry at the time—the Board of Ordnance—to map the Scottish Highlands in response to a rebellion, and then to survey England's southern coasts as the nearby French Revolution broke out. The first Ordnance Survey maps were published in 1801, and over time, the design of OS maps came to focus less on military users and more on a diverse set of general users, particularly as the emergence of railroads, cycling and automobiles increased public demand for accurate maps.¹ These maps have long

¹ “Our history.” Ordnance Survey. <https://www.ordnancesurvey.co.uk/about/overview/history.html>

been popular with the private and public sector alike, generally admired for both their technical and artistic elements.²

Historically, OS has been proactive in integrating new technologies —e.g., aerial photography—that have held potential to improve and enhance the accuracy of maps. In 1971, computers and digital mapping were introduced. In 1974, the traditional one-inch map was replaced by a metric 1:50,000 map,³ one of the large-scale maps OS continues to use today, which portray real-world features larger on the map to provide a very high level of detail.⁴ Large areas of 1:25,000 scale mapping were also used to create a series of Outdoor Leisure Maps, which remain popular among outdoor enthusiasts and travelers.⁵

In 1990, now a fully civilian agency, Ordnance Survey began using the National Global Positioning Network (i.e., British GPS) to replace the triangulation network (a surveying method that relies on the angles in a triangle formed by three control points).⁶ OS launched its website in 1995 and digitized its remaining 230,000 maps, making Britain the “first country in the world to complete a large-scale electronic mapping program.”⁷

By the end of the 1990s, advances in computing and the use of hand-held pen computers by surveyors made OS data routinely available to customers within 24 hours of surveys being completed. With these capabilities, OS solidified its reputation as a leader in producing highly detailed, technically impressive and accurate maps. In 1999 the British government designated OS as a Trading

2 Wainwright, Martin. “Ordnance Survey maps are worth paying for.” *The Guardian*. April 3, 2010. <http://www.theguardian.com/commentisfree/2010/apr/03/ordnance-survey-maps-online>

3 This means that a distance of 1cm on the map represents 50,000cm (500m) in the real world. “Our history.” Ordnance Survey. <https://www.ordnancesurvey.co.uk/about/over-view/history.html>

4 “Products and services FAQs.” Ordnance Survey. <http://www.ordnancesurvey.co.uk/support/products-services.html>

5 “Our history.” Ordnance Survey. <https://www.ordnancesurvey.co.uk/about/overview/history.html>

6 “Surveying for Mapping—Section 2, Surveying Methods.” ICSM. <http://www.icsm.gov.au/mapping/surveying2.html>

7 “Our history.” Ordnance Survey. <https://www.ordnancesurvey.co.uk/about/overview/history.html>

Fund, requiring OS to cover its costs by charging for its products and giving a portion of its profits to the U.K. Treasury (although OS had always charged nongovernment users for maps and geospatial data).⁸ The effects of this self-financing requirement continue to significantly affect the culture and approach to open data at OS today.

In 2001 OS launched its flagship product, OS MasterMap.⁹ OS MasterMap is “an intelligent geospatial database offering definitive, consistent, and maintained referencing to more than 460 million manmade and natural landscape features in Britain.”¹⁰ OS MasterMap also allows external information to be integrated into the map as separate layers directly on the map, or as links to other OS map products, enabling the creation of detail-rich, interactive maps.

Importance of Geospatial Data

It is unlikely that the average citizen thinks about geospatial data very often, despite using it daily. Perhaps the most obvious reason to use spatial data is to locate a particular point of interest, yet even that activity can range from the personal (e.g., How do I get to 123 Main St?) to national government planning (e.g., Which is the best route along which to build a highway?). Other examples of use include real estate developers surveying land for purchases and construction; seismologists, geologists and climatologists analyzing environmental events and patterns; public health officials tracking disease outbreaks and identifying correlating environmental variables; local police analyzing crime patterns and statistics by neighborhood; and retailers or marketers planning a new campaign or store launch.¹¹

When asked to describe the importance of geospatial data, Alyssa Wright, president of the U.S. Board for OpenStreetMap says, “That’s like asking ‘What’s the purpose of a map?’ It’s so fundamental to

⁸ “Our history.” Ordnance Survey. <https://www.ordnancesurvey.co.uk/about/overview/history.html>

⁹ <https://www.ordnancesurvey.co.uk/business-and-government/products/mastermap-products.html>

¹⁰ “Our history.” Ordnance Survey. <https://www.ordnancesurvey.co.uk/about/overview/history.html>

¹¹ Bailey, Trevor C. and Anthony C. Gatrell. *Interactive Spatial Data Analysis*. (New York: Routledge), 1995.

understanding who people are, how they relate, where they're going to go." The potential impact of geospatial data is amplified when made available on the Web, and exponentially so if it is made open.¹² Citizens use spatial data on mobile devices constantly today, from locating public transportation options, hailing an Uber or reading Yelp reviews of the nearest Thai restaurant. For researchers and policymakers, integrating geospatial data with demographic and health surveys can enable analysis by spatial characteristics, such as proximity to roads, health clinics, public transportation or levels of urban development, making their services more cost-efficient and effective.¹³

As the Open Data Institute explains, geospatial data is "incredibly valuable because it is the underlying data for so much information. It's the glue that can bind together multiple different data sets, and make the dream of a truly linked web of data come true."¹⁴ Nigel Clifford, CEO of OS, highlights this concept, stating, "Everything happens somewhere. We can tell you where that thing is and connect lots of different pieces of information."¹⁵ OS understands that the energy and infrastructure, land and property, financial services, and retail categories are the largest users of their data. However, particularly with the advent of open data, John Carpenter, OS director of Strategy, notes that there are likely to be "new uses, new products and industries that we hadn't even thought of before," illustrating the enormous potential of geospatial data.¹⁶

12 GovLab interview with Alyssa Wright, President of the U.S. Board for OpenStreetMap, September 25, 2015.

13 Fischer, Alex. "Critical Role for Geospatial Data in Advancing Sustainable Development Goals." State of the Planet Blog, Earth Institute, Columbia University. May 8, 2015. http://blogs.ei.columbia.edu/2015/05/08/sdg_geospatial_data2015/

14 Tennison, Jeni and Stuart Harrison. "Collaborative Geospatial Data." Open Data Institute. March 2014. http://www.w3.org/2014/03/lgd/papers/lgd14_submission_55

15 Cave, Andrew. "Why apps won't be ousting maps any time soon as Ordnance Survey chief charts a digital future." *The Telegraph*. October 18, 2015. <http://www.telegraph.co.uk/finance/11939740/Why-apps-wont-be-ousting-maps-any-time-soon-as-Ordnance-Survey-chief-charts-a-digital-future.html>

16 GovLab interview with John Carpenter, Director of Strategy and Robert Andrews, Head of Corporate Communications, Ordnance Survey, September 16, 2015.

Open Data in the United Kingdom

Given that spatial data is used in so many contexts, it follows that many actors, across industries, want access to the best data available. Having always been a leader in spatial data, OS is now under tremendous pressure from its users, the broader open data community, and in some ways the U.K. government, to release its data publicly. Such pressure often occurs against the backdrop of a more general push by the U.K. government toward more open and transparent government. For example, in 2006, *The Guardian* began a “Free Our Data” campaign, calling upon the U.K. government to make its data open and free for download, noting that “government-funded and approved agencies such as Ordnance Survey and UK Hydrographic Office and Highways Agency are government-owned agencies; they collect data on our behalf. So why can’t we get at that data as easily as we can Google Maps or the Xtides program?”¹⁷

Open government momentum grew significantly under the Gordon Brown administration (2007-2010), which launched several acclaimed open government and open data initiatives.¹⁸ These include the open data portal data.gov.uk, which was launched in 2010 by Sir Tim Berners-Lee, inventor of the World Wide Web.¹⁹ The U.K. portal offers almost three times the number of data sets available on the comparable U.S. site (data.gov). This dedication to being a global leader in open government continued under Prime Minister David Cameron, who pledged to make the U.K. government “the most open and transparent in the world.”²⁰ These efforts are evidenced by the U.K. being ranked as first out of 86 countries in the Open Data Barometer by the World Wide Web Foundation as of 2015,²¹ and by the U.K. being one of the eight founding members of

17 <http://www.freeourdata.org.uk/>

18 Butcher, Mike. “Gordon Brown launches big shift to open gov data and broadband but where’s the detail?” *TechCrunch*. March 22, 2010. <http://techcrunch.com/2010/03/22/gordon-brown-launches-big-shift-to-open-gov-data-and-broadband-but-wheres-the-detail/>

19 Anderson, Kevin. “Tim Berners-Lee launches UK public data website.” *The Guardian*. January 21, 2010. <http://www.theguardian.com/technology/blog/2010/jan/21/timbernerslee-government-data>

20 <http://www.opengovpartnership.org/country/united-kingdom>

21 <http://barometer.opendataresearch.org/report/analysis/rankings.html>

the global Open Government Partnership.²² In a related project also commissioned by Omidyar Network, researcher Becky Hogge, who studied six cases of open data in the U.K., finds that, “the theory of change that drives government open data policy is working, and makes a strong case for continuing to pursue the policy.”²³

Project Description and Inception

OS Today

Today Ordnance Survey’s Public Task Statement includes:

To maintain and develop the underlying physical infrastructure which is needed to support mapping applications; To create, maintain and provide governance for the National Geographic Database made up of geographic information datasets with coverage of all of Great Britain to a consistent specification; To make the content of the relevant datasets widely available as products which can be used by customers of all types for a wide range of purposes; To provide advice and support to the UK Government on all aspects of survey, mapping and geographic information.²⁴

To support this public task, OS collects and records: the topography and topology of natural and manmade features of the landscape; descriptive annotations and distinctive names/postal numbers of buildings, places and features; the connectivity of water and communication networks including the routing attribution of roads, tracks and paths; the alignment of administrative and electoral boundaries; and the extent and classification of certain types of land cover and use.²⁵

Ordnance Survey invests significantly in collecting, maintaining, and improving its geospatial data and mapping products, which helps to uphold its status as a leader in geospatial data, even against

22 <http://www.opengovpartnership.org/country/united-kingdom>

23 Hogge, Becky. “Open Data: Six Stories About Impact in the UK” Omidyar Network. November 2015. https://www.omidyar.com/sites/default/files/file_archive/insights/Open%20Data%20Stories%20About%20Impact%20in%20the%20UK/OpenData_CaseStudies_Report_complete_DIGITAL_102715.pdf

24 “Our public task.” Ordnance Survey. <https://www.ordnancesurvey.co.uk/about/governance/public-task.html>

25 “Our public task.” Ordnance Survey. <https://www.ordnancesurvey.co.uk/about/governance/public-task.html>

the rise of other mapping entities developed by the private sector, like Google Maps, and through crowdsourcing, such as OpenStreetMap. As OS Commercial Director Andrew Loveless describes, “The journey that we’ve been on, [as] an organization that’s been in existence for 224 years, we have been through the complete transformation from almost the calligraphy pen, to analog, to digital, and I think that’s a journey that’s probably about 900 million pounds of investment. We spend tens of millions of pounds in our refresh capability in terms of capturing change.”²⁶

Specifically, OS has a team of almost 300 field surveyors and seven aircraft²⁷ (two OS aircraft supported by five contracted aerial imagery companies) surveying all 243,241 square kilometers of Great Britain. OS’s team of surveyors produce 10,000 updates a day into the database which supplies OS MasterMap, resulting in a vast yet granular geospatial dataset.²⁸ OS states on its website, “We’ve mapped the location of every fixed physical object in Great Britain, from the ground upwards, to within one metre accuracy.” In total, OS has logged 460 million individual geographical features, including over 35 million buildings and over 27 million residential addresses.²⁹

OS OpenData

Following the broader shift toward open data across government culture in the U.K., and to comply with a direct mandate from the Brown administration, OS launched its first open data platform, OS OpenData, in 2010, giving the public free access to more OS mapping than ever before.³⁰ The data, “made available via the website for viewing, download or order, had almost 50,000 orders placed in

26 GovLab interview with Andrew Loveless, Commercial Director, Ordnance Survey, October 1, 2015.

27 GovLab interview with Andrew Loveless, Commercial Director, Ordnance Survey, October 1, 2015.

28 “We’re all about location.” Ordnance Survey. <https://www.ordnancesurvey.co.uk/about/overview/what-we-do.html>

29 “We’re all about location.” Ordnance Survey. <https://www.ordnancesurvey.co.uk/about/overview/what-we-do.html>

30 GovLab interview with John Carpenter, Director of Strategy and Robert Andrews, Head of Corporate Communications, Ordnance Survey, September 16, 2015.

2013 alone.”³¹ These data sets and open maps have scales as detailed as 1:10,000 inches,³² which is usually adequate for the average citizen’s or mapping enthusiast’s needs.³³ The OS OpenData platform includes an open data map viewer, downloadable CSV files, the OS OpenSpace API to embed and annotate maps, downloadable software to help users manipulate the data, along with an OS community forum and support center to provide guidance and answer questions.³⁴ As of 2015, OS OpenData has had over a million downloads.³⁵

The range of OS OpenData products includes 16 data sets covering various spatial data: street-level and vector mapping; place names, road numbers and postcodes; water networks; road networks; U.K. administrative and electoral boundaries. In addition, OS OpenData provides customizable vector, terrain and environmental data, which allows users to layer multiple data sets onto maps and conduct their own spatial analysis.³⁶

However, not all OS data or mapping products are accessible as open data products in OS OpenData. For instance, OS’ flagship product, OS MasterMap, is conspicuously absent. The data sets provided for free are less detailed than the data sets available for purchase or licensing. It is worth noting that, prior to OS OpenData, OS did allow developers to access some data for free in order to create or test new products, provided that they did not commercialize them. Otherwise, users would have to pay for access and to license the data.³⁷ The reason for the disparity begins with OS’ roots as a

31 “Our history.” Ordnance Survey. <https://www.ordnancesurvey.co.uk/about/overview/history.html>

32 Wainwright, Martin. “Ordnance Survey maps are worth paying for.” *The Guardian*. April 3, 2010. <http://www.theguardian.com/commentisfree/2010/apr/03/ordnance-survey-maps-online>

33 GovLab interview with John Carpenter, Director of Strategy and Robert Andrews, Head of Corporate Communications, Ordnance Survey, September 16, 2015.

34 <https://www.ordnancesurvey.co.uk/business-and-government/products/opendata-products-grid.html>

35 GovLab interview with Andrew Loveless, Commercial Director, Ordnance Survey, October 1, 2015.

36 “OS OpenData.” Ordnance Survey. <https://www.ordnancesurvey.co.uk/business-and-government/products/opendata-products.html>

37 GovLab interview with John Carpenter, Director of Strategy and Robert Andrews, Head of Corporate Communications, Ordnance Survey, September 16, 2015.

Trading Fund and OS' self-financing requirement. Having historically met this requirement by selling access to its cutting-edge geo-spatial data and products, Neil Ackroyd, the OS chief operating officer, argues that it would be difficult to generate revenue if OS made all its data truly open, especially given recent austerity measures in the U.K.³⁸

Beginning in April 2015, OS has operated as a government-owned limited company. The Secretary of State for Business, Innovation and Skills (BIS) fully owns OS, and OS must follow a Crown Rights Agreement regarding copyright and licensing of OS data, in addition to its previously established self-financing requirement.³⁹ OS' status as a government-owned company also requires OS to pay tax, "as well as the annual dividend of about £20m that it returns to the Government."⁴⁰ To help balance the opposing forces of an open data mandate and a self-financing mandate, OS applies a "freemium model" to its products. That is, any data or products included in the OS OpenData portal are free to all users, but there is a fee to "upgrade" access and license OS' most accurate, advanced mapping data and products.⁴¹

OS came to this model by working with government stakeholders, who also helped identify which data would be the most valuable to users based on user analytics such as downloads and page visits,⁴² with an understanding that the new data ecosystem would have to be "quite sophisticated, driven by use cases and recognizing user value."⁴³ OS was and remains strategic in determining which data to open and in developing new products specifically for the open data

38 GovLab interview with Neil Ackroyd, Chief Operating Officer, Ordnance Survey, September 14, 2015.

39 "Governance and legal status." Ordnance Survey. <https://www.ordnancesurvey.co.uk/about/governance/index.html>

40 Cave, Andrew. "Why apps won't be ousting maps any time soon as Ordnance Survey chief charts a digital future." *The Telegraph*. October 18, 2015. <http://www.telegraph.co.uk/finance/11939740/Why-apps-wont-be-ousting-maps-any-time-soon-as-Ordnance-Survey-chief-charts-a-digital-future.html>

41 "Our public task." Ordnance Survey. <https://www.ordnancesurvey.co.uk/about/governance/public-task.html>

42 GovLab interview with John Carpenter, Director of Strategy and Robert Andrews, Head of Corporate Communications, Ordnance Survey, September 16, 2015.

43 GovLab interview with Neil Ackroyd, Chief Operating Officer, Ordnance Survey, September 14, 2015.

platform—for instance, if a gap in the commercial market is identified.⁴⁴ OS also commissioned studies to assess the cost of releasing these data sets to the public for free, and whether releasing the data sets would still create value through other means, such as promoting innovation (discussed further below).⁴⁵ Based on these discussions and assessments, the U.K. government agreed to a 10-year deal to compensate OS for this data, which was estimated to be worth £20m a year in lost sales and data management costs.⁴⁶

While the advanced data in premium products such as OS MasterMap is not typically used by average citizens, it is used heavily by many private sector entities (e.g., utilities, finance, real estate, etc.), other government agencies and scientists.⁴⁷ Therefore, since the launch of OS OpenData and this mixed-cost model, OS has focused on growing revenue from the private sector, rather than from taxpayers or other government agencies. At launch time, about 50 percent of revenue came from the private sector. Today, it is closer to 60 percent of revenue.⁴⁸ For public sector customers in England and Wales OS has a collective agreement called the Public Sector Mapping Agreement (PSMA).⁴⁹ On behalf of the nearly 4,000 members of the PSMA, as of October 2015, the British government funds OS (about £55m a year) to supply mapping data to nearly all England

44 GovLab interview with John Carpenter, Director of Strategy and Robert Andrews, Head of Corporate Communications, Ordnance Survey, September 16, 2015.

45 GovLab interview with Neil Ackroyd, Chief Operating Officer, Ordnance Survey, September 14, 2015.

46 Ballard, Mark. “Ordnance Survey accused of stifling competition in open data row.” *Computer Weekly*. November 14, 2014. <http://www.computerweekly.com/news/2240234712/Ordnance-Survey-accused-of-stifling-competition-in-open-data-row>; GovLab interview with James Norris, Policy Analyst, Ordnance Survey, February 3, 2015.

47 It is worth noting, however, that academics and researchers usually have an OS use contract for “almost free,” if not free. GovLab interview with John Carpenter, Director of Strategy and Robert Andrews, Head of Corporate Communications, Ordnance Survey, September 16, 2015.

48 GovLab interview with Andrew Loveless, Commercial Director, Ordnance Survey, October 1, 2015.

49 “Who can join the PSMA?” Ordnance Survey. <https://www.ordnancesurvey.co.uk/business-and-government/public-sector/mapping-agreements/psma-who-can-join.html>

and Wales government agencies, national and local.⁵⁰ A similar collective agreement, known as the One Scotland Mapping Agreement is in place in Scotland.⁵¹

John Carpenter, OS director of Strategy, describes the policy question behind the mixed model: “Is there a happy medium where those who have big pockets and will profit ... pay for it, but those who won’t profit, or are just starting out, don’t have to pay, or pay very little?”⁵² OS argues that since its fees are an “incidental cost for most big businesses”⁵³ and high volume users, such as a large utility company, there is not a strong argument for providing such advanced, highly maintained (and therefore costly) data for free.⁵⁴

As an example of this model, one of the most popular open OS products is Code-Point Open, which includes 1.7 million postcode units with an average of 15 adjoining addresses.⁵⁵ This enables users to integrate postcode searches onto their own maps to add features such as route-planning or simple “find my nearest” searches. However, data sets providing a higher level of detail and additional features, such as PO Box indicators, are only available through the paid version, Code Point.⁵⁶ As described in following sections, this mixed model arrangement helps OS meet its self-financing requirement, but it creates significant tension between OS and the broader open

50 Ballard, Mark. “Ordnance Survey accused of stifling competition in open data row.” *Computer Weekly*. November 14, 2014. <http://www.computerweekly.com/news/2240234712/Ordnance-Survey-accused-of-stifling-competition-in-open-data-row>; GovLab interview with James Norris, Policy Analyst, Ordnance Survey, February 3, 2015.

51 “One Scotland Mapping Agreement: Licensing.” Ordnance Survey. <https://www.ordnancesurvey.co.uk/business-and-government/public-sector/mapping-agreements/osma-licensing.html>

52 GovLab interview with John Carpenter, Director of Strategy and Robert Andrews, Head of Corporate Communications, Ordnance Survey, September 16, 2015.

53 GovLab interview with John Carpenter, Director of Strategy and Robert Andrews, Head of Corporate Communications, Ordnance Survey, September 16, 2015.

54 GovLab interview with John Carpenter, Director of Strategy and Robert Andrews, Head of Corporate Communications, Ordnance Survey, September 16, 2015.

55 “Code-Point Open: Benefits and key features.” Ordnance Survey. <https://www.ordnancesurvey.co.uk/business-and-government/products/code-point-open.html#Benefits-and-key-features>

56 “Code-Point Open: Benefits and key features.” Ordnance Survey. <https://www.ordnancesurvey.co.uk/business-and-government/products/code-point-open.html#Benefits-and-key-features>

data community—tensions that OS hopes will be mitigated by recent efforts outlined in detail below.

This tension with the open data community does not, however, extend to the vast majority of OS customers and partners. Despite customers having to pay to “upgrade” their data access, OS has over 350 business partners, and Loveless notes that OS consistently scores “very, very highly in our customer surveys,” based on their range of customer service analytics and evaluations, including scoring, random sampling or advisory groups and panels to learn specifics about customer experiences, perceptions and areas for improvement. For example, a March 2015 report finds that OS meets or exceeds customer service target levels in areas such as answering phone calls to Customer Services within 15 seconds (98.8 percent), responding to all written enquiries or responding to written complaints within 10 working days (99.9 percent and 99.7 percent), and fulfilling digital data orders from business partners within five days (98.9 percent).⁵⁷ Loveless further notes: “As an organization, [customer relationships and feedback] is something we are continually improving and we’re pretty restless in thinking through and improving at every opportunity we can.”⁵⁸

Impact

Economic Impact

From March 2011 - March 2012, OS commissioned a study to assess the broader economic impact of releasing OS data through the new OS OpenData platform. The study, conducted by firms Consulting-Where and ACIL Tasman, uses the Computable General Equilibrium model, chosen because it “overcomes problems with simple benefit/multiplier approaches” and allows “analysis of changes in macroeconomic aggregates and in potential changes to government

⁵⁷ “Our customer service performance.” Ordnance Survey. <https://www.ordnancesurvey.co.uk/about/governance/customer-service-performance.html>

⁵⁸ GovLab interview with Andrew Loveless, Commercial Director, Ordnance Survey, October 1, 2015.

revenues.”⁵⁹ The report focuses on the costs and savings at Ordnance Survey related to releasing and disseminating open data; costs and benefits among users in accessing and using the products; and the wider economic and social impacts resulting from the release of OS OpenData. The report was independently reviewed by members of the Advisory Panel on Public Sector Information to confirm its validity.⁶⁰

The report concluded that making the nine OS OpenData products available in 2012 free at the point of entry would directly improve the level of productivity in the economy and higher overall levels of output. Specific estimates calculated throughout the report include:

⁶¹

- OS OpenData will directly deliver a net £13 million - £28.5 million increase in GDP by 2016. The main components of this increase are net productivity gains (£8.1 million - £18.2 million) and additional tax revenues (£4.4 million - £8.3 million).
- The GDP increase is net of £3.7 million per year, applied as a negative to U.K. exports to account for OS OpenData being integrated into products of companies paying taxes abroad. Despite this loss of export income, overall the value of exports to the economy increases by £6.1 million - £10.3 million as other sectors of the economy expand.
- The increased exports will enable U.K. residents to purchase more foreign goods, increasing real imports by £3.6 million - £7.1 million.

⁵⁹ Carpenter, John and Phil Watts. “Assessing the Value of OS OpenData to the Economy of Great Britain—Synopsis.” 2013. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/207692/bis-13-950-assessing-value-of-opendata-to-economy-of-great-britain.pdf

⁶⁰ Carpenter, John and Phil Watts. “Assessing the Value of OS OpenData to the Economy of Great Britain—Synopsis.” 2013. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/207692/bis-13-950-assessing-value-of-opendata-to-economy-of-great-britain.pdf

⁶¹ Carpenter, John and Phil Watts. “Assessing the Value of OS OpenData to the Economy of Great Britain—Synopsis.” 2013. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/207692/bis-13-950-assessing-value-of-opendata-to-economy-of-great-britain.pdf

- Real national disposable income (real GNP) will increase £10.2 million - £24.1 million by 2016, indicating an increase in the economic welfare for British society as a whole.
- Real consumption will increase £8.1 million - £20.3 million and real investment will increase by £2.3 million - £5.1 million as result of OS OpenData.

OS is currently determining how best to measure economic impacts in more specific terms, for example based on each data set or product. However, quantifying the impact of open data, particularly geo-spatial data, is challenging because spatial data typically is embedded within larger applications, products or business processes. As Ed Parsons, geospatial technologist of Google and former CTO of OS puts it, “I often characterize it as it’s being like a gene in the DNA sequence. Lots of organisms might well have this gene and it might well be quite important but it’s just facing all elements of lots of other things. It often is quite hard to measure the incremental benefit of making this data openly accessible.”⁶² John Carpenter, OS director of Strategy, echoes this sentiment by explaining, “We know there are a lot of people using [our data]—we don’t know exactly how, or how they are profiting from it—but they acknowledge that they’re using it, and a lot are not paying for it.”⁶³

OS Minecraft

In 2013, OS aimed to build excitement and promote learning tools around OS OpenData by tapping into one of the most active and popular video game communities around, Minecraft, a mapping game set in custom-built worlds comprising blocks that can be mined for raw materials.⁶⁴ The OS Minecraft map of Great Britain, made with 22 billion blocks representing all 224,000 square kilometers of the U.K., was downloaded over 100,000 times within months

⁶² GovLab interview with Ed Parsons, Geospatial Technologist, Google, former CTO of Ordnance Survey, September 25, 2015.

⁶³ GovLab interview with John Carpenter, Director of Strategy and Robert Andrews, Head of Corporate Communications, Ordnance Survey, September 16, 2015.

⁶⁴ GovLab interview with John Carpenter, Director of Strategy and Robert Andrews, Head of Corporate Communications, Ordnance Survey, September 16, 2015.

of release,⁶⁵ going on to win a Guinness World Record as the largest real-world place represented in Minecraft.⁶⁶ Following the launch of OS Minecraft, demand indicators for other OS data increased significantly. OS views the surprising success of this initiative as a positive indicator of the enthusiasm surrounding mapping and data among the general public, and as a creative example of how it can continue building interest in its products and services, particularly among those who may be new to spatial data manipulation.⁶⁷

GeoVation Program

To more directly support innovation, in 2010 OS launched its GeoVation program, which runs challenges to “address specific problems within communities which may be satisfied, in part, through the use of geography.”⁶⁸ OS collaborates with the data and information innovation companies and organizations, Open Geospatial Consortium, Springwise, Nominet and Esri, to run and promote the GeoVation program. Since launching, GeoVation Challenges have awarded £718,000 in funding 31 new ventures,⁶⁹ and addressed varied social impact questions such as “How can Britain feed itself? How can we improve transport in Britain? How can we help British business improve environmental performance?” Winners include:

- Carbon Profit: which combines soil testing and OS mapping data to enable farmers to monitor carbon levels in their soil.⁷⁰
- The Green Alchemist: uses OS mapping to show the location of businesses, local waste carrier services and reprocessing companies to improve recycling and waste management.⁷¹

⁶⁵ Ward, Mark. “Minecraft map of the UK upgraded to include houses.” *BBC News*. September 24, 2014. <http://www.bbc.com/news/technology-29330265>

⁶⁶ “Largest real-world place created in Minecraft.” *Guinness World Records*. September 23, 2014. <http://www.guinnessworldrecords.com/world-records/117557-largest-real-world-place-created-in-minecraft>

⁶⁷ GovLab interview with Neil Ackroyd, Chief Operating Officer, Ordnance Survey, September 14, 2015.

⁶⁸ <https://www.ordnancesurvey.co.uk/innovate/geovation/>

⁶⁹ <https://theodi.org/blog/guest-post-five-years-and-1-million-downloads-the-os-opendata-journey>

⁷⁰ <http://www.arcarbon.co.uk/about-arcarbon/>

⁷¹ Page, Richard. “Using the GeoVation Hub as a GeoVation winner.” *Geovation*. July 20, 2015. <https://www.geovation.org.uk/using-geovation-hub-geovation-winner/#more-11068>

- Growing Routes: helps businesses map and identify opportunities on the Wales Coast Path, bringing new jobs and opportunities to local residents.⁷²
- Democratising Development: uses OS and Land Registry data to identify small-scale disused/unmanaged sites that can be brought into the housing market.⁷³
- MyHome Energy Planner: allows homeowners to measure current house energy usage and prescribe a set of costed retrofit measures to reduce this usage.⁷⁴

Other Impacts

Since launching, OS OpenData has received over a million downloads.⁷⁵ OS collects user analytics such as number of downloads, volume of downloads, number of data sets accessed and number of users on API downloads. This user behavior information, in addition to its customer satisfaction analytics, drives future strategy and product development.⁷⁶

While more precise impact measures are under consideration, there are broader, albeit less tangible, social, process or human indicators of OS OpenData's impact. Some examples, which indicate the wide variety of uses (and impacts) resulting from OS' using open data, include:

- IDC Consulting, an environment and sustainable energy consulting firm, uses OS OpenData to identify optimal locations and terrain for renewable energy sites such as wind and solar

⁷² "Growing Routes for business." GeoVation. July 22, 2014. <https://www.geovation.org.uk/growing-routes-for-business/>

⁷³ "Winners of the GeoVation Housing Challenge." GeoVation. March 12, 2015. <https://www.geovation.org.uk/winners-of-the-geovation-housing-challenge/>

⁷⁴ "Winners of the GeoVation Housing Challenge." GeoVation. March 12, 2015. <https://www.geovation.org.uk/winners-of-the-geovation-housing-challenge/>

⁷⁵ Henderson, David. "Guest post: Five years and 1 million downloads—Ordnance Survey's open data journey." Open Data Institute. March 26, 2015. <https://theodi.org/blog/guest-post-five-years-and-1-million-downloads-the-os-opendata-journey>

⁷⁶ GovLab interview with John Carpenter, Director of Strategy and Robert Andrews, Head of Corporate Communications, Ordnance Survey, September 16, 2015.

farms.⁷⁷ For such energy sites, location is critical to success, and by integrating OS data into its process, IDC can more quickly and accurately identify potential sites at a wide scale.⁷⁸

- Local fire department West Midlands Fire Service used OS OpenData (and other OS mapping) to develop an integrated open-source solution to manage all its mapping data holdings, reducing the cost, and increasing the efficiency and ease of mapping, to offer live-mapping capabilities to assist in decision-making directly at an incident.⁷⁹
- Engineering company CH2M HILL used OS OpenData to create a cost-effective automated tool to draw a multimodal bus network from any part of Great Britain, reducing the time taken to analyze an entire county bus network by 90 percent.⁸⁰
- Citrus Analytics, a customer and marketing analytics firm, used OS OpenData to improve its data, offering the ability to visualize customers on a map and to be used as the basis of qualitative research groups, for example, accurately and efficiently identifying if there are geographic clusters of certain types of customers.
⁸¹
- Rubicon Heritage Services Ltd, an archeology firm, used OS OpenData to enable background mapping in their survey equipment, providing real-time verification that the primary data is being collected correctly, consistently and accessibly.⁸²

⁷⁷ GovLab interview with John Carpenter, Director of Strategy and Robert Andrews, Head of Corporate Communications, Ordnance Survey, September 16, 2015.

⁷⁸ “Open for Business?” IDC Consulting. March 24, 2015. <http://idconsulting.net/2015/03/open-for-business/>

⁷⁹ “OS mapping ‘makes huge impact’ on fire safety.” Ordnance Survey. 2015. <https://www.ordnancesurvey.co.uk/business-and-government/case-studies/west-midlands-fire-service-psma-os-opendata.html>

⁸⁰ “Unlocking the potential of open data in transport planning.” Ordnance Survey. 2015. <https://www.ordnancesurvey.co.uk/business-and-government/case-studies/halcrow-ch2mhill-unlocking-opendata-transport.html>

⁸¹ “Visualising customer data to provide valuable insights.” Ordnance Survey. 2015. <https://www.ordnancesurvey.co.uk/business-and-government/case-studies/citrus-analytics-visualising-customer-data-valuable-insights.html>

⁸² “OS OpenData mapping significantly improves cost and time efficiency.” Ordnance Survey. 2015. <https://www.ordnancesurvey.co.uk/business-and-government/case-studies/rubicon-heritage-opendata-mapping-improves-efficiency.html>

- UIA, a mutual insurance company, used OS OpenData to set more accurate prices, drilling down to individual buildings rather than zip codes, making the underwriting process quicker and more accurate, thereby saving the company money.⁸³
- T4 Media, an advertising firm, used OS OpenData to automate the process of identifying suitable outdoor advertising spaces, allowing employees to analyze multiple points from their desks.⁸⁴
- Local county council for Warwickshire used OS OpenData and other OS products to develop an open-source solution for internal Web geographical information system (GIS), allowing the GIS team to make department-specific projects, saving over 65 percent in license and maintenance costs, improving control over data performance issues, and an up-front saving of £15 - £20,000 by developing a custom in-house system.⁸⁵
- Many health care providers use OS OpenData, with other OS products, to better pinpoint patient locations and routes, enabling more accurate evidence-based analysis and decision-making to deliver efficiency savings and improved services.⁸⁶
- OS is celebrated for supporting outdoors enthusiasts by providing accurate, detailed maps, for example, recently promoting an OS Photo Contest for users to submit photos taken on their adventures, helping others find some of the most beautiful spots.⁸⁷

⁸³ “UIA: Reducing poor risks and improving customer service.” Ordnance Survey. <https://www.ordnancesurvey.co.uk/business-and-government/case-studies/uia-reducing-poor-risks.html>

⁸⁴ “Putting the local into location-based advertising to target specific audiences.” Ordnance Survey. 2015. <https://www.ordnancesurvey.co.uk/business-and-government/case-studies/t4-media-putting-the-local-into-location.html>

⁸⁵ “Warwickshire County Council’s new web GIS increases users with no extra cost.” Ordnance Survey. 2015. <https://www.ordnancesurvey.co.uk/business-and-government/case-studies/warwickshire-county-council-new-web-gis.html>

⁸⁶ “Pinpointing patients and practice areas to improve services and efficiency.” Ordnance Survey. 2015. <https://www.ordnancesurvey.co.uk/business-and-government/case-studies/nhs-swcs-pinpointing-patients-practice-areas.html>

⁸⁷ “Using OS OpenData to explore top Peak District photo spots.” Ordnance Survey Blog. September 7, 2015. <http://www.ordnancesurvey.co.uk/blog/2015/09/using-os-opendata-to-explore-top-peak-district-photo-spots/>

Challenges

Cost

The greatest challenge to scaling OS OpenData is cost. As Andrew Loveless, OS commercial director, explains: “Someone has to pay for data at some stage, and as we see in the marketplace, it’s not actually the release of data, it’s the maintenance of data,” as in regularly updating, cleaning and running quality assurance tests, particularly at the high levels of OS’ current standards, that is costly. Since measuring the impact of open data is so difficult, “How does that value get realized? Open knows no global boundaries—is the beneficiary, from an OS perspective, is it a U.K. entrepreneur? Is it the U.K. system? Or is it international organizations? I think the concept of a full open data model needs to be fully thought through and then obviously fully funded.”⁸⁸

OS views the current mixed model, as justified by the 2011 - 12 economic impact study, as the best option for balancing the conflicting demands of opening data and self-financing. OS argues that the freemium model offers a sustainable way of dealing with the economic realities it operates in, and that it is designed to “meet the needs of various customer segments including the public, businesses and government,” and that startup businesses in particular are supported by their innovation activities.⁸⁹ More stakeholder assessments and impact studies should be conducted in order to continually reassess whether this remains the best option, particularly as other mapping services such as Google Maps and OpenStreetMap (described below) continue to grow and offer a free or possibly less expensive alternative to currently paying OS customers. Still, it is important to note that Google Maps and some other mapping and satellite navigation companies also use OS OpenData to some extent.⁹⁰

⁸⁸ GovLab interview with Andrew Loveless, Commercial Director, Ordnance Survey, October 1, 2015.

⁸⁹ GovLab interview with Robert Andrews, Head of Corporate Communications, Ordnance Survey, October 28, 2015.

⁹⁰ Cave, Andrew. “Why apps won’t be ousting maps any time soon as Ordnance Survey chief charts a digital future.” *The Telegraph*. October 18, 2015. <http://www.telegraph.co.uk/finance/11939740/Why-apps-wont-be-ousting-maps-any-time-soon-as-Ordnance-Survey-chief-charts-a-digital-future.html>

OS would also have to assess whether it would be able to provide such quality data in the first place if it lost such revenue. The U.S. for example, historically has always offered geospatial data for free, albeit with less funding. The quality and scope of that data, however, may not realistically be up to par with OS, which updates a third of the country, minimum, every year, in a highly detailed, near real-time database—i.e., the factors that likely create such a high demand for OS data in the first place.⁹¹

Addressing Open Data Principles

To many of those active in the geospatial and open data communities, OS is claiming rights to “the ultimate public good” of *where we are* by claiming copyright on spatial information. This raises the question asked by Peter Rabley of Omidyar Network: “What is your unique position here? How on earth can you copyright that, the actual coordinates, which is the mathematical measurement from the center of the earth?”⁹² [Disclosure: Omidyar Network funded this case study.] *The Guardian’s* “Free Our Data Campaign” specifically calls out OS for asking taxpayers to pay for access to data that OS, being a government agency, already collected directly on their behalf through taxes, arguing that besides being unfair, this “stifles innovation, enterprise and the creativity that should be the lifeblood of new business.”⁹³ However, as discussed further below, the ongoing fees for OS data go toward addressing needs like maintenance and dissemination, not the actual geospatial data collection.⁹⁴

⁹¹ GovLab interview with John Carpenter, Director of Strategy and Robert Andrews, Head of Corporate Communications, Ordnance Survey, September 16, 2015.

⁹² GovLab interview with Peter Rabley, Director, Investments, Omidyar Network, September 23, 2015.

⁹³ Arthur, Charles and Michael Cross. “Give us back our crown jewels.” *The Guardian*. March 9, 2006. <http://www.theguardian.com/technology/2006/mar/09/education.epublic>

⁹⁴ GovLab interview with Andrew Loveless, Commercial Director, Ordnance Survey, October 1, 2015.

One illustration of the community's frustration with OS licensing and copyright hurdles is the rise of OpenStreetMap (OSM), a volunteer-driven, crowdsourced mapping initiative founded in the U.K. in 2004 originally as a direct response to OS' lack of open data.⁹⁵ According to a report from *National Geographic*, OSM adheres to true open innovation and open data principles, being a "Web-based project that aggregates hundreds of users' personally collected GPS data into master files that trace out thousands of streets and byways."⁹⁶ The OpenStreetMap Foundation, founded in 2006, focuses its services on public sector, nonprofit and humanitarian uses without profit.⁹⁷ It is worth noting, again, that OSM incorporates OS OpenData into its data sets in some instances.

OSM exemplifies the demand for freely available, accurate mapping for all users, regardless of ability to pay. As Alyssa Wright of OSM explains, location data represents "a shared experience" that is "so fundamental to understanding who people are, how they relate, where they're going to go. ... Mapping is universal need."⁹⁸ Therefore, many in the open data community believe that geospatial data should be treated no differently than other government data. For example, OSM and open innovation supporters argue that, "When programmers can read, redistribute and modify the source code for a piece of software, the software evolves. People improve it, people adapt it, people fix bugs. And this can happen at a speed that, if one is used to the slow pace of conventional software development, seems astonishing."⁹⁹ OSM attempts to replicate this same kind of open innovation process in geospatial data by freely offering its

95 "History of OpenStreetMap." OpenStreetMap. http://wiki.openstreetmap.org/wiki/History_of_OpenStreetMap

96 Anderson, Mark. "Global Positioning Tech Inspires Do-It-Yourself Mapping Project." *National Geographic News*. October 18, 2006. <http://news.nationalgeographic.com/news/2006/10/061018-street-maps.html>

97 GovLab interview with Alyssa Wright, President of the U.S. Board for OpenStreetMap, September 25, 2015.

98 GovLab interview with Alyssa Wright, President of the U.S. Board for OpenStreetMap, September 25, 2015.

99 Anderson, Mark. "Global Positioning Tech Inspires Do-It-Yourself Mapping Project." *National Geographic News*. October 18, 2006. <http://news.nationalgeographic.com/news/2006/10/061018-street-maps.html>

mapping data through an open license and encouraging users to add and manipulate spatial data.¹⁰⁰

Furthermore, Charles Arthur and Michael Cross of *The Guardian* point to the U.S. as a positive example for always offering its spatial data as free and open, noting that “It is no accident that it is also the country that has seen the rise of multiple mapping services (such as Google Maps, Microsoft’s MapPoint and Yahoo Maps) and other services—mashups—that mesh government-generated data with information created by the companies.”¹⁰¹ Given that OS’ own economic impact study found that releasing some data would be positive for the economy, it naturally raises the question as to whether releasing more data would be even more beneficial to the economy, and to spurring innovation.

In response to these critiques, OS argues that the existence of intellectual property rights in OS geospatial data (via copyright and database rights in U.K. law), is “accepted by the HMSO [Her Majesty’s Stationery Office] Controller from who we take delegation to licence the data. The open data community also would appear to implicitly accept this by promoting open licences.”¹⁰² Furthermore, in October 2015, the Court of Justice of the European Union (CJEU) supported this position in a similar case between Verlag Esterbauer, an Austrian publishing company specializing in tour map books, concerning an application for a cease-and-desist order under the German Law on copyright and related rights.¹⁰³ The CJEU found that “geographical information extracted from a topographic map, used by a third party to produce and market another map, retains sufficient informative value, which in turn constitutes ‘independent materials’ of a database.”^{104,105}

100 https://wiki.openstreetmap.org/wiki/Main_Page

101 Arthur, Charles and Michael Cross. “Give us back our crown jewels.” *The Guardian*. March 9, 2006. <http://www.theguardian.com/technology/2006/mar/09/education.epublic>

102 GovLab interview with Robert Andrews, Head of Corporate Communications, Ordnance Survey, October 28, 2015.

103 “Case C-490/14, Verlag Esterbauer—a map is not a database.” *EU Law Radar*. July 8, 2015. <http://eulawradar.com/case-c-49014-verlag-esterbauer-a-map-is-not-a-database/>

104 GovLab interview with Robert Andrews, Head of Corporate Communications, Ordnance Survey, October 28, 2015.

105 “Case C-490/14, Verlag Esterbauer—a map is not a database.” *EU Law Radar*. July 8, 2015. <http://eulawradar.com/case-c-49014-verlag-esterbauer-a-map-is-not-a-database/>

OS also contends that these open data principles—i.e., public data should be universally and readily accessed, used and redistributed free of charge—apply to data sets, not organizations or business models, and that OS’ open data (e.g., in OS OpenData) scores highly against these principles. Furthermore, in the nine years since the “Free Our Data” campaign launched, many of the campaign’s grievances have been addressed with the release of OS OpenData, licensing simplification and amendments to the GeoVation program.¹⁰⁶

Private Sector Competition

The tension between OS and the open data community carries over to the private sector, as many businesses believe that they could reduce costs and grow even more quickly if the data in OS MasterMap, for example, and not just OS OpenData were free. Although OS OpenData offers data on a scale of 1:10,000, Ed Parsons of Google says that the “real juice is in the 1:5,000, 1:2,500, 1:1,000 or 1:500 [scales]. The larger the scale gets, the smaller the area, and the more valuable the data sets.”¹⁰⁷ Critics describe OS as a monopoly, as there is no competitor to offer larger scales, lower pricing or greater efficiency, which allows OS to set their own prices while making it easier to claim they only charge exactly what their services cost.¹⁰⁸ They complain that, currently, trying to access certain OS data sets and embed OS products within your own programs can mean entering into a very bureaucratic process, with a lot of forms to fill, followed by licensing fees.¹⁰⁹ However, it is important to note that as a public sector body, all OS accounts are fully audited.

Some businesses also accuse OS of stifling innovation and actively trying to prevent private sector competitors from succeeding. Exemplifying this sentiment is a 2014 complaint filed by aerial photography and mapping company Getmapping. The complaint argues that through its contracts with the U.K. government (the Public Sector

¹⁰⁶ GovLab interview with Robert Andrews, Head of Corporate Communications, Ordnance Survey, October 28, 2015.

¹⁰⁷ GovLab interview with Ed Parsons, Geospatial Technologist, Google, former CTO of Ordnance Survey, September 25, 2015.

¹⁰⁸ GovLab interview with Peter Rabley, Director, Investments, Omidyar Network, September 23, 2015.

¹⁰⁹ GovLab interview with Ed Parsons, Geospatial Technologist, Google, former CTO of Ordnance Survey, September 25, 2015.

Mapping Agreement and the compensation agreement at initial release), OS illegally uses £800m of government contracts to stifle competition.¹¹⁰ Getmapping chairman Tristram Cary has called these contracts a “national scandal” because “Ordnance Survey has been granted these contracts, which has made it highly profitable—much more profitable than most industry companies. It is using its state aid to compete with the rest of the partner network. Ordnance Survey is a serial abuser of its partners. And the government knows it.”¹¹¹ Cary, who has seen previous legal complaints against OS rejected, still argues that by investing in its own research and development, OS has become “far too big and is competing with British technology companies instead of, as they are supposed to do, promoting them. It is also hopelessly inefficient.”¹¹² OS refutes Cary’s claims and, as of November 2015, there has been no judgment in this case. Also, as noted above, the CJEU recently ruled that there can be intellectual property rights in geospatial data.¹¹³

In *The Guardian’s* “Free Our Data” campaign, Cary argues that OS being an “almost monopoly” in the provision of large-scale maps allows OS to aggressively enforce copyright fees, even on derived products.¹¹⁴ For example, if a utility uses an OS map to draw its pipeline map, OS will indefinitely enforce a fee unless the utility re-creates their own data. This “acts as an effective barrier to the creation of a truly competitive market” because customers, such as the utility in this case, are locked in, and because other map suppli-

¹¹⁰ Milmo, Cahal. “Has the Ordnance Survey lost its moral compass?” *The Independent*. November 28, 2014. <http://www.independent.co.uk/travel/news-and-advice/has-the-ordnance-survey-lost-its-moral-compass-9891770.html>

¹¹¹ Milmo, Cahal. “Has the Ordnance Survey lost its moral compass?” *The Independent*. November 28, 2014. <http://www.independent.co.uk/travel/news-and-advice/has-the-ordnance-survey-lost-its-moral-compass-9891770.html>

¹¹² Milmo, Cahal. “Has the Ordnance Survey lost its moral compass?” *The Independent*. November 28, 2014. <http://www.independent.co.uk/travel/news-and-advice/has-the-ordnance-survey-lost-its-moral-compass-9891770.html>

¹¹³ GovLab interview with Robert Andrews, Head of Corporate Communications, Ordnance Survey, October 28, 2015.

¹¹⁴ “Tristam Cary of Getmapping responds on the OS consultation: ‘British mapping is fundamentally flawed.’” *The Guardian Free Our Data Blog*. March 2013. <http://www.freedata.org.uk/blog/2010/03/tristram-cary-of-getmapping-responds-on-the-os-consultation-british-mapping-is-fundamentally-flawed/>

ers cannot obtain any derived data without also paying OS a fee and negotiating licensing terms.¹¹⁵

Critics also argue that being an effective monopoly inherently discourages OS from investing in new research and development, new technologies or in new processes with the same vigor as private businesses. Although OS invests in research and development, by being a closed organization, it can be argued that these efforts do not match the kind of innovation that occurs through more competitive or crowdsourced models. They argue that this disadvantage will only increase, as advances in geospatial technology give rise to other sources of equally accurate and timely spatial data, at a lower cost than OS. Rabley offers that OS needs to be careful and ask, “What’s our core value to the U.K. taxpayer?”¹¹⁶ That is, OS shouldn’t be trying to own a market and compete with businesses with new products and services, but rather OS should focus on how it enables businesses to be innovative, which is through its impressive geospatial data. Therefore, Rabley argues, more data should be opened, and *then* businesses will be their “biggest support in ensuring that Ordnance Survey continues, whereas right now, the complete opposite is the case. The businesses want to get rid of and shut down Ordnance Survey.”¹¹⁷

These challenges of cost, adhering to open data principles and navigating private sector competition highlight the larger policy issue as to whether or not taxpayers should be funding government agencies to build developer resources, or if OS should focus more on supporting outside experts to build their own products, using OS data, as needed.¹¹⁸ OS therefore, should consider what its relationship should be with the open data community and with the private sector, particularly as potential competitors continue to emerge with

¹¹⁵ “Tristam Cary of Getmapping responds on the OS consultation: ‘British mapping is fundamentally flawed.’” *The Guardian Free Our Data Blog*. March 2013. <http://www.freeourdata.org.uk/blog/2010/03/tristam-cary-of-getmapping-responds-on-the-os-consultation-british-mapping-is-fundamentally-flawed/>

¹¹⁶ GovLab interview with Peter Rabley, Director, Investments, Omidyar Network, September 23, 2015.

¹¹⁷ GovLab interview with Peter Rabley, Director, Investments, Omidyar Network, September 23, 2015.

¹¹⁸ GovLab interview with Ed Parsons, Geospatial Technologist, Google, former CTO of Ordnance Survey, September 25, 2015.

rivaling technology, and whether aggressively enforcing licensing fees is worth the risk to its relationships, and whether its actions do in fact stifle innovation and competition. In response, OS argues that it does prioritize these considerations, and that as a member of the Open Data Institute (ODI), OS regularly engages with the open data community on policy issues.¹¹⁹ Moreover, OS has been accredited to the Information Fair Traders Scheme since 2003.¹²⁰

Measuring Impact and User Demands

As described, OS is working toward developing more precise impact and use indicators in the future, in addition to its current sophisticated user surveys and analytics. Through these efforts, John Carpenter, OS director of Strategy, described two key learning points: 1) it is challenging for some users to discover what data is available in the first place; and 2) simply publishing the data is not enough. As Carpenter puts it: “It has to be discoverable, usable, and in our case, mapping data is not terribly easy to use. We’ve had to put a lot of effort to make it easier to use, and have to invest to get people to use it.”¹²¹

The future priorities of OS are “going to be much more about getting what’s out there *now* used vs. getting *more* data out there to be used” as OS continues to progress beyond the mindset of “if you put it out there, they will come.”¹²² While customer service teams and user forums help address these issues, OS has also launched several larger initiatives to not only help users learn more about using and navigating the data, but also to spur engagement among existing and new users, including open data classes, blogs, wikis and the development of open standards, among others. For example, OS also sponsored an Open Data Camp in February 2015 for 200 developers,

¹¹⁹ GovLab interview with Clare Hadley, Policy and Engagement Manager, and James Norris, Policy Analyst, Ordnance Survey, October 23, 2015.

¹²⁰ GovLab interview with James Norris, Policy Analyst, Ordnance Survey, February 3, 2015; “IFTS members.” The National Archives. <http://www.nationalarchives.gov.uk/information-management/re-using-public-sector-information/ifts-and-regulation/ifts-members/>

¹²¹ GovLab interview with John Carpenter, Director of Strategy and Robert Andrews, Head of Corporate Communications, Ordnance Survey, September 16, 2015.

¹²² GovLab interview with John Carpenter, Director of Strategy and Robert Andrews, Head of Corporate Communications, Ordnance Survey, September 16, 2015.

innovators and entrepreneurs to test new OS data to create fresh insights and innovative products and services.¹²³ Carpenter describes the approach to promoting innovation as, “This is for life, not just Christmas. [We] have to go big or don’t bother; a half-hearted approach is a waste of time for everybody.”¹²⁴

Looking Forward

Evolving GeoVation Hub

To continue scaling excitement and impact of the GeoVation program, in March 2015, OS opened a new GeoVation Hub in London. Cathrine Armour of OS’ Innovation Programme, described, “One of the biggest challenges facing the geospatial industry and all those operating in it is: Innovating. ... To increase the relevance of our industry we need to expose ourselves to new thinking, to undertake thought leadership and share this in a way that as an industry we lift our gaze. The Hub will enable OS and its partners to work collaboratively with fringe industries, new media, digital and emerging technologies.”¹²⁵

OS OpenData Additions

In 2015, OS also announced additions to its OS OpenData platform, including:

- OS OpenMap: “a ‘street level’ vector dataset designed to be the most detailed open data mapping product available, providing a backdrop for integrating and visualising analytical data.”¹²⁶

¹²³ Scott, Anna. “X marks the spot: Ordnance Survey and Esri open up more geospatial data.” Open Data Institute. February 21, 2015. <https://theodi.org/blog/broadening-horizons-ordnance-survey-esri-put-open-data-on-their-maps>

¹²⁴ GovLab interview with John Carpenter, Director of Strategy and Robert Andrews, Head of Corporate Communications, Ordnance Survey, September 16, 2015.

¹²⁵ “Ordnance Survey to open hub dedicated to innovation.” Ordnance Survey. February 27, 2015. <http://www.ordnancesurvey.co.uk/about/news/2015/geospatial-innovation-hub-announced.html>

¹²⁶ “Britain leads the way as Ordnance Survey helps to drive economic growth and digital innovation through open data.” Ordnance Survey. February 21, 2015. <http://www.ordnancesurvey.co.uk/about/news/2015/os-drives-economic-growth-and-digital-innovation-through-open-data.html>

- A new Open Water Network: “a generalised network product covering Great Britain’s rivers which will deliver a national view of our watercourses.”¹²⁷
- A new and improved Gazetteer: for use by people who need the most up-to-date place names, road names, road numbers and postcodes for use in location searches for both mobile and online uses.”¹²⁸

Looking Beyond the U.K.

OS has been expanding the marketplace for its data services outside of the U.K. For example, in 2013, OS International LLP, a subsidiary of Ordnance Survey, provided geospatial services and strategy consulting to the Kingdom of Bahrain.¹²⁹ OS created the first 3D-enabled national spatial data model of the Kingdom of Bahrain to provide government and industry stakeholders with consistent, intelligent and reliable 2D and 3D data to support “data sharing, analysis and decision-making” using “intelligent geospatial data.”¹³⁰ In April 2015, OS developed a GIS Roadmap Framework for The Environment Agency of Abu Dhabi to help the agency establish and manage a “single source of accurate, reliable environmental information,” share geospatial information with stakeholders, set goals and improve the “consistency and quality of geospatial data capture, processing, analysis and dissemination” of its environmental and spatial data.¹³¹

¹²⁷ “Britain leads the way as Ordnance Survey helps to drive economic growth and digital innovation through open data.” Ordnance Survey. February 21, 2015. <http://www.ordnancesurvey.co.uk/about/news/2015/os-drives-economic-growth-and-digital-innovation-through-open-data.html>

¹²⁸ “Britain leads the way as Ordnance Survey helps to drive economic growth and digital innovation through open data.” Ordnance Survey. February 21, 2015. <http://www.ordnancesurvey.co.uk/about/news/2015/os-drives-economic-growth-and-digital-innovation-through-open-data.html>

¹²⁹ “About OS International.” Ordnance Survey. <https://www.ordnancesurvey.co.uk/international/about/index.html>

¹³⁰ “Creating a 3D data model for the Kingdom of Bahrain.” Ordnance Survey. <https://www.ordnancesurvey.co.uk/international/case-studies/creating-3d-data-model.html>

¹³¹ “Using geospatial support to safeguard the environment for future generations.” Ordnance Survey. <https://www.ordnancesurvey.co.uk/international/case-studies/using-geospatial-support-to-safeguard-environment.html>

Supporting Smart Cities

With the rise of smart cities—cities that use innovative technology (i.e., sensors, meters, applications) to create more efficient and interactive infrastructure and information systems—OS is examining the role of geospatial data and products in driving the continued development of smart cities, including: providing connectivity between assets, networks and sensor data; sharing and integrating data services; data analytics in smart energy, water and transport management; and data visualizations and public engagement.¹³² Recently, OS has been working with Glasgow to deliver a new trial license so that data produced using OS' maps will be made available through a new open data portal for Glasgow's "Future City/Glasgow" initiative. In particular, OS is focusing on opportunities in smart grids, health/assisted living, intelligent transport, smart water management and smart water management for Future City/Glasgow.¹³³

Additionally, OS is supporting the Digital Greenwich initiative by, among other things, enabling the borough's greater use of predictive analytics and "providing support for machine-to-machine data services" to allow Internet of Things (IoT) devices to more efficiently connect to each other.¹³⁴

These recent efforts illustrate that OS values innovation, and suggests that in addition to maintaining high-quality data and impressive service levels for its current customers and government partners, OS may be able to mitigate concerns voiced by members of the open data community by focusing more on supporting outside actors in their own development and growth. However, the challenges to scaling, particularly the self-funding requirement, and the strong criticisms of its mixed model and its fierce copyright protections, will need to be reassessed, particularly if OS, as an organization, hopes someday to embody the principles of open data fully.

¹³² "Smart Cities." Ordnance Survey. <https://www.ordnancesurvey.co.uk/about/thinking/smart-cities/>

¹³³ "Smart Cities." Ordnance Survey. <https://www.ordnancesurvey.co.uk/about/thinking/smart-cities/>

¹³⁴ "Greenwich Smart City Strategy." Digital Greenwich. 2014.

CHAPTER 13

United States' New York City Business Atlas

Leveling the Playing Field

*By Andrew Young, David Sangokoya and
Stefaan Verhulst*

Summary: While retail entrepreneurs are experts in their respective trades, they often lack access to high-quality information about economic conditions in the neighborhoods where they operate or are considering operating. The NYC Business Atlas is designed to alleviate that information gap by providing a public tool that gives small businesses access to high-quality data to help them decide where to establish a new business or expand an existing one. The tool brings together a diversity of data, including business-filing data from the Department of Consumer Affairs, sales tax data from the Department of Finance, demographic data from the census and traffic data from Placemeter, a New York City startup focusing on real-time traffic information.

Dimension of Impact: Creating Opportunity—Economic Growth

Key Takeaways:

- The impact of open data can be amplified when government works directly with private business on targeted initiatives. Such “data collaboratives” represent a new form of collaboration, beyond the public-private partnership model, in which participants from different sectors—including private companies,

research institutions and government agencies—can exchange data to help solve public problems.¹

- Although a large number of early stage open data projects around the world focus on simply pushing information out, the next stage should revolve around targeted, user-centered release. In the example discussed here, the user-centered work done by the Department of Small Business Services helped ensure that the Business Atlas was designed in a way that made it particularly useful for the New York small business community.
- New York City's Mayor's Office of Data Analytics (MODA) provides an example of how governments can go beyond providing data in raw formats to the public by also performing the analytical work needed for those within and outside government to gain new insights from the data.

Context and Background

In recent years there has been growing recognition that urban life is being transformed by data. From Chicago to London to Singapore, city administrators and planners are turning to data to help plan the future and address mundane, everyday issues like potholes and waste collection. Underlying such trends is an awareness of the vast amounts of data being generated (often passively) in urban centers, through devices like smartphones and sensors. In the words of *The Economist* magazine, cities today are “open air computers” and “data factories.”²

In 2002, Michael Bloomberg assumed office as the 108th mayor of New York City. Bloomberg had made his fortune providing data and sophisticated analytics to financial traders. It was probably inevitable that, under his administration, New York would join the many

1 Verhulst, Stefaan and David Sangokoya. “Data Collaboratives: Exchanging Data to Improve People’s Lives,” *Medium*, April 22, 2015. <https://medium.com/@sverhulst/data-collaboratives-exchanging-data-to-improve-people-s-lives-d0fcfc1bdd9a>.

2 “Open-air Computers,” *The Economist*, October 27, 2012. Accessed July 14, 2015. <http://www.economist.com/news/special-report/21564998-cities-are-turning-vast-data-factories-open-air-computers>.

cities around the world seeking to extract greater value from the terabytes of data being created every day by their citizens.

In 2013, through Executive Order No. 306, New York City created the Mayor's Office of Data Analytics (MODA).³ The stated goal of the office was "leveraging City data for more effective, efficient, and transparent government."⁴ Today the office comprises a team of analysts, based in City Hall, who collect and analyze data from a wide variety of sources. Among other areas, MODA works on crime prevention, disaster response, improving public services, and economic development. MODA also played a key role in setting up New York's Open Data Portal (<https://nycopendata.socrata.com>), which, as of early 2016, houses over 12,000 data sets related to health, business, public safety and much more. In addition, MODA helped establish DataBridge, a single, unified repository of information that aims to enhance data sharing and interoperability among various NYC organizations.⁵ In July 2015, the city released its updated "Open Data for All" strategy document, which focuses on two central "beliefs": that every New Yorker can benefit from open data; and that open data can benefit from every New Yorker.⁶

3 Exec. Order No. 306, 3 C.F.R. (2013), http://www.nyc.gov/html/om/pdf/eo/eo_306.pdf.

4 Flowers, Michael. "NYC by the Numbers Annual Report." New York City Government. December 2013. http://www.nyc.gov/html/analytics/downloads/pdf/annual_report_2013.pdf.

5 Yasin, Rutrell. "How analytics is making NYC's streets and buildings safer." GCN, October 4, 2013. <http://gcn.com/articles/2013/10/04/gcn-award-nyc-databridge.aspx>.

6 "Open Data for All" New York City Government. 2015. <http://www1.nyc.gov/assets/home/downloads/pdf/reports/2015/NYC-Open-Data-Plan-2015.pdf>

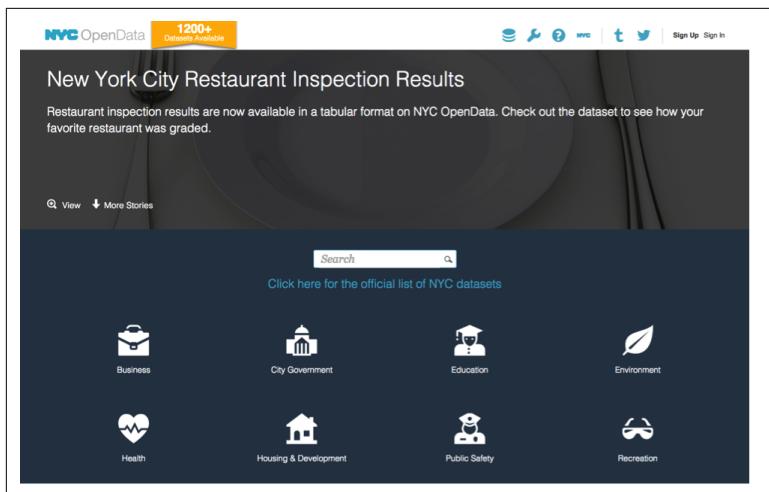


Figure 13-1. New York City’s Open Data Portal

Because of these and other efforts, New York City is generally considered a leader in open data initiatives in the United States (itself the second-ranked country on the Open Data Barometer).⁷ MODA, in particular, is a pioneering and increasingly emulated entity in the open data ecosystem. It has played an important role not only in releasing open data to increase accountability and innovation, but also in doing analytical work on that data. This work includes measuring the efficiency of city services, providing data-driven predictions and, as with the Business Atlas, combining high-value data sets from a diversity of sources to provide new insights and visualizations to government agencies and the public.

MODA’s analytics efforts, led at the time of writing by New York City’s Chief Analytics Officer, Dr. Amen Ra Mashariki, are deployed to aid disaster response and recovery, improve delivery of city agencies and services, enable data sharing among city agencies, crystallize best practices in data analysis and, as evidenced in the case described here, spur economic development.⁸ Within these domains of focus and types of analytical work, MODA subscribes to four cen-

⁷ <http://barometer.opendataresearch.org/report/analysis/rankings.html>

⁸ Flowers, Michael. “NYC by the Numbers Annual Report.” New York City Government. December 2013. http://www.nyc.gov/html/analytics/downloads/pdf/annual_report_2013.pdf.

tral, overarching goals: improving awareness, measuring success, maximizing impact and increasing engagement.⁹

Mike Flowers, Mashariki's predecessor and the first New York City chief analytics officer, described the integral role played by MODA in the city's data operations in an annual report released after MODA's first year of operations: "Over the last three terms, our agencies have developed information systems that they use to make our streets safer, our businesses vibrant, and our parks cleaner. Through a blend of statistical analysis, engineering skills, and deep investigation of the missions and organizational structure of the agencies—the why, what and how of city government—MODA ties these systems together, enabling the City to tap into our collective knowledge and experience to tackle our thorniest challenges."¹⁰

According to MODA officials, its mission and projects are focused on the "thorny" challenges at hand for New Yorkers while its efforts are using new analytical capabilities. Lindsay Mollineaux, director of analytics at MODA, for instance, notes that, "[Addressing real need] is very much how we think about things at MODA—every project is addressing need. We want to make sure what we're doing is useful."¹¹

Project Description and Inception

The New York City Business Atlas, initiated in 2013, is part of a broader effort by MODA aimed at "driving small business growth with analytics."¹² This broader effort also includes the Comprehensive Business Census, which arose in the aftermath of Superstorm Sandy, when the city struggled to assess the storm's full impact on businesses and the economy. Before MODA began working in this area, there existed no comprehensive record of businesses in the

9 "NYC MODA Presentation Federal Summit." New York City: NYC Analytics, February 2015. http://lnwprogram.org/sites/default/files/NYC_MODA_Presentation_Federal_Summit.pdf

10 Flowers, Michael. "NYC by the Numbers Annual Report." New York City Government. December 2013. http://www.nyc.gov/html/analytics/downloads/pdf/annual_report_2013.pdf.

11 GovLab interview with Lindsay Mollineaux, Director of Analytics, Mayor's Office of Data Analytics, New York City, July 2, 2015.

12 Flowers, Michael. "NYC by the Numbers Annual Report." New York City Government. December 2013. http://www.nyc.gov/html/analytics/downloads/pdf/annual_report_2013.pdf.

city. MODA sought to fill this information gap by working with PLUTO, a database of land use and geographic data, to assemble a more “complete picture” of businesses and business activity in New York City.

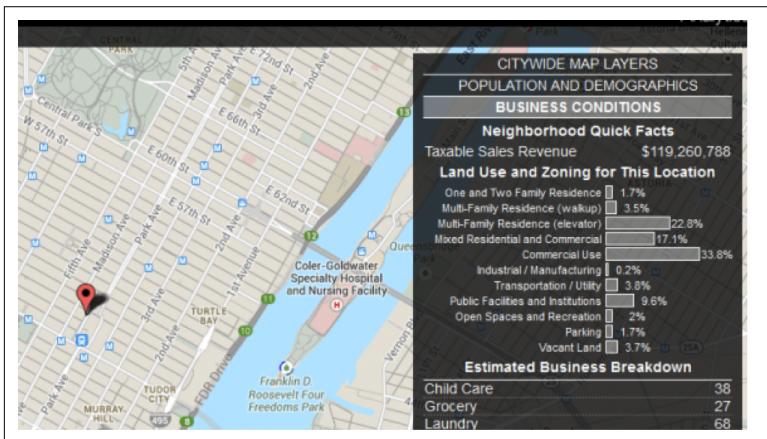


Figure 13-2. New York City Business Atlas with Business Conditions Sidebar

The New York City Business Atlas grew out of a recognition among city officials that, when it comes to data, large businesses often have an edge over smaller ones. While large businesses can afford to hire expensive consultants and commission data-driven research, smaller businesses must rely on “gut feeling” to make important business decisions, such as where to open a new location or how to navigate regulatory challenges. Mike Flowers explains the advantages held by large businesses as follows: “In many parts of Manhattan, you can’t swing a dead cat without hitting a Starbucks. Those guys have robust infrastructure, capacity to help them figure out two things: a) where to open up in the first place; and b) the piece of this about navigating the regulatory challenges of opening a place.”¹³ He adds that, for small businesses in particular, data paucity is a “chronic” problem, and “probably [has] been chronic since Emperor Augustus was trying to incentivize small business in Rome.”¹⁴

¹³ GovLab interview with Mike Flowers, former Chief Analytics Officer, Mayor’s Office of Data Analytics, New York City, August 14, 2015.

¹⁴ GovLab interview with Mike Flowers, former Chief Analytics Officer, Mayor’s Office of Data Analytics, New York City, August 14, 2015.

Though followed by a number of twists and turns, work on the Business Atlas began following a discussion within MODA on how small business owners often feel under siege from the city government, rather than supported by it. Flowers noted that the city restaurant rating system, which assigns letter grades to restaurants based on their compliance with health regulations, tends to benefit large chains and restaurants, which typically have “the wherewithal and institutional experience and institutional resources to bake into their infrastructure code compliance.”¹⁵ The Business Atlas represents a pivot away from the specific inspiration regarding code compliance, but remains in line with the focus on arming small business owners with new tools to compete with bigger chains.¹⁶

In the words of John Feinblatt, chief policy adviser to Mayor Bloomberg, the Business Atlas “democratizes” information, “putting quality research into the hands of small business owners.”¹⁷ It is important to note that much (but not all) of the data included in the Business Atlas already existed—for example, through the city’s open data portal—and was theoretically at least available to small business owners. As noted, however, it was often in fragmentary form, and without the sophisticated analytics and visualization layer contained within the Business Atlas, both of which make the data far more accessible and useful for entrepreneurs. To use the tool, businesspeople visit maps.nyc.gov/businessatlas, and select a neighborhood. The data pulled up by the app includes population, population distribution by age, median household income, how many households have children, homeowners vs. renters, and much more specific to

15 GovLab interview with Mike Flowers, former Chief Analytics Officer, Mayor’s Office of Data Analytics, New York City, August 14, 2015.

16 In a separate initiative, MODA is collaborating with the New York City Small Business Services “to reduce the regulatory burden on small businesses in New York City” by sending “trained client managers to neighborhoods across the five boroughs to provide business owners with targeted education informed by data showing the specific needs in a given neighborhood—including top violations, areas of noncompliance, new business growth data, 311 complaints, and more.”

“Small Business Services Launches Proactive Education and Outreach Efforts to Help Ease Regulatory Burden on New York City Small Businesses.” New York City Small Business Services. May 27, 2015. http://www.nyc.gov/html/sbs/html/pr/2015_05_27_SB1.shtml

17 Schweidel, D. A. *Profiting from the data economy: Understanding the roles of consumers, innovators, and regulators in a data-driven world* (Upper Saddle River, NJ: Pearson Education).

that neighborhood. Not only is the Business Atlas itself free to use, but users can also sign up for free training sessions, held in city business centers, that will help them derive the most from the tool.¹⁸

One of the platform's most important pieces of data is the foot traffic in various neighborhoods. To collect this information, New York partnered with a local startup, Placemeter, a self-styled "urban intelligence platform."¹⁹ Placemeter uses cameras (including existing municipal street traffic cameras and sensor-laden IP cameras) to assess population movement through neighborhoods. The resulting information includes both pedestrian and vehicular traffic data. While much of the analytical work is done algorithmically, Placemeter also relies on humans to analyze videos and perform random quality checks of the work being done by the algorithms.²⁰ The resulting data gives businesspeople an indication of prospective customer numbers, thus helping guide location-relevant business decisions.²¹ The city also has plans to supplement the data using crowdsourced information. Although an important piece of the Business Atlas, Placemeter's work on quantifying public spaces could lead to privacy concerns directed at the city down the road. That said, Placemeter has taken concrete steps to mitigate those concerns by: a) processing video in real time so that less than 0.01 percent of all video is recorded or stored—and only for processing and quality assurance purposes; and b) only providing anonymized counts of pedestrians, with no specific identities attached. Former U.S. Deputy Chief Technology Officer Nicole Wong also acts as a privacy adviser to the company.²²

In addition to Placemeter data, the Atlas includes data pulled from a variety of government departments and agencies. These include the Department of Consumer Affairs, the Department of Finance (e.g.,

18 Furman, Phyllis. "Map this! New city tech tool lets small businesses compete with the big guys by dishing data." *New York Daily News*. December 16, 2013. <http://www.nydailynews.com/new-york/map-new-city-tech-tool-lets-small-businesses-compete-big-guys-dishing-data-article-1.1559044>

19 GovLab interview with Florent Peyre, Chief Operating Officer and Co-Founder, Placemeter, August 19, 2015.

20 GovLab interview with Florent Peyre, Chief Operating Officer and Co-Founder, Placemeter, August 19, 2015.

21 "Platform - How It Works." Placemeter. <https://www.placemeter.com/platform>

22 "Privacy by Design." Placemeter. <http://www.placemeter.com/privacy>

sales tax information), and demographic data from census results. The Atlas supplements this data with information shared from the Department of Health and Mental Hygiene (DOHMH), Business Integrity Commission (BIC), Department of Environmental Protection (DEP), Department of City Planning (DCP) and Department of Buildings (DOB) in New York City, as well as state and national open data.²³ In many cases, MODA's task involved combining and analyzing data sets that were already open and accessible to the public. In other cases, additional effort was required from MODA in order to secure the release of data. Sales tax data from the Department of Finance, for example, is protected due to the inclusion of personally identifiable information. In order to include the data in the Atlas, MODA first had to strip away personal information through an anonymization process.²⁴

“Some of the needed data in designing the Atlas was obvious to us, but the question was what is useful to entrepreneurs versus information overload?”

—Lindsay Mollineaux, Mayor’s Office of Data Analytics

In order to combine all this data in a single place, the team creating the Atlas had to overcome several technical and conceptual challenges. For example, while as much data as possible was pulled from the city’s DataBridge (described above), there were inevitable issues concerning the compatibility of data sets. Differences between data standards and formats create a major and time-consuming challenge in the effort to combine multiple data streams into one usable tool. In addition, finding accurate data for local businesses proved more challenging than anticipated. As Mollineaux explained, each industry has its own particular licensing regulations (and some businesses, for example bookstores, have no licensing requirements at all), making it difficult to accurately represent and synthesize local business information across sectors.²⁵

²³ “NYC MODA Presentation Federal Summit.” New York City: NYC Analytics, February 2015. http://lnwprogram.org/sites/default/files/NYC_MODA_Presentation_Federal_Summit.pdf

²⁴ GovLab interview with Mike Flowers, former Chief Analytics Officer, Mayor’s Office of Data Analytics, New York City, August 14, 2015.

²⁵ GovLab interview with Lindsay Mollineaux, Director of Analytics, Mayor’s Office of Data Analytics, New York City, July 2, 2015.

User-Centered Design and Partnering with Department of Small Business Services

Although the general value proposition of the NYC Business Atlas was clear from the outset, MODA decided to partner with the New York City Department of Small Business Services (SBS) to ensure that the needs of small businesses (its target audience) were truly met. As Mollineaux noted: “Some of the needed data in designing the Atlas was obvious to us, but the question was what is useful to entrepreneurs versus information overload? SBS served as our subject matter experts who interfaced with actual entrepreneurs (for example, people might come to them about opening a bakery) and could use the Atlas to directly serve these needs. ... We always partner with client agencies who are subject matter experts and can help define what success looks like.”²⁶

Through ethnographic research and interviews, SBS was able to help MODA determine what was most relevant for various types of users. For example, MODA originally focused on displaying some of the business and demographic information as a score for a given geographic location. User feedback, gathered with the assistance of SBS, helped MODA to recognize that, in fact, entrepreneurs would be more interested in less-aggregated data; most businesspeople wanted the underlying data rather than a one-size-fits-all score.²⁷ Rather than a simple score or grade, therefore, the data is now mapped in a disaggregated, “raw” form, allowing users to draw their own conclusions.

In addition to partnering with SBS to supplement the informational base of the platform, MODA partnered with the New York City library system to drive use. Research had indicated to Flowers and his team that many entrepreneurs rely on their local library to gain insight into how to start a new business. With this audience of potential users in mind, MODA worked with and trained library

²⁶ GovLab interview with Lindsay Mollineaux, Director of Analytics, Mayor’s Office of Data Analytics, New York City, July 2, 2015.

²⁷ GovLab interview with Lindsay Mollineaux, Director of Analytics, Mayor’s Office of Data Analytics, New York City, July 2, 2015.

staff to introduce the platform to potential entrepreneurs and essentially serve as “small business counselors.”²⁸

Overall, MODA’s approach of partnering with different agencies and institutions has proven remarkably fruitful. According to Flowers, it is part of a well-thought-out strategy to ensure the longevity of the Business Atlas. As Flowers puts it: “You have to get the civil service on board. ... If you don’t have them on board as major participants, then in the next election everything you worked on is gone.”²⁹

Impact

Like many data-driven urban projects around the world, the Business Atlas has benefited from the existence of copious amounts of data and a relatively sophisticated user base that is well-informed and aware about the potential of open data. This conducive ecosystem has translated into tangible—and almost immediate—impact for the project’s intended beneficiaries.

Intended Beneficiaries

Entrepreneurs and Small Business Owners

- Community that will make the most direct use of data housed in the Business Atlas.
- Improved decision-making capabilities engendered through free access to market research data and analytics that would normally come at great cost.
- Evidence of market opportunities provided by the Atlas can be useful for securing financing and investment for new businesses.

²⁸ GovLab interview with Mike Flowers, former Chief Analytics Officer, Mayor’s Office of Data Analytics, New York City, August 14, 2015.

²⁹ GovLab interview with Mike Flowers, former Chief Analytics Officer, Mayor’s Office of Data Analytics, New York City, August 14, 2015.

Citizens of New York City

- By providing small business entrepreneurs with decision-making support, the Business Atlas seeks to enable the creation of new jobs across New York City.
- As a result of the new businesses opened based on Business Atlas insights, consumers should see an influx of new businesses targeted to the needs of their communities.
- In particular, residents of traditionally underserved neighborhoods in the city stand to benefit from new businesses being opened in their area as a result of a greater understanding of community-level needs and opportunities.

Leveling the Playing Field for Market Research

One of the most important impacts of the Business Atlas is the way in which it levels the playing field between large and small businesses. MODA's 2013 annual report points out that, "When a major national retailer looks to open a new storefront, they often commission sophisticated neighborhood market research that helps the company decide where to locate the new business."³⁰ That type of research is often too expensive for smaller businesses. But the Business Atlas, as John Feinblatt, former Mayor Bloomberg's chief policy adviser, puts it, "democratizes ... research, putting quality research into the hands of small business owners."³¹

Even when small businesses do have access to data (for example, through public feeds or other sources), they may lack the analytical skills to process and understand it. Here, too, the Business Atlas plays a powerful role, its sophisticated analytics and visualization tools further leveling the playing field between larger and smaller players. At a meeting of municipal chief data officers, Amen Ra Mashariki, New York City's chief analytics officer, pointed out the many ways in which such data and analysis can empower small businesses. He cited the example of an entrepreneur approaching a bank for a loan. With the information contained within the Business Atlas, the entrepreneur can make a far more compelling case,

³⁰ Neubauer, Miranda. "With Business Atlas, NYC Analytics Office Looks to 2014." *Tech-President*. January 2, 2014. <http://techpresident.com/news/24635/business-atlas-nyc-analytics-office-looks-2014>

³¹ Krasny, Jill. "NYC Data Tool Gives Small Business an Edge." *Inc.* December 27, 2013. <http://www.inc.com/jill-krasny/nyc-business-atlas-helps-small-businesses.html>

backed by real evidence, for the sustainability and potential of the business.³²

Enabling Business Improvement District (BID) Analysis

A further impact of the Business Atlas is evident in SBS' intentions to deploy the Atlas for its own work. Currently, SBS is planning to use data contained within the Atlas to aid in the analysis of how Business Improvement Districts (BIDs) are fueling economic growth in New York. BIDs are public-private partnerships "in which property and business owners elect to make a collective contribution to the maintenance, development and promotion of their commercial district."³³ The data now in place, thanks to the Business Atlas, will allow SBS to compare BID neighborhoods in terms of economic change, commercial investments and business activities; this will in turn permit SBS to identify which BIDs have been most impactful to date and develop best practices to replicate their success across the city.

The deployment of the Business Atlas to fuel the growth of BIDs points to another particular community that stands to benefit from new availability of market research data: residents of underserved NYC neighborhoods. As Mashariki has pointed out, "city agencies can also use Business Atlas to address large businesses and show them that there is good reason for them to open locations in neighborhoods which they may have otherwise avoided."³⁴ Instead of making location-based decisions based purely on intuition (or media-driven biases), companies can now take a closer look at data and find underserved areas that offer a compelling business case. This is just one more way in which the Business Atlas holds poten-

32 "Towards Data Driven Cities? Meet up with Chief Data Officers." Proceedings of La Fabrique De La Cite, Paris. March 23, 2015. [http://www.lafabriquedelacite.com/fabrique-de-la-cite/data.nsf/FDD3CB2E8CEA41D2C1257E0F00324482/\\$file/actes_cdo_02062015_def_web.pdf](http://www.lafabriquedelacite.com/fabrique-de-la-cite/data.nsf/FDD3CB2E8CEA41D2C1257E0F00324482/$file/actes_cdo_02062015_def_web.pdf).

33 New York City Department of Small Business Services. "Starting a Business Improvement District: A Step-by-Step Guide." Nyc.gov. 2003. http://www.nyc.gov/html/sbs/downloads/pdf/bid_guide_complete.pdf.

34 "Towards Data Driven Cities? Meet up with Chief Data Officers." Proceedings of La Fabrique De La Cite, Paris. March 23, 2015. [http://www.lafabriquedelacite.com/fabrique-de-la-cite/data.nsf/FDD3CB2E8CEA41D2C1257E0F00324482/\\$file/actes_cdo_02062015_def_web.pdf](http://www.lafabriquedelacite.com/fabrique-de-la-cite/data.nsf/FDD3CB2E8CEA41D2C1257E0F00324482/$file/actes_cdo_02062015_def_web.pdf).

tial to level the playing field—for consumers as well as for businesses.

Influencing Data Analytics Innovation in New York and Abroad

As with many of the examples included in this report, MODA’s work has had important ripple effects, spurring the development of other similar open data projects. Flowers notes that the Business Atlas “certainly has this burst through the wall capacity to show that open data can mean a lot more than simply building a Yelp app.”³⁵ Recently, for example, the Fire Department of New York (FDNY) set up an analytics unit modeled on MODA’s own analytics team. The efforts of the FDNY team include the development and use of a Risk Based Inspection System (RBIS), which “enables the Department to identify buildings most at risk for fire and prioritizes those for fire inspections.”³⁶ Data is pooled using DataBridge from an FDNY data warehouse and other city databases including City Planning, Buildings and others.³⁷ In setting up the unit and its analytics platform, FDNY worked directly with MODA, providing an example of a constructive partnership and synergy across city departments.³⁸

Another example of MODA’s ripple effects within NYC can be found in a NYC Buildings Department project to manage complaints about illegal building conversions, a “311 City Pulse” program that live-feeds city 311 activities, and a data collection and sharing mechanism on disaster response.³⁹ All of these programs used lessons and principles that have been applied and tested by MODA.

³⁵ GovLab interview with Mike Flowers, former Chief Analytics Officer, Mayor’s Office of Data Analytics, New York City, August 14, 2015.

³⁶ “FireCast: Leveraging Big Data for Mitigating Fire Risks.” The Innovation Enterprise. January 22, 2014. <https://icondemand.com/divisions/big-data/events/4/presentations/firecast-leveraging-big-data-for-mitigating-fire-risks#sthash.FidXBzgc.dpuf>.

³⁷ Yasin, Rutrell. “How analytics is making NYC’s streets and buildings safer.” GCN, October 4, 2013. <http://gcn.com/articles/2013/10/04/gcn-award-nyc-databridge.aspx>.

³⁸ Flowers, Michael. “NYC by the Numbers Annual Report.” New York City Government. December 2013. http://www.nyc.gov/html/analytics/downloads/pdf/annual_report_2013.pdf.

³⁹ <http://gcn.com/articles/2014/02/19/new-york-city-geek-squad.aspx>

Other cities have also taken notice of New York's open data efforts. The London-based Capital City Foundation has suggested, for instance, that London should look to projects like the Business Atlas in its efforts to become a "smart city." In a recent report, the foundation argued: "If a business wanted to appeal to customers from certain parts of London, data from Transport for London (TfL) shows exactly where people touch in and touch out of the transport network. Maps can thereby be created showing where people move from and to. This could be helpful to know which tube or bus stops to place a business near. Creating an online tool to make these kinds of data sets available would build on ideas started in New York City."⁴⁰ It is likely that the appetite for similar data analytics teams will continue to spread, as the lessons learned and best practices from MODA are increasingly shared with innovators and policymakers around the world.⁴¹

Challenges

Communicating Opportunity

A tool is only useful if people actually use it. So while the Business Atlas presents a major opportunity for giving more businesspeople an understanding of the contexts in which they might consider opening a business, communicating that opportunity to the public will be an important ongoing challenge to ensure wide use. As Mike Flowers put it: The Business Atlas is part of an effort to give (a hypothetical) "Nadine's Burritos" the type of market research insights and capabilities that have been enjoyed by the likes of McDonald's and Subway for years.⁴² But making sure that Nadine—and thousands of small entrepreneurs like her—is aware of the availability of this information remains something of a challenge.

40 Coepland, Eddie. "Big Data in the Big Apple." Capital City Foundation. 2015. <http://capitalcityfoundation.london/big-data-in-the-big-apple-web-version/>.

41 Flowers, Mike and Lauren Talbot. "Building a Gov Data Skunkworks." Code for America. February 18, 2014. <http://www.codeforamerica.org/peer-network-training/02-18-2014/>.

42 GovLab interview with Mike Flowers, former Chief Analytics Officer, Mayor's Office of Data Analytics, New York City, August 14, 2015.

To that end, awareness-raising and the types of outreach already conducted with city libraries will be essential. In addition, Flowers believes that the creation of an application program interface (API) to enable developers to take the data housed on the Business Atlas and create new apps could also help disseminate that data more widely.⁴³

Addressing Technical Challenges

In order to achieve its many ambitions for growing the Business Atlas, MODA will also need to address a number of technical challenges that could serve as stumbling blocks. For example, as noted previously, different types of businesses typically have different licensing requirements. As Mollineaux points out, this is just one instance of a more general issue—the different “data contexts” that exist for different categories of businesses, and that make it challenging to pull together and meaningfully analyze data from disparate sources.⁴⁴

In the past, MODA has written a number of proprietary algorithms to overcome such difficulties. But challenges remain, and as the agency plans to add more data, they could possibly grow. Finding new ways to synthesize and harmonize large sources of data, pulled from different agencies and groups, is one of the key tasks confronting the agency as it seeks to expand its reach and scale up its efforts.

Improving Granularity of Data

The information contained on the Business Atlas has proven to be very helpful for businesspeople seeking to identify suitable (or unsuitable) neighborhoods in which to operate. But as Flowers points out, “a neighborhood in New York City is bigger than most towns in America.”⁴⁵ He adds that, to increase its usefulness, the Business Atlas could seek to provide a finer level of granularity in the analysis it provides its users. For example, it could move beyond

⁴³ GovLab interview with Mike Flowers, former Chief Analytics Officer, Mayor’s Office of Data Analytics, New York City, August 14, 2015.

⁴⁴ GovLab interview with Lindsay Mollineaux, Director of Analytics, Mayor’s Office of Data Analytics, New York City, July 2, 2015.

⁴⁵ GovLab interview with Mike Flowers, former Chief Analytics Officer, Mayor’s Office of Data Analytics, New York City, August 14, 2015.

neighborhood-level information and, perhaps, focus on areas that included five or 10 blocks.

Looking Forward

MODA has in many respects changed the way citizens and policy-makers in New York make decisions. The Business Atlas is just one example, albeit one with particularly significant potential. Given the early success of and positive response to Business Atlas, MODA has plans to scale it up and extend its reach in the coming years.

Business Atlas 2.0: New Tools and Features

Business Atlas 2.0 is the name being applied to a suite of new tools and enhancements MODA is currently planning to add to the original Business Atlas. Among the new features that may be included in the planned update:⁴⁶

- a feature that will allow businesspeople to compare multiple locations;
- more traffic information, including daily subway rider information;
- crowdsourced data on foot and vehicular traffic, with the intention of adding accuracy and completeness to Placemeter's traffic numbers; and
- a “solver” tool that will allow entrepreneurs to enter their needs or specifications and identify potentially suitable locations for their businesses.

MODA is also considering adding a predictive analytics tool that would identify businesses at risk of failure, and proactively target them by offering assistance from other small businesses. This “targeted assistance” would, for example, take the form of loan offers or other financial assistance.⁴⁷

⁴⁶ Flowers, Michael. “NYC by the Numbers Annual Report.” New York City Government. December 2013. http://www.nyc.gov/html/analytics/downloads/pdf/annual_report_2013.pdf.

⁴⁷ Flowers, Michael. “NYC by the Numbers Annual Report.” New York City Government. December 2013. http://www.nyc.gov/html/analytics/downloads/pdf/annual_report_2013.pdf.

Partnerships and Collaborations

In addition to rolling out new features, MODA also plans to enhance existing, and initiate new, partnerships designed to increase the reach and usefulness of the Business Atlas. As discussed above, MODA is already partnering with SBS to analyze the economic growth of Business Improvement Districts in New York. This partnership will be continued and extended, with a particular focus on identifying common traits and behaviors that are most conducive to economic growth. One of the goals is to identify a series of Best Practices that can guide economic growth in the city.

Additionally, MODA plans to continue and initiate partnerships with a number of outside agencies, institutions, businesses and individuals.⁴⁸ In the academic and research sector, MODA already partners with New York University's Center for Urban Science and Progress (CUSP), Columbia's Center for Data Science and the Rensselaer Polytechnic Institute. MODA's collaboration with Microsoft Labs is specifically focused on a project regarding automated responses to 311 SMS messages. MODA also works closely with average citizens through hackathons during which new projects are developed and feedback from users is collected.

Although much of the work with Placemeter occurred as a single, one-time collaboration, new work by the startup could lead to future opportunities to supplement the Business Atlas. The next big step for Placemeter “is the ability to actually measure in real time the speed of cars in your neighborhood, then gather that data for the benefit of your neighborhood and more widely than city officials.”⁴⁹ This could not only play a major role in Vision Zero—Mayor Bill de Blasio’s mission to reduce traffic deaths to zero in New York City—but also provide additional traffic flow information to businesses.

These and other changes are either planned or already underway. But many of the most important changes and additions in the years ahead remain unforeseen, and will likely arise directly from users. MODA’s “Open Data for All” strategy document outlines a concer-

48 Flowers, Michael. “NYC by the Numbers Annual Report.” New York City Government. December 2013. http://www.nyc.gov/html/analytics/downloads/pdf/annual_report_2013.pdf.

49 GovLab interview with Florent Peyre, Chief Operating Officer and Co-Founder, Placemeter, August 19, 2015.

ted effort to learn by watching how the site is used, and by whom.⁵⁰ What types of data and analytical tools are most useful? What aspects of the site seem to pose difficulties or represent friction for users? How do people discover the site, and what converts them into repeat visitors (as opposed to single-occasion users)? These are some of the questions MODA will be asking—and building on—as it moves forward.

⁵⁰ “Open Data for All” New York City Government. 2015. <http://www1.nyc.gov/assets/home/downloads/pdf/reports/2015/NYC-Open-Data-Plan-2015.pdf>

CHAPTER 14

United States' NOAA: Opening Up Global Weather Data in Collaboration with Businesses

Creating a New Industry Through Access to Weather Data

By Christina Rogawski, Stefaan Verhulst and Andrew Young

Summary: Opening up weather data through the United States National Oceanic and Atmospheric Administration (NOAA) has significantly lowered the economic and human costs of weather-related damage through more accurate forecasts; the development of a multi-billion-dollar weather derivatives financial industry; and the growth of a million-dollar industry of tools and applications derived from NOAA's real-time data. In many ways, the industry built around NOAA's weather data is seen as the paradigmatic example of how the release of open data can yield major economic impacts. To further scale the impact of its data, NOAA has launched the Big Data Project (BDP), which provides an opportunity to combine NOAA's tremendous volume of high-quality environmental data and advanced data products, private industry's vast infrastructure and technical capacity, and the U.S. economy's innovation and energy.

Dimension of Impact: Creating Opportunity—Economic Growth

Key Takeaways:

- The impacts of a given open data set can span an incredible array of sectors and users—with NOAA’s real-time data able to both help an individual decide whether or not to bring an umbrella on her commute or enable a farmer to better prepare for this season’s crop yield, to name just two.
 - NOAA collects over 3.5 billion weather observations per day. Around 96 percent of the U.S. public obtains 301 billion forecasts each year, providing \$31.5 billion in benefits, far exceeding the \$5.1 billion spent annually by both private and public weather bureaus on generating forecasts.
- These estimated benefits are available because of NOAA’s use of a variety of quantitative estimates of the value of its products and services using the Value of Information (VOI) methodologies. This focus on assessing the impacts of information on decisions and on the effect of those decisions on real-world outcomes sets NOAA apart from many other open data initiatives in terms of quantified impacts.
- Collaborations between public data providers and private industry actors can create new value and opportunities from open data.
- Crowdsourcing ideas for how to best make use of open data could help to ensure that data owners build fruitful partnerships and/or release data in ways that can create maximum value without requiring maximum resource investment.

Context and Background

The National Oceanic and Atmospheric Administration (NOAA) traces its roots to 1807 as the Coast and Geodetic Survey.¹ Over time, numerous other scientific and environmentally focused federal agencies formed. In 1970, President Nixon proposed the creation of NOAA to “unify the nation’s piece-meal environmental activities and provide a rational and systematic approach to understanding,

¹ “A History of NOAA.” NOAA History - NOAA Legacy/Agency History/NOAA History. http://www.history.noaa.gov/legacy/noaahistory_1.html

protecting, developing and enhancing the total environment.”² Today, NOAA is housed under the U.S. Department of Commerce, and its stated mission is “to understand and predict changes in Earth’s environment and conserve/manage coastal and marine resources to meet the nation’s economic, social and environmental needs.”³

With an annual budget over \$5 billion, NOAA deploys weather-monitoring satellites and high-tech instrumentation and sensors to collect large amounts of climate and environmental data.⁴ Today, its operating units include the National Weather Service (which conducts forecasts, and broadcasts severe weather watches and warnings), the National Environmental Satellite and Data Service (which operates satellites and the atmospheric data center), the National Ocean Service (which performs hydrographic, geodesy and maritime services), the National Marine Fisheries Service (which manages commercial fisheries and protects marine species) and the Office of Oceanic and Atmospheric Research (which conducts applied research).⁵

Data is at the heart of NOAA’s mission and central to virtually all its activities. As of 2014, “about 100 petabytes of environmental data currently are stored in NOAA data centers,” with another 30 petabytes added every year.⁶ For perspective, 1 petabyte = 1 million gigabytes; or about the equivalent of 20 million four-drawer filing cabinets filled with texts, or 13.3 years of HD-TV video. The “entire written works of mankind from the beginning of recorded history” is about 50 petabytes.⁷ Over 3.5 billion observations are collected

2 “A History of NOAA.” NOAA History - NOAA Legacy/Agency History/NOAA History. http://www.history.noaa.gov/legacy/noahistory_1.html

3 “Assessing the Economic and Social Benefits of NOAA Data.” NAS/OECD Conference. February 2008. <http://www.oecd.org/sti/ieconomy/40066192.pdf>

4 Chopra, Aneesh P. and Ethan Skolnick. *Innovative State: How New Technologies Can Transform Government.* (New York, NY: Atlantic Monthly Press), 2014.

5 “Assessing the Economic and Social Benefits of NOAA Data.” NAS/OECD Conference. February 2008. <http://www.oecd.org/sti/ieconomy/40066192.pdf>

6 “NOAA Looks for Advice to Make Its Data Easier to Use.” *Eos*, Vol. 95, No. 11. March 18, 2014. <http://onlinelibrary.wiley.com/doi/10.1002/2014EO110003/pdf>

7 Diaz, Jesus. “How Large Is a Petabyte?” *Gizmodo*. July 8, 2009. <http://gizmodo.com/5309889/how-large-is-a-petabyte>

each day⁸ from over 90 operational and research observing systems as well as 100 real- and near-real-time information systems, including: satellite data, regional monitoring (e.g., coastal, arctic), model data and archives, ocean profiles, climatology, etc.

The resulting information underlies huge swathes of economic, social and political life in America (and, indeed, around the world). Among other activities, they inform weather forecasts, climate predictions, ship and aircraft navigation, and conservation work involving marine populations. Thus NOAA's data helps guide decisions that are both momentous (e.g., how much electricity to produce on a given weekend) and more trivial (e.g., should I make outdoor plans this weekend?).

NOAA also plays a significant role in limiting losses from natural disasters by working with the Federal Emergency Management Agency (FEMA). In this partnership, NOAA doesn't simply provide data. It is also involved in the analysis of information, converting raw data into an "index of severity" that helps guide FEMA's planning for and responses to disasters.

NOAA's data is widely consumed by actors in both the private and public sectors. In funding and supporting NOAA, and especially in opening up its information systems to private and public actors, the United States government is performing a vital social, economic, political and cultural service that could not easily be replaced by private sector actors. The data that NOAA publishes is best handled by a public sector entity for a variety of reasons: a) it is non-exclusive (anyone can observe and record the environment); b) it is non-rival (one actor using that data does not subsequently make it any less useful to others); c) it requires a high cost of the infrastructure (expensive scientific equipment); and d) it is reproducible at near zero-marginal cost (for most purely information goods, once produced, the marginal cost of redistribution is close to zero; therefore they cannot be created and produced by firms that use revenues obtained from sales to cover costs. For firms to make revenue, there must be a value-added service, as seen in The Weather Channel reformatting NOAA data into concise, localized formats for TV,

⁸ Breggin, Linda and Judith Amsalem. "Big Data and Environmental Protection: An Initial Survey of Public and Private Initiatives." Environmental Law Institute. 2014. <https://www.eli.org/sites/default/files/eli-pubs/big-data-and-environmental-protection.pdf>

websites and mobile phones).⁹ As NOAA CIO Zach Goldstein describes, “It’s our job to get that data out there. The data doesn’t belong to us, it belongs to the American people.”¹⁰

Project Description and Inception

Since its inception, NOAA has boasted a strong open data culture and is considered a leader in open data, if not *the* leading open data example among government agencies. When the Obama administration launched data.gov as part of its flagship Open Government Initiative in January 2009, NOAA was cited as the paradigmatic example as to how government agencies can both publish data and make that data accessible for the private sector to use and build a multi-billion-dollar industry.¹¹

It follows that NOAA’s data products and services have always been user-driven, determined by demand among consumers and citizens who have a need for weather-related information. Demand has been particularly strong from certain sectors that are highly affected by weather events (e.g., energy, agriculture, water resources and emergency management). These sectors have in many ways helped shape the way NOAA provides data, and have been at the forefront of pushing the agency to open up its information pool and services. As NOAA Data Management Architect Jeff de La Beaujardiere explains, if NOAA measures something to predict the weather, users will get their weather forecast for tomorrow, but those observations might have other uses, like long-term forecasting for planting crops. “The broader societal objective is to make sure that data gets out there so that as many people and companies and stakeholders can get to them and make decisions based on the data, and make new information products based on the data they find.”¹²

⁹ Rifkin, Jeremy. “Capitalism is making way for the age of free.” *The Guardian*. March 31, 2014. <http://www.theguardian.com/commentisfree/2014/mar/31/capitalism-age-of-free-internet-of-things-economic-shift>

¹⁰ GovLab interview with Zachary Goldstein, Chief Information Officer, NOAA, September 3, 2015.

¹¹ GovLab interview with Zachary Goldstein, Chief Information Officer, NOAA, September 3, 2015.

¹² GovLab interview with Jeff De La Beaujardiere, Data Management Architect, NOAA, September 18, 2015.

In its early years, NOAA collected its data in a proprietary format, making it difficult for private users or companies to download and manipulate it independently. This primarily was due to technical limitations of the time. NOAA otherwise, culturally and policy-wise, has always prioritized making its data as accessible to users as possible.¹³ When NOAA first began developing a centralized open data portal in the early 2000s, its goal was to make the portal as technologically “cutting edge as possible.”¹⁴ In particular, the organization hoped to transition from being simply a clearing house for raw data to “actually a menu of html and formats to be pulled directly into applications,” in response to user demand.¹⁵

As advances in data technology proliferated, and the sophistication of users continued to increase, NOAA recognized the need for more standardized goals and policies regarding its relationship with users. In 2003, The National Research Council (NRC), the working arm of the U.S. National Academies, therefore conducted a study examining the respective roles of government, academic and private sectors in the weather industry. The study found that “advances in science and technology have blurred the distinctions between the sectors,” and provided recommendations on how “partnership can effectively move forward in an era of rapid advances.”¹⁶ The NRC identified the need for a policy that would “define processes for decision-making rather than defining roles.”¹⁷

13 GovLab interview with Tim Owen, Chief of Climate Information Services Division, NOAA, July 27, 2015.

14 GovLab interview with Tim Owen, Chief of Climate Information Services Division, NOAA, July 27, 2015.

15 GovLab interview with Tim Owen, Chief of Climate Information Services Division, NOAA, July 27, 2015.

16 “NOAA Issues New Partnership Policy: Response to ‘Fair Weather’ Recommendations Strengthens Relationships Among Government, Universities and the Private Sector.” *NOAA Magazine*. December 1, 2004. <http://www.noaanews.noaa.gov/stories2004/s2348.htm>

17 “NOAA Issues New Partnership Policy: Response to ‘Fair Weather’ Recommendations Strengthens Relationships Among Government, Universities and the Private Sector.” *NOAA Magazine*. December 1, 2004. <http://www.noaanews.noaa.gov/stories2004/s2348.htm>

In response, in 2004 NOAA announced a new partnership policy, which included the following provisions:¹⁸

- NOAA will adhere to applicable law regarding government information, based on the premise that government information is a valuable national resource and the benefits to society are maximized when such information is available to all.
- NOAA will carry out activities that contribute to its mission and provide open and unrestricted access to publicly funded information at the lowest possible cost.
- NOAA will provide information in forms accessible to the public as well as the underlying data in forms convenient to additional processing.
- NOAA will give due consideration to the ability of private sector entities and the academic and research community to provide diverse services and will consider the effects of its decision on the activities of these entities, to serve the public interest and advance the environmental information enterprise as a whole.
- NOAA will not haphazardly institute significant changes in existing information dissemination activities, or introduce new services without first carefully considering the full range of views and capabilities of all parties.
- NOAA will use appropriate mechanisms to encourage input from and collaboration with others on decisions affecting the environmental information enterprise.
- NOAA will promote the open and unrestricted exchange of environmental information worldwide.
- NOAA's participation in the environmental information enterprise will be based on the principles of mission connection, consultation, open information dissemination, equitable dealings and recognition of the role of others.

These provisions were designed to help “nurture the growth of a complex and diverse environmental information enterprise and to

¹⁸ “NOAA Issues New Partnership Policy: Response to ‘Fair Weather’ Recommendations Strengthens Relationships Among Government, Universities and the Private Sector.” *NOAA Magazine*. December 1, 2004. <http://www.noaanews.noaa.gov/stories2004/s2348.htm>

serve the public interest by giving our nation the best environmental information services in the world,” as described by Conrad C. Lautenbacher, undersecretary of commerce for Oceans and Atmosphere and NOAA administrator at the time.¹⁹ Accordingly, in 2004, NOAA changed its data format to the widely used XML, significantly lowering barriers to data use and expanding the base of potential information consumers tremendously. These actions set the stage for NOAA to be a leader in open data policies, and more specifically, for the NOAA Data Catalog that is available today (data.noaa.gov/).²⁰

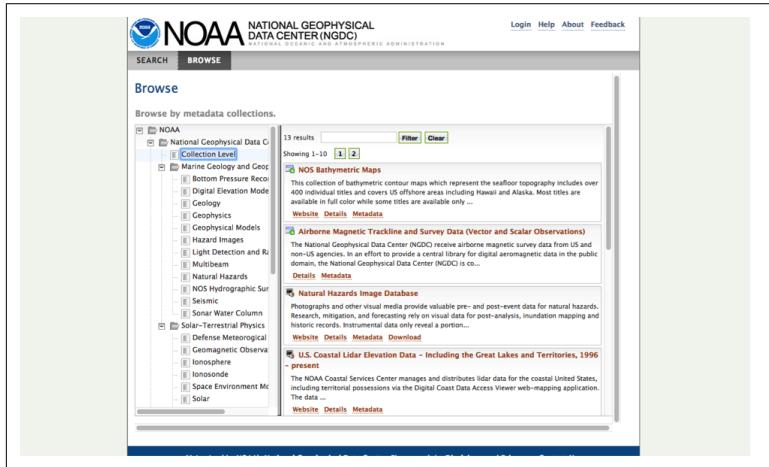


Figure 14-1. NOAA’s Geoportal

Today, NOAA data is archived at the National Centers for Environmental Information (NCEI), which “is committed to full and open data access in support of its community of Data Producers and Data Consumers.” NCEI archives data in collaboration with a number of international and national organizations “dedicated to the exchange and open access of ocean- and climate-related data.” NOAA continues to work hard to make its forecasting information more relevant and precise, corresponding to advances in technology and in user

¹⁹ “NOAA Issues New Partnership Policy: Response to ‘Fair Weather’ Recommendations Strengthens Relationships Among Government, Universities and the Private Sector.” *NOAA Magazine*. December 1, 2004. <http://www.noaanews.noaa.gov/stories2004/s2348.htm>

²⁰ Chopra, Aneesh P. and Ethan Skolnick. *Innovative State: How New Technologies Can Transform Government*. (New York, NY: Atlantic Monthly Press), 2014.

feedback on data service or product needs. Once a service has been developed, NOAA promotes it through formal and informal avenues, such as its Web channels and list-servs.²¹ NOAA's early attempts to use standard and modern Web services to distribute the data, as opposed to "dumping it all on an FTP site," helped reinforce the perception of NOAA as an open data innovator, both externally and within internal culture.²²

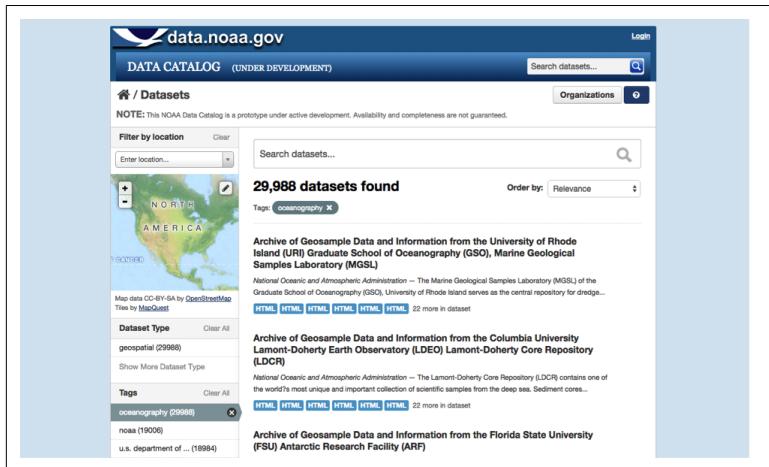


Figure 14-2. Oceanography Data Sets on NOAA's Data Portal

Data sets and data products on the NOAA portal are organized:

- by categories
 - Satellite Data
 - Regional Products (e.g., Harmful Algal Blooms Observing Systems, Arctic Regional Climatology, Gulf of Mexico Data Atlas)
 - Observational & Near-Real-Time Data
 - Instrument Types (e.g., Acoustic Doppler Current Profiler Data, Buoy Data)

²¹ GovLab interview with Tim Owen, Chief of Climate Information Services Division, NOAA, July 27, 2015.

²² GovLab interview with Jeff De La Beaujardiere, Data Management Architect, NOAA, September 18, 2015.

- Model Data (e.g., NOAA Operational Model Archive and Distribution System)
- Ocean Profile Data
- Ocean Climatology
- Project Data (e.g., Coral Reef Information System, Joint Archive for Sea Level)
- International Ocean Atlas Series (e.g., Climatic Atlas of the North Pacific Seas 2009)
- by data types (e.g., Temperature, Oxygen, Ocean Currents, Waves)
- by specific data services (e.g., Live Access Server, HTTP)²³

However, of the 30 petabytes of NOAA data collected and archived every year, only about 10 percent is openly available for public access on its websites.²⁴ Again, this is due to the technical challenges inherent in providing access to the 20 terabytes of data produced daily,²⁵ rather than cultural or policy barriers. Still, NOAA's home page, noaa.gov and the more weather-specific data site, weather.gov, are the most visited websites for the U.S. federal government overall.²⁶ The average citizen looking for basic information about weather, such as beach forecasts or severe weather alerts, knows to go to those two sites, or to climate.gov for more detailed climate research data. Indeed, climate data has become "very important" to NOAA's communications with both the public and the private sector. Officials are learning that the demand for this data is very high and the expectations surrounding the access, availability and products from NOAA data are increasing.²⁷

²³ <https://data.noaa.gov/dataset>

²⁴ "NOAA Looks for Advice to Make Its Data Easier to Use." *Eos*, Vol. 95, No. 11. March 18, 2014. <http://onlinelibrary.wiley.com/doi/10.1002/2014EO110003/pdf>

²⁵ "U.S. Secretary of Commerce Penny Pritzker Delivers Remarks on the Power and Potential of Open Government Data." U.S. Department of Commerce. July 14, 2014. <http://www.commerce.gov/news/secretary-speeches/2014/07/us-secretary-commerce-penny-pritzker-delivers-remarks-power-and>.

²⁶ GovLab interview with Zachary Goldstein, Chief Information Officer, NOAA, September 3, 2015.

²⁷ GovLab interview with Tim Owen, Chief of Climate Information Services Division, NOAA, July 27, 2015.

Industry users (e.g., utilities, energy, agriculture, space) usually access NOAA data through a more advanced open data application called the Geoportal, which attempts to centralize access to distributed geospatial data, tools, applications and services, allowing users to search and access data via a master catalog.²⁸ However, NOAA recognizes that much of its data remains siloed, which can make searching difficult for all users, and therefore launched a website redesign in late 2015, described in more detail below.

Users can directly use these sites and data sets on their own, or they can partner with NOAA to develop applications and other products to access and analyze the data to fit their specific needs. Some of these partnerships are listed below. These partnerships have changed NOAA's own perceptions of its data, nudging it to improve its services and the quality of the information and tools on its site. Overall, say NOAA insiders, partnerships have greatly expanded the scope and the potential of the organization's data. As Tim Owen, chief of the NOAA Climate Information Services Division, puts it: The move toward openness is "as much of a cultural change as it is a policy change. ... Open data brought collaborative channels to work that haven't worked before, or weren't even possible to work 10-15 years ago."²⁹

Private partnerships have also helped shift the internal culture to be more willing to allow industry and economic forces to find new ways to publish and use the data, with NOAA filling in the gaps as necessary. However, NOAA firmly believes that its data is a public good, open and already paid for by the taxpayers, and so, as NOAA CIO Zach Goldstein explains, "NOAA will always focus on advancing science and fulfilling its duty to serving the public interest."³⁰

"Open data brought collaborative channels to work that haven't worked before, or weren't even possible to work 10-15 years ago."

—Tim Owen, NOAA Climate Services Division

²⁸ <https://catalog.data.gov/dataset/noaas-national-ocean-service-nos-data-explorer-geoportal>

²⁹ GovLab interview with Tim Owen, Chief of Climate Information Services Division, NOAA, July 27, 2015.

³⁰ GovLab interview with Zachary Goldstein, Chief Information Officer, NOAA, September 3, 2015.

NOAA Partnerships^{31,32,33,34}

An example of NOAA's partnerships can be seen in the Climate Data Initiative, launched in March 2014 by the Obama administration with the goal of leveraging the government's data resources to stimulate innovation and entrepreneurship in response to climate change. As a part of the president's Climate Action Plan, the Climate Data Initiative launched a number of significant commitments from federal agencies (including NOAA), private-sector collaborators and research institutions to combat climate change through data-driven innovation.³⁵

Google

In support of the White House Climate Data Initiative, Google is donating one petabyte of cloud storage to house satellite observations, digital elevation data and climate/weather model data sets and 50 million CPU hours of high-performance cloud computing resources on the Google Earth Engine geospatial analysis platform. The goal is to make climate information "as accessible to the public as using Google Maps to get driving directions."

Climate Central

Climate Central, a nonprofit news organization that performs analysis on climate science to inform reporting, will release a "free Web tool providing local projections, maps and assessments of exposure to sea level rise and coastal flooding tabulated for every coastal zip code, municipality, county and state in

31 "Google's Participation in the Climate Data Initiative." https://docs.google.com/document/d/15BH83pdO0-hY-vnznLxCU6mH9sZ_8mkSZHBIJsOPqU4/pub.

32 "Surging Seas." Surging Seas: Sea Level Rise Analysis by Climate Central. <http://sealevel.climatecentral.org/>.

33 "Creating a Climate-Resilient Planet with Maps, Apps, and Data." ArcWatch: GIS News, Views, and Insights. April 1, 2014. <http://www.esri.com/esri-news/arcwatch/0414/creating-a-climate-resilient-planet-with-apps-maps-and-data>

34 "Creating a Climate-Resilient Planet with Maps, Apps, and Data." ArcWatch: GIS News, Views, and Insights. April 1, 2014. <http://www.esri.com/esri-news/arcwatch/0414/creating-a-climate-resilient-planet-with-apps-maps-and-data>

35 "FACT SHEET: The President's Climate Data Initiative: Empowering America's Communities To Prepare For The Effects Of Climate Change." The White House, 2014. <http://www.whitehouse.gov/the-press-office/2014/03/19/fact-sheet-president-s-climate-data-initiative-empowering-america-s-comm>.

the U.S., along with planning, legislative and other geographic districts.” Exposure assessments will cover more than 100 demographic, economic, infrastructure and environmental variables using data drawn mainly from federal sources, including NOAA open data.

Esri

Esri, a data and geography research group focused on sustainability, is offering geospatial technology and expertise to 12 coastal cities to help them build maps and applications with NOAA data to help plan for climate change’s impact. Esri will also host a new climate-focused geo-collaboration portal and sponsor a Climate Resilience App Challenge, awarding prizes to applications that focus on solutions to climate-related issues.

Impact

There are many ways to measure the impact of NOAA’s data and its data portal. One fairly conventional way is simply to look at data consumption—to count visitors to the site and data downloads. NOAA officials remark that these usage numbers reveal an increase over time, as user interest in and sophistication with data increases, as does the amount of data and products NOAA offers.³⁶ Tim Owen, chief of the NOAA Climate Information Services Division, explains that open data has “raised our awareness of the level of sophistication of the use-community and in turn forces the government to step up the response.”³⁷

Economic Benefits

However, data consumption does not on its own provide an accurate indicator of value created. What matters is not *how much* data users download or analyze, but *what they do* with that data. More sectors of the U.S. economy are recognizing the impacts of weather, water and climate on their operations and are becoming more sophisticated at using weather-related information to make better decisions.

³⁶ GovLab interview with Tim Owen, Chief of Climate Information Services Division, NOAA, July 27, 2015.

³⁷ GovLab interview with Tim Owen, Chief of Climate Information Services Division, NOAA, July 27, 2015.

NOAA's National Weather Service (NWS) collects environmental information and provides services for other governmental agencies, emergency managers, the private sector, the public and the global community. NOAA's National Ocean Service (NOS) provides ocean and coastal science, tools and services to address threats to coastal areas such as climate change, population growth, port congestion and contaminants in the environment, all working toward healthy coasts, coastal populations and coastal economies.

To better understand the impact of its data, NOAA has made a number of quantitative estimates of the value of its products and services using the Value of Information (VOI) methodologies. These methodologies are based on the premise that the information informs decisions, and that these decisions have real-world economic outcomes.³⁸

Several methodologies can be employed to estimate VOI, including:³⁹

- The first is to model and compare the outcomes of decisions taken with information vs. those taken without information.
- The second is to ask stakeholders (e.g., those who have taken decisions using information) to assess the value of the information they used—i.e., a self-assessment methodology.
- The third is by comparing data from actual events—i.e., observed effects of, for example, weather phenomena with and without forecasts or warnings.^{40,41}

³⁸ "Value of a Weather Ready Nation." NOAA. October 13, 2011. <http://www.ppi.noaa.gov/wp-content/uploads/PPI-Weather-Econ-Stats-10-13-11.pdf>

³⁹ Adams, R.M., C.C. Chen, B.A. McCarl, and R.W. Weiher, "The economic consequences of ENSO events for agriculture," *Climate Research*, 13 (December 10, 1999). Available at: <http://www.int-res.com/articles/cr/13/c013p165.pdf>

⁴⁰ "Value of a Weather Ready Nation." NOAA. October 13, 2011. <http://www.ppi.noaa.gov/wp-content/uploads/PPI-Weather-Econ-Stats-10-13-11.pdf>

⁴¹ Weiher, Rodney. "Assessing the Economic & Social Benefits of NOAA Data." NOAA. February 2008. <http://www.oecd.org/sti/ieconomy/40066192.pdf>

Based on these various methodologies (and sometimes a combination of them), several estimates have been made of the VOI of NOAA's data products:^{42,43}

- NOAA real-time data supplies a burgeoning private weather service industry with well over \$700 million in value added annually.
- The United States' \$8-10 billion and growing annual Weather Derivatives financial industry relies on NOAA's seasonal weather data and records.
- NOAA's forecasts and warnings and associated emergency responses result in \$3 billion in a typical hurricane season.
- United States electricity generators save \$166 million annually using 24-hour temperature forecasts.
- Economic benefits of new investments in United States coastal ocean observing systems from improved coastal marine information is estimated at over \$700 million annually.
- Benefits to world ship routing from NOAA polar satellite data is estimated at \$95 million annually.
- Every dollar invested in mitigation of storm-surge effects on coastal communities saves the U.S. taxpayer four dollars in losses from natural hazards.
- Installation of NWS Doppler radars reduced tornado fatalities and injuries by 40 percent from the levels in late 80s and early 90s.

Other studies have attempted to quantify the value of accurate weather data over the years. Notable findings include:

- The agriculture industry has always been among the most important consumers of NOAA's products. By providing more accurate climate forecasts and more timely warnings of adverse weather, NOAA has helped the industry improve decision-

⁴² "Value of a Weather Ready Nation." NOAA. October 13, 2011. <http://www.ppi.noaa.gov/wp-content/uploads/PPI-Weather-Econ-Stats-10-13-11.pdf>

⁴³ "The Economic Value of Resilient Coastal Communities." NOAA. March 18, 2013. <http://www.ppi.noaa.gov/wp-content/uploads/EconomicValueofResilientCoastalCommunities.pdf>.

making and crop yields. According to one estimate, data provided by the NWS Climate Prediction Center—a component of NOAA⁴⁴—has benefited U.S. agriculture by over \$460 million by helping guide planting decisions in El Niño, normal and La Niña years.⁴⁵

- The U.S. electricity industry is estimated to save \$166 million every year due to better forecasts that allow it to estimate and plan for demand. Each percentage point of improvement in the quality of forecasts saves an additional \$1.4 million; each 1°C improvement results in approximately \$59 million of additional savings annually.⁴⁶
- According to a nationwide survey, 96 percent of the U.S. public obtains, either actively or passively, 301 billion forecasts each year. Based on an average annual household value of \$286 placed on weather information,⁴⁷ the American public collectively receives \$31.5 billion in benefits from forecasts each year. These benefits far exceed the \$5.1 billion spent annually by both private and public weather bureaus on generating forecasts.⁴⁸
- Data from NOAA (and other government agencies) has also created value by spurring the development of a booming weather-related industry that creates jobs and new economic opportunities. According to Aneesh Chopra, the nation's former chief technology officer, "Weather is a \$5 billion-a-year indus-

⁴⁴ <http://www.weather.gov/about>

⁴⁵ "Value of a Weather Ready Nation." NOAA. October 13, 2011. <http://www.ppi.noaa.gov/wp-content/uploads/PPI-Weather-Econ-Stats-10-13-11.pdf>

⁴⁶ Teisberg, T., R. Weiher, and A. Khotanzad. "The Economic Value of Temperature Forecasts in Electricity Generation." *Bulletin of the American Meteorological Society*, Vol. 86, No. 12. December, 2005. <http://journals.ametsoc.org/doi/pdf/10.1175/BAMS-86-12-1765>

⁴⁷ Lazo, Jeffrey K., Rebecca E. Morss, and Julie L. Demuth. "300 Billion Served: Sources, Perceptions, Uses, and Values of Weather Forecasts." *Bulletin of the American Meteorological Society*, Vol. 90. June 2009. <http://journals.ametsoc.org/doi/pdf/10.1175/2008BAMS2604.1>

⁴⁸ Lazo, Jeffrey K., Rebecca E. Morss, and Julie L. Demuth. "300 Billion Served: Sources, Perceptions, Uses, and Values of Weather Forecasts." *Bulletin of the American Meteorological Society*, Vol. 90. June 2009. <http://journals.ametsoc.org/doi/pdf/10.1175/2008BAMS2604.1>

try.”⁴⁹ Some prominent examples of companies built around weather data include The Weather Channel, which uses NOAA data to reach some 97.3 million American households; and the Climate Corporation, which used weather data to provide “weather insurance” to businesses and was sold in 2013 to Monsanto for \$930 million.⁵⁰

- Using the self-assessment methodology mentioned above to estimate VOI, a survey was conducted of households in states prone to hurricane damage to learn how much taxpayers would be willing to pay for enhanced hurricane forecasts. Researchers found that, on average, households in at-risk states were willing to pay an additional \$14.34 per year.⁵¹

Additionally, some of NOAA’s impacts are less obviously economic in nature. But while fighting erosion and lessening toxins in coastal waters clearly have wider benefits, these results can help to bolster, for example, tourism industries and lessen disaster-related economic losses.

- Penny Pritzker, the U.S. Secretary of Commerce, recently pointed out that more accurate weather forecasts have saved lives by improving warning-times for events like tornadoes, resulting in individuals having more time to get to safety, and also for businesses to prepare their properties and activities to mitigate damage.⁵²

49 “Government Data Could Unleash Economic Growth, Expert Says,” *Government Technology*. December 11, 2014. <http://www.govtech.com/data/Government-Data-Could-Unleash-Economic-Growth-Expert-Says.html>.

50 Upbin, Bruce. “Monsanto Buys Climate Corp for \$930 Million.” *Forbes*. October 2, 2013. <http://www.forbes.com/sites/bruceupbin/2013/10/02/monsanto-buys-climate-corp-for-930-million/>

51 Lazo, J.K., D.M. Waldman, B.H. Morrow, and J.A. Thacher. “Assessment of Household Evacuation Decision Making and the Benefits of Improved Hurricane Forecasting.” *Weather and Forecasting*, Vol. 25, No. 1. February, 2010. <http://journals.ametsoc.org/doi/pdf/10.1175/2009WAF2222310.1>

52 “U.S. Secretary of Commerce Penny Pritzker Delivers Remarks on the Power and Potential of Open Government Data.” 2014 Esri International User’s Conference, July 14, 2014. <http://www.commerce.gov/news/secretary-speeches/2014/07/us-secretary-commerce-penny-pritzker-delivers-remarks-power-and>

- NOAA's data has proved instrumental in a number of environmental and conservation efforts in recent years whether through one of operating units or other government agencies using NOAA data. For example, the data has been used to track and respond to coastal erosion;⁵³ to predict and forecast areas at risk from wildfires, particularly in southern California;⁵⁴ to protect ecosystems in the Gulf of Mexico that are adversely affected by toxins released by algae blooms in coastal waters;⁵⁵ and in a host of additional settings and applications designed to protect water bodies, forests, animal life and other natural phenomena, key factors in many regional economies, including seafood, recreation, tourism, property and development.

“... [E]verything that we do is of service to someone. From kindergarten students to aviators, every spectrum of the public has something that NOAA touches in their lives.”

—Allison Soussi-Tanani, NOAA Digital Strategy Lead and Web Committee Co-Chair

Challenges

NOAA's ultimate ambition is to make more data available at any time, in any format. While the organization has made significant strides toward that goal, several challenges remain. These include:

Scaling

Despite NOAA's leadership in open data, some data remains locked up due to technical and resource-related challenges. As data collection and analysis grows both in volume and complexity, NOAA will have to continue integrating new technologies to provide accurate, useful data services and keep up with user demand and sophistication. NOAA's Big Data Partnership, a plan to collaborate more closely with the private sector, described further below, should prove central to meeting this challenge. Through the Partnership, NOAA

⁵³ “Stories from the Field.” Digital Coast, Office for Coastal Management, NOAA. <http://coast.noaa.gov/digitalcoast/stories/list>

⁵⁴ “Assessing Fire Hazard Risk in Southern California.” Digital Coast, Office for Coastal Management, NOAA. <http://coast.noaa.gov/digitalcoast/stories/californiofire>

⁵⁵ “Harmful Algal Blooms Observing System—HABSOS.” NOAA: Harmful Algal Blooms Observing System. <http://habsos.noaa.gov/>.

will be able to crowdsource and assess innovative ideas directly from the private sector to help meet existing and emerging demands. Crowdsourcing solutions among partners and the public will also help NOAA to more quickly identify gaps or quality issues with the data, as more users mean more frequent testing of the data.⁵⁶

To help scale, NOAA is also working to measure how open its data is by taking an inventory of all its data and analyzing how accessible that data is through its websites. Asking for example, how many data sets have an online link to the data? Where do those links lead to exactly: a standardized Web service that lets users download data or a project home page that requires additional navigation? NOAA hopes this assessment will provide a better understanding of exactly which data sets are open and to what extent that data meets accepted open standards and usability, in order to better scale and meet user needs—perhaps by writing software that automatically accesses the catalog holdings, services offered and formats used for each user need.⁵⁷

Data Security

To address data security challenges, from hackers to the broader system risks inherent in such large volumes of data processing, NOAA has a dedicated Cyber Security Division, which serves as the NOAA IT security officer. The division oversees the accreditation of NOAA's IT systems, including the “development of computer security plans, risk assessments, development and testing of contingency/disaster recovery plans, and system certification.” NOAA also has an Information Technology (IT) Security Program with a dedicated Computer Incident Response Team, which provides “reasonable and acceptable assurance that IT systems are performing as specified; that information is provided adequate protection; that data and software integrity is maintained; and, that unplanned disruptions of processing will not seriously impact mission accomplishment.” While data security will likely always be a threat, as it is with any data service, continued investment in the Cyber Security Division can help NOAA mitigate any risks.

⁵⁶ GovLab interview with Maria Patterson, Scientific Lead, Open Science Data Cloud, University of Chicago, Open Cloud Consortium, October 1, 2015.

⁵⁷ GovLab interview with Jeff De La Beaujardiere, Data Management Architect, NOAA, September 18, 2015.

Understanding User Needs

In order to keep growing and serving its users, NOAA must also develop more accurate data use measurement tools and better understand its user capabilities and needs. To those ends, NOAA has already expanded upon rudimentary measures of usage (e.g., volume and downloads) to more detailed customer analytics over the past decade, and has provided more data formats, visualizations and applications in response to user requests. To remain useful and meet its full potential for users, NOAA will have to continue investing in customer analytics and apply the insights learned from analyzing user behavior toward developing more products and services, as user needs inevitably evolve and grow.⁵⁸

Better understanding of user needs is also one of the goals of the Big Data Project (detailed below.) By partnering with key private sector actors involved in big data, open data and cloud computing, NOAA can directly collaborate with experts representing different user bases and their specific needs regarding open data access, types, standards and products. Building an active “ecosystem” around the data, with an active feedback loop among partners, is “very high on [NOAA]’s list of priorities,”⁵⁹ and will help NOAA better address this challenge as its open data initiatives continue to scale in size and scope.

Data Navigation and User Experience

NOAA is also seeking to improve user experience by redesigning its entire Web presence, a process launched in 2014. Given NOAA’s history as an aggregation of different agencies, its Web presence grew in a similar way, with many sub-sites with siloed information that may be difficult to find and/or repeat elsewhere, making navigation difficult. To make the Web experience more cohesive, NOAA has been working to promote itself both internally and externally as more of a unified entity, shifting from “this is *my* [agency’s] content,

⁵⁸ GovLab interview with Tim Owen, Chief of Climate Information Services Division, NOAA, July 27, 2015.

⁵⁹ GovLab interview with Maria Patterson, Scientific Lead, Open Science Data Cloud, University of Chicago, Open Cloud Consortium, October 1, 2015.

to this is *our* [NOAA's] content.”⁶⁰ Redesigning its websites and portals under this mindset will hopefully help unify NOAA’s messaging in addition to improving user experience and understanding of NOAA’s offerings.⁶¹

The other cultural driver to the website redesign is viewing NOAA as a service agency, with the “understanding that everything that we do is of service to someone. From kindergarten students to aviators, every spectrum of the public has something that NOAA touches their lives,” and therefore the website, and NOAA data, should be as accessible and open to users as possible.⁶²

Data Awareness

Finally, raising and promoting data awareness remains a challenge. Although NOAA’s customer service teams help customers navigate data sets and tools they may not be aware of through Web channels, list-servs, targeted outreach initiatives and surveys,⁶³ there may be new potential clients or users who could benefit from NOAA’s data, and who are currently unaware of the portal, do not have a formal partnership with NOAA, and/or are not as sophisticated as a typical NOAA data-user and so would need more assistance.

For the end user, who’s not necessarily the “expert data manipulator,” according to NOAA’s de La Beaujardiere: “The really key thing is to be able to see yourself in the data or have the data apply to you or where you are. If people are given a global figure, e.g., temperature around the globe is going to go up by a tenth of a degree, then that doesn’t seem to have much of an impact. If you can see what’s going to happen, has been happening, might happen at your location, that is probably the most important contextual variable.”⁶⁴

60 GovLab interview with Allison Soussi-Tanani, Digital Strategy Lead and Web Committee Co-Chair, NOAA, September 16, 2015.

61 GovLab interview with Allison Soussi-Tanani, Digital Strategy Lead and Web Committee Co-Chair, NOAA, September 16, 2015.

62 GovLab interview with Allison Soussi-Tanani, Digital Strategy Lead and Web Committee Co-Chair, NOAA, September 16, 2015.

63 GovLab interview with Tim Owen, Chief of Climate Information Services Division, NOAA, July 27, 2015.

64 GovLab interview with Jeff De La Beaujardiere, Data Management Architect, NOAA, September 18, 2015.

Therefore, to reach these users, NOAA plans to continue growing its outreach and customer services teams to identify demands and needs that users may not even be aware of themselves.⁶⁵ Additionally, NOAA plans to continue supporting social media and marketing efforts, particularly around the more day-to-day stories of not just how NOAA data is impacting people's lives, but how the individual people of NOAA prioritize serving citizens.⁶⁶

Looking Forward

Key among NOAA's strategies to meet these various challenges is a plan to pursue more partnerships with private sector companies. In a 2014 speech, the U.S. Secretary of Commerce Penny Pritzker extolled the virtues of NOAA's work and the weather industry it had spurred. But she pointed out that, despite all this work, "the public has access to just 10 percent of NOAA's more than 20 terabytes of data produced daily." (Twenty terabytes of data is equivalent to 20,480 gigabytes—in other words, NOAA produces the storage capacity of 1,280 standard iPhones each day.) She went on to suggest that in order to tap into the potential of the remaining 90 percent, "we must partner with the private sector to make our data even more useful to businesses, communities, individuals and decision-makers of all types."⁶⁷ Her remarks point to a broader interest in so-called "data collaboratives," innovative solutions for cross-sectoral data collaboration to benefit the public good.

At the heart of NOAA's plans to develop more partnerships is its "Big Data Project."⁶⁸ In February 2014, NOAA issued a request for information (RFI) to businesses and researchers to solicit ideas from the private sector on ways to better improve access to, and the usability of, its data and data products,⁶⁹ opening access to its 30 peta-

⁶⁵ http://www.cio.noaa.gov/cio_orgs.html

⁶⁶ GovLab interview with Allison Soussi-Tanani, Digital Strategy Lead and Web Committee Co-Chair, NOAA, September 16, 2015.

⁶⁷ "U.S. Secretary of Commerce Penny Pritzker Delivers Remarks on the Power and Potential of Open Government Data." U.S. Department of Commerce. July 14, 2014. <http://www.commerce.gov/news/secretary-speeches/2014/07/us-secretary-commerce-penny-pritzker-delivers-remarks-power-and>.

⁶⁸ <https://data-alliance.noaa.gov/>

⁶⁹ "NOAA announces RFI to unleash power of 'big data.'" NOAA. February 24, 2014. http://www.noaanews.noaa.gov/stories2014/20140224_bigdata.html.

bytes of annual environmental data to spur innovation and economic growth.⁷⁰ The RFI focused on developing physical infrastructure for storing and sharing NOAA's data within a publicly accessible cloud that would allow industry experts and scientists to extrapolate vast amounts of data. After receiving 70 responses from individuals, academia and industry organizations, NOAA developed an initial concept for its big data partnership business model called The Cooperative Research & Development Agreement (CRADA).⁷¹

Currently, there are two major pipelines of NOAA data. The first is station-based data via weather services channels, which has an established pipeline that can publish all data, in near real time.⁷² The other data, consisting of remote sensing, radar and other advanced forecast and climate research data, however, has a much higher volume and can only be handled by a supercomputer. Providing unlimited access therefore poses a problem, as NOAA currently does not have the resources to allow unlimited access without risking slowing down the entire data system. As NOAA CIO Zach Goldstein explains, "The last thing I want to do is run the risk of preventing citizens from getting information on hurricanes, for example. People's lives depend on getting information from NOAA."⁷³

As part of CRADA, private companies IBM, Amazon, Google, Microsoft and nonprofit Open Cloud Consortium (OCC) will help NOAA publish this supercomputer data with almost no latency, while innovating, researching and providing feedback on the data.⁷⁴ The CRADA model allows NOAA to explore what works from a cloud perspective without risking slowing down its current data publishing processes. Instead, the Big Data Project will take an itera-

70 "NOAA Looks for Advice to Make Its Data Easier to Use." *Eos*, Vol. 95, No. 11. March 18, 2014. <http://onlinelibrary.wiley.com/doi/10.1002/2014EO110003/pdf>

71 "NOAA asks industry to fund open data program." *Federal Times*. November 18, 2014. <http://www.federaltimes.com/article/20141114/FEDIT03/311140011/NOAA-asks-industry-fund-open-data-program>

72 GovLab interview with Tim Owen, Chief of Climate Information Services Division, NOAA, July 27, 2015.

73 GovLab interview with Zachary Goldstein, Chief Information Officer, NOAA, September 3, 2015.

74 GovLab interview with Maria Patterson, Scientific Lead, Open Science Data Cloud, University of Chicago, Open Cloud Consortium, October 1, 2015.

tive approach, starting with a few large data sets and allowing CRADA members themselves to determine how to most efficiently and effectively publish the data into the public domain.⁷⁵ NOAA chose this incremental approach to help maximize value by better understanding which data sets are most valuable, and to encourage a feedback loop among Big Data Project partners.⁷⁶

CRADA is also innovative in that it asks the private companies involved to pay the marginal cost for the data, with Goldstein noting that “it may be open but it’s not free.”⁷⁷ Dr. Maria Patterson of the Open Cloud Consortium (OCC) echoes this sentiment, noting that data being “freely available doesn’t mean there is not a cost associated with it,” and in order to have a lot of data open, someone has to be paying for storage, access and sharing. To balance this cost, therefore, NOAA will not charge users for access to the data itself, but if the cost of providing *new* access requires *additional* resources, CRADA members will absorb that marginal cost. If successful, the Big Data Project will represent a new and cost-effective approach to publishing and utilizing NOAA’s vast data resources.

It is important to highlight that in addition to offering an innovative public-private business model, the research component of the Big Data Project is equally valuable, illustrated by the fact that the partners involved are not “contractors” but participants in a “cooperative *research* and development agreement.”⁷⁸ As Dr. Patterson explains, NOAA was very deliberate in developing the CRADA approach, and that “the entire project itself is its own research experiment—asking how can NOAA work with partners in a mutually beneficial arrangement to release data into an ecosystem. If the business model fails, the research project can still succeed in identifying what works and what does not work.”⁷⁹ NOAA CIO Zach Goldstein expands

⁷⁵ GovLab interview with Zachary Goldstein, Chief Information Officer, NOAA, September 3, 2015.

⁷⁶ GovLab interview with Maria Patterson, Scientific Lead, Open Science Data Cloud, University of Chicago, Open Cloud Consortium, October 1, 2015.

⁷⁷ GovLab interview with Zachary Goldstein, Chief Information Officer, NOAA, September 3, 2015.

⁷⁸ GovLab interview with Maria Patterson, Scientific Lead, Open Science Data Cloud, University of Chicago, Open Cloud Consortium, October 1, 2015.

⁷⁹ GovLab interview with Maria Patterson, Scientific Lead, Open Science Data Cloud, University of Chicago, Open Cloud Consortium, October 1, 2015.

upon this sentiment, “No one else in government has looked at building open data with big data in this way. What we’re doing that’s unique is putting it all together and updating it with today’s technology. It’s possible it won’t work, and NOAA may eventually have to resort to a more traditional pay-per-service model, but the only way we’ll fail is if we fail to learn—if we fail to try.”⁸⁰

As anticipated, the partnership with the private sector combines three powerful resources: NOAA’s tremendous volume of high-quality environmental data and advanced data products, private industry’s vast infrastructure and technical capacity, and the American economy’s innovation and energy. The overall goal is to create a sustainable, market-driven ecosystem that lowers the cost barrier to data publication, spurs economic development and creates new jobs.

From its early, closed inception to today’s Data Portal to tomorrow’s Big Data Project, the evolution of NOAA’s data efforts represents a powerful illustration of how opening data can have significant economic, social, environmental and other impacts. The changes wrought by NOAA’s 30 petabytes of annual data have rippled through American (and, indeed, global) life, often under the surface, but no less powerful for that. With NOAA’s plans for expansion and new partnerships, there is every reason to believe that further transformations await.

⁸⁰ GovLab interview with Zachary Goldstein, Chief Information Officer, NOAA, September 3, 2015.

CHAPTER 15

United States Opening GPS for Civilian Use

Creating a Global Public Utility

*By Andrew Young, Christina Rogawski and
Stefaan Verhulst*

Summary: The Global Positioning System (GPS) has had a major impact on both business and society—with estimated losses reaching \$96 billion were the system to be somehow discontinued. GPS is a U.S. government-owned technology that provides military and civilian users with positioning, navigation and timing (PNT) services. The system was developed by the U.S. Department of Defense in 1978 and originally restricted to military use. Following the Korean Airlines disaster in 1983, the Reagan administration announced that GPS would be available for civilian use. In 2000, President Clinton announced a commitment to grant civilians access to un-degraded GPS signals on par with those used by the military, and in 2007, under President Bush, the Department of Defense made that commitment permanent. Over the past 20 years, GPS technology has led to a proliferation of commercial applications across industries and sectors, including agriculture, construction, transportation, aerospace and—especially with the proliferation of portable devices—everyday life. In addition to creating new efficiencies and reducing operating costs, the adoption of GPS technology has improved safety, emergency response times and environmental quality, and has delivered many other less-readily quantifiable benefits.

Dimension of Impact: Creating Opportunity—Economic Growth

Key Takeaways:

- One of the paradigmatic examples pointed to by open data advocates, the GPS system is not truly a data system. Rather, it is a service—like a dial-tone. That said, its role in demonstrating the incredible value and potential impacts of opening information to the public cannot be discounted.
- Open data initiatives can be self-reinforcing. When GPS was opened for civilian use, aviation was the only explicitly named industry that would gain access to the signal, and there was no mention of any kind regarding the system being free of charge. Once the system was opened, however, the impacts across industries and sectors were so massive that it would be nearly impossible to turn back to a more constrained system.
- Location information enabled by governments can be put to potentially troubling ends. While GPS has created incredible value for the global public, it has also enabled a diversity of privacy-intruding technologies—from smartphone games that share a user’s location with unnamed third parties to employers tracking every move of their factory employees.

Context and Background

GPS in Our Lives Today

For most of human history, the ability to determine our exact location in the world at almost any time only existed in the realm of fantasy. The benefits and uses of such a capability—from easing travel headaches to ensuring kindergartners do not venture too far from the playground to uncovering elephant poaching rings¹—are limited only by the imagination.

The Global Positioning System (GPS) has become such a central piece of life for so many across the globe that, despite the system

¹ “GPS Trackers in Fake Elephant Tusks Reveal Ivory Smuggling Route.” *NPR*. August 12, 2015. <http://www.npr.org/sections/goatsandsoda/2015/08/12/431908397/gps-trackers-in-fake-elephant-tusks-reveal-ivory-smuggling-route>

only becoming fully operational in the mid 1990s, “the Associated Press uses the GPS acronym without elaboration.”² The story of how GPS came to occupy a nearly ubiquitous place in global life, creating impacts across regions and sectors, touches upon the Space Race and Cold War and involves a diversity of major technical, geopolitical and governing decisions.

Early History of Satellite Navigation: Sputnik

The roots of this system can be traced back to October 1957, when the Soviet Union launched Sputnik, mankind’s first satellite. A few days after Sputnik entered orbit and captured the world’s attention, two physicists at Johns Hopkins University’s Applied Physics Laboratory, George Weiffenbach and William Guier, realized that they were able to predict Sputnik’s orbits by analyzing the Doppler shift of its radio signal as the satellite circled the earth.³ By relying on the frequency of the Sputnik signals, which shifted between 500 to 1,500 hertz, they were able to calculate the satellite’s orbital path with great reliability.

Weiffenbach and Guier were still in the midst of their work when the Sputnik radio went dead. However, a month later, in November 1957, the Russians launched a second satellite, Sputnik II. This one broadcast on two radio frequencies at once, a fact which not only made the signal more reliable but also helped the scientists get a more accurate fix on the satellite’s location.

Even more importantly, the early work with Sputnik I and II led to a central insight that helped lay the groundwork for modern GPS. Scientists soon realized that if satellites could be tracked from the ground by measuring the frequency of their radio signals, then conversely, the locations of receivers on the ground could also be tracked by their distance from the satellites. This realization has been described as the “conceptual foundation of modern GPS”; indeed, much the same system is used today by the GPS in our smartphones or cars.

² Easton, Richard D. Easton and Eric F. Frazier. *GPS Declassified: From Smart Bombs to Smartphones*. (Lincoln: Potomac Books, University of Nebraska Press), 2013.

³ Easton, Richard D. Easton and Eric F. Frazier. *GPS Declassified: From Smart Bombs to Smartphones*. (Lincoln: Potomac Books, University of Nebraska Press), 2013.

Early History of Satellite Navigation: Transit

During the 1960s, Guier and Weiffenbach built on the research they had conducted with Sputnik to develop the first satellite-based navigation system. This work was done in collaboration with the Defense Advanced Research Projects Agency (DARPA) which led to the deployment of the first operational satellite navigation system, Transit. Although Transit was invented by the United States as a military technology, it was accessible to any user—including both Argentina and Great Britain who used Transit as a navigational aid for their navies during the 1982 Falklands War.⁴

Transit was an extraordinarily effective technology, but it did have a couple of serious limitations. First, Transit satellites were not always within range, meaning that a ship at sea might have to wait several hours before one of the Transit satellites passed overhead. This was obviously a problem for any potential user or application that relied on continuous updates. In addition, because Transit was a two-dimensional system, a receiving unit could calculate only its latitude and longitude, but not altitude. So, for example, a user would know how far north, south, east or west he was, but not how high above the land or sea.⁵

Project Description and Inception

The Arrival of GPS

In 1978, the first GPS satellite was launched into orbit—20 years after Sputnik first inspired the system to be developed, and 14 years after Transit paved the way for space-based navigation. The GPS system began with a \$150 million budget. This relatively meager funding, “and the fact that this program lacked the self-evident drama of the Apollo program,” led to the completion of the fully operational GPS Network taking “about twice as long as putting a man on the moon.”⁶ Indeed, it was not until 1995—three decades after the idea

⁴ Bray, Hiawatha. *You Are Here: From the Compass to the GPS, the History and Future of How We Find Ourselves*. (New York: Basic Books), 2014.

⁵ Bray, Hiawatha. *You Are Here: From the Compass to the GPS, the History and Future of How We Find Ourselves*. (New York: Basic Books), 2014.

⁶ Bray, Hiawatha. *You Are Here: From the Compass to the GPS, the History and Future of How We Find Ourselves*. (New York: Basic Books), 2014.

for GPS was initially developed—that constant, uninterrupted worldwide GPS coverage was achieved shortly after the 24th GPS satellite began to orbit the Earth.⁷

Who Deserves the Credit?

To this day, significant debate exists regarding who deserves the lion's share of the credit for GPS. Bradford Parkinson, the air force engineer who brought together the joint program that helped to initiate GPS, is one likely candidate. Parkinson and his collaborator, Aerospace Corporation president Ivan Getting, have received numerous honors "for conceiving and creating the GPS system." However, Navy engineer Roger Easton was in many ways the individual who put those concepts into practice. Easton created and helped to deploy TIMATION satellites. TIMATION was a precursor to GPS that built on Transit's early advances and included many of the central components of GPS, including space-based atomic clocks. All three of the men are now in the National Inventors Hall of Fame.

⁸

The ideas behind GPS date back even before Parkinson, Getting and Easton began to move GPS forward in earnest. In 1964, Roy E. Anderson of the General Electric Corporate Research and Development Center was the first person to propose a navigation system based on 24 satellite coverage in a mid-altitude configuration. This passive design is a central feature of today's GPS system.⁹ Moreover, the spread-spectrum signal structure that enables GPS signals to be received on Earth was invented in World War II by actress Hedy Lamarr and composer George Antheil.¹⁰

⁷ Easton, Richard D. and Eric F. Frazier. *GPS Declassified: From Smart Bombs to Smartphones*. (Lincoln: Potomac Books, University of Nebraska Press), 2013.

⁸ Bray, Hiawatha. *You Are Here: From the Compass to the GPS, the History and Future of How We Find Ourselves*. (New York: Basic Books), 2014.

⁹ Easton, Richard D. and Eric F. Frazier. *GPS Declassified: From Smart Bombs to Smartphones*. (Lincoln: Potomac Books, University of Nebraska Press), 2013.

¹⁰ Easton, Richard D. and Eric F. Frazier. *GPS Declassified: From Smart Bombs to Smartphones*. (Lincoln: Potomac Books, University of Nebraska Press), 2013.

GPS Today

Today, GPS is enabled by a collection of satellites weighing around two tons each that orbit the Earth at an altitude of around 12,500 miles. In the interest of ensuring constant signals from the satellites reaching Earth, the GPS satellites are separated across six orbital planes. The satellites orbit the earth approximately every 12 hours.¹¹ The system available for military, commercial and civilian use is highly accurate, with the typical GPS signal in space providing a “worst case” pseudorange accuracy of 7.8 meters at a 95 percent confidence level. Note that this is not the same as user accuracy, as pseudorange is the distance from a GPS satellite to a receiver. The actual accuracy for users depends on uncontrollable variables such as atmospheric effects, sky blockage and receiver quality; however, data from the FAA shows that their high-quality GPS receivers provide better than 3.5 meter horizontal accuracy.¹² The average citizen or commercial users typically does not need higher levels of accuracy; however, greater accuracy is attainable through GPS augmentation systems, which in some cases enable real-time positioning to within a few centimeters.¹³

Currently, the GPS system is governed by a diversity of U.S. Government departments and agencies, which are coordinated by the National Executive Committee for Space-Based Positioning, Navigation and Timing, which was created by President George W. Bush in 2004. The secretaries of defense and transportation were tasked with jointly chairing the committee, along with officials from the State, Interior, Agriculture, Commerce and Homeland Security Departments, as well as representatives from the Joint Chiefs of Staff and NASA acting as members. Individuals from the White House and FCC also act as liaisons.¹⁴ The permanent staff of the Committee make up the National Coordination Office Space-Based PNT (NCO).

In addition to collaborating with industry to ensure the system serves the needs of GPS users, the central role of the NCO is to

¹¹ Easton, Richard D. and Eric F. Frazier. *GPS Declassified: From Smart Bombs to Smartphones*. (Lincoln: Potomac Books, University of Nebraska Press), 2013.

¹² “GPS Accuracy.” GPS.gov. <http://www.gps.gov/systems/gps/performance/accuracy/>

¹³ “Augmentation Systems.” GPS.gov. <http://www.gps.gov/systems/augmentations/>

¹⁴ <http://www.gps.gov/governance/excom/>

coordinate the efforts of the many agencies involved in the GPS program—with particular focus on “modernization, funding, policy, interference detection, spectrum management, international cooperation, and applications,” and on providing transparency among the agencies involved in governing the system.¹⁵

The Nature of GPS Data

One of the key features of GPS, responsible in many ways for its widespread adoption, is the fact that it has a “passive” character. Hiawatha Bray, technology columnist for *The Boston Globe* and author of *You Are Here: From the Compass to GPS, the History and Future of How We Find Ourselves*, likens it to “the same way any number of people can tune in to a radio station at the same time.”¹⁶ In other words, since a GPS user does not directly interact with the navigational satellites, but rather simply accesses the signals provided by those satellites, the number of simultaneous users is unlimited, a fact that makes the whole system much more robust and scalable.

Despite its robustness and scalability, however, there are some aspects of GPS that are surprisingly fragile. For example, GPS signals are relatively low-power “line-of-sight” communications. This means that while they can pass through clouds and other thin materials with little trouble, solid objects (e.g., buildings, walls and even tree foliage) can significantly disrupt reception.¹⁷ In addition, the precision needed for GPS to successfully complete its many duties is extremely high. A time signal error of a billionth of a second can result in a GPS receiver missing its mark by about a foot.¹⁸

Finally, it is important to recognize that, while references to “GPS data” are common, as Jason Kim, senior adviser at the NCO states, GPS is actually “not really a data service at all. ... GPS is a measure-

¹⁵ GovLab interview with Jason Kim, Senior Advisor, National Coordination Office for Space-Based Positioning, Navigation, and Timing, September 22, 2015.

¹⁶ Easton, Richard D. and Eric F. Frazier. *GPS Declassified: From Smart Bombs to Smartphones*. (Lincoln: Potomac Books, University of Nebraska Press), 2013.

¹⁷ Easton, Richard D. and Eric F. Frazier. *GPS Declassified: From Smart Bombs to Smartphones*. (Lincoln: Potomac Books, University of Nebraska Press), 2013.

¹⁸ Easton, Richard D. and Eric F. Frazier. *GPS Declassified: From Smart Bombs to Smartphones*. (Lincoln: Potomac Books, University of Nebraska Press), 2013.

ment tool. It's just a beacon that, instead of being on the ground, it's in space—like a lighthouse or a radio tower, and you just measure your distance from it.”¹⁹ So while GPS is often seen as a paradigmatic example of how government openness can create major global impacts, the tendency to label the system as an “open data” initiative is not entirely accurate.

How GPS Became Open

While not truly an open data service, as Kim puts it: “[GPS has] been our gift to the world.”²⁰ This is all the more remarkable given that GPS originally emerged as an American military technology. Its general adoption for civilian purposes, by companies and governments around the world, is a significant development.

In fact, although the popular narrative of GPS often describes a system created exclusively for military use that was later made available for civilian use, GPS was always intended to be opened for civilians as well. Co-developers Bradford Parkinson and Stephen Powers have written that, “Contrary to some versions of GPS history, from the very beginning, GPS was configured to be a dual-use system.”²¹ Nonetheless, it was a major civilian tragedy, the 1983 downing of Korean Air Lines (KAL) Flight 007, that provided the moment during which to push GPS from being primarily a military technology to widespread civilian adoption.

The KAL disaster was a significant milestone in Cold War history. It occurred on September 1, 1983, when the civilian airliner, on a flight from New York City to Seoul, strayed accidentally into Soviet airspace and was shot down by a Soviet fighter jet. All 269 people on board died. The fact that the plane deviated from its flight plan despite having “triple-redundant inertial navigation systems (INS) on board”²² quickly made the potential value of GPS in averting such

19 GovLab interview with Jason Kim, Senior Advisor, National Coordination Office for Space-Based Positioning, Navigation, and Timing, September 22, 2015.

20 GovLab interview with Jason Kim, Senior Advisor, National Coordination Office for Space-Based Positioning, Navigation, and Timing, September 22, 2015.

21 “Part 1: The Origins of GPS, and the Pioneers Who Launched the System.” GPS World. May 1, 2010. <http://gpsworld.com/origins-gps-part-1/>

22 Easton, Richard D. and Eric F. Frazier. *GPS Declassified: From Smart Bombs to Smartphones*. (Lincoln: Potomac Books, University of Nebraska Press), 2013.

future tragedies clear. Fifteen days after the disaster, White House Press Secretary Larry Speakes announced that: “World opinion is united in its determination that this awful tragedy must not be repeated. As a contribution to the achievement of this objective, the president has determined that the United States is prepared to make available to civilian aircraft the facilities of its Global Positioning System when it becomes operational in 1988.”²³

Bray, author of *You Are Here*, argues that, “causes of the crash aside, Reagan’s seemingly magnanimous offer of GPS technology was a clever propaganda coup, no more.” He points to a notice from the Federal Register posted two years prior saying the completed network “will be made available to the worldwide civil/commercial community within the limits of national security considerations.”²⁴ This posting was developed by the National Oceanic and Atmospheric Administration (NOAA) and included detailed technical standards for the civilian use of GPS and, at least in part, was meant to assure the surveying industry—eventually a key user of GPS, as discussed below—that GPS would be available for its use when ready.²⁵

Since the administration’s statement came at a time when only a few GPS satellites were in the air, and GPS did not yet exist as a fully operational service, it was, as Kim puts it, “kind of a futuristic statement.” In fact, “it was really addressing civil aviation, but people read it to mean all civil applications,” which obviously was the eventual outcome, but the statement “didn’t say anything about free” at the time.²⁶

It is only since that time that GPS has grown into the ubiquitous “global utility”²⁷ used in millions of applications and devices today. During that time, the nature of GPS has also changed, growing more

²³ Easton, Richard D. and Eric F. Frazier. *GPS Declassified: From Smart Bombs to Smartphones*. (Lincoln: Potomac Books, University of Nebraska Press), 2013.

²⁴ Bray, Hiawatha. *You Are Here: From the Compass to the GPS, the History and Future of How We Find Ourselves*. (New York: Basic Books), 2014.

²⁵ Easton, Richard D. and Eric F. Frazier. *GPS Declassified: From Smart Bombs to Smartphones*. (Lincoln: Potomac Books, University of Nebraska Press), 2013.

²⁶ GovLab interview with Jason Kim, Senior Advisor, National Coordination Office for Space-Based Positioning, Navigation, and Timing, September 22, 2015.

²⁷ Easton, Richard D. and Eric F. Frazier. *GPS Declassified: From Smart Bombs to Smartphones*. (Lincoln: Potomac Books, University of Nebraska Press), 2013.

recognizably “open.” Indeed, in 2004, the Bush administration released a policy adding another layer of openness to GPS. The U.S. Space-Based Positioning, Navigation, and Timing Policy enshrined a number of principles, including a commitment to “provide on a continuous, worldwide basis civil space-based, positioning, navigation and timing services free of direct user fees for civil, commercial and scientific uses, and for homeland security through the Global Positioning System and its augmentations, and provide open, free access to information necessary to develop and build equipment to use these services.”²⁸

According to Kim, the relevance of this announcement goes beyond just a commitment to sharing technical specifications; it also shows that innovators can depend on “the U.S. government to meet certain standards of performance in terms of accuracy, availability, reliability, continuity. ... The stability of all those commitments over time ... has really helped the private sector rely on and trust GPS.”²⁹

Continuously Operating Reference Station (CORS)

While there is significant debate regarding whether or not the GPS system truly fits the open data model, the closely related Continuously Operating Reference Station (CORS) system, on the other hand, exists comfortably within its contours. The CORS system is a massive network of GPS receivers and antennas that constantly collect data at specific sites. The satellite data collected is then transmitted to the National Geodetic Survey (NGS) and NOAA, where quality control is conducted before making that data available to the user.

The network was originally designed by the father of CORS, Bill Strange, the chief geodesist and scientist at NGS and NOAA in 1993. The Network started with a single site in Gaithersburg, Maryland.

28 “FACT SHEET: U.S. Space-Based Positioning, Navigation, and Timing Policy.” http://www.aopa.org/-/media/Files/AOPA/Home/News/All-News/News-Archives/2004/Importance-of-GPS-to-GA-is-affirmed-by-Bush/041216space_facts.pdf

29 GovLab interview with Jason Kim, Senior Advisor, National Coordination Office for Space-Based Positioning, Navigation, and Timing, September 22, 2015.

Today, the 2,000 sites found in the CORS network cover nearly the entirety of the U.S.³⁰

CORS users process GPS data that they have collected at a location of interest, together with associated GPS data from a CORS site, to calculate the coordinates of their data-collection points relative to the CORS site. With its associated tools such as OPUS (Online Positioning User Service), CORS provides free access to highly accurate (centimeter level) positions in the National Spatial Reference System (NSRS) using GPS, yielding a substantial improvement over “stand alone” GPS which can have meters of inaccuracy.³¹

The CORS Network not only provides information to users free of charge, it also exists as a cross-sector collaboration in itself. In addition to partnering with the Coast Guard, which has over 50 stations around the coast and inland waterways, “we partner with academic institutions, universities, state [Departments of Transportation] and the private sector. They follow a set of guidelines to build a uniform infrastructure, and then we manage it from there,” says Dr. Neil Weston, chief scientist of the National Geodetic Survey, part of the National Ocean Service within NOAA.³² Today, the CORS network has over 230 partners. This system allows NGS and NOAA to oversee a largely uniform network without being responsible for deploying and maintaining the equipment found at each site. That information was made available to the public almost from the start, when the “Internet was still in its infancy.”³³

Given its user-centricity, NGS’s Geodetic Services Division has made a significant effort “to educate the public on how they could use this data; how it could be more efficient for them to collect it, what the benefits were, and what the numbers meant.”³⁴ This focus on prepar-

³⁰ GovLab interview with Dr. Neil D. Weston, Chief Scientist, National Geodetic Survey, National Ocean Service, NOAA, September 24, 2015.

³¹ Leveson, Irving. “Socio-Economic Benefits Study: Scoping the Value of CORS and GRAV-D.” Final Report Prepared for the National Geodetic Survey. January 2009. http://www.ngs.noaa.gov/PUBS_LIB/Socio-EconomicBenefitsofCORSandGRAV-D.pdf

³² GovLab interview with Dr. Neil D. Weston, Chief Scientist, National Geodetic Survey, National Ocean Service, NOAA, September 24, 2015.

³³ GovLab interview with Dr. Neil D. Weston, Chief Scientist, National Geodetic Survey, National Ocean Service, NOAA, September 24, 2015.

³⁴ GovLab interview with Dr. Neil D. Weston, Chief Scientist, National Geodetic Survey, National Ocean Service, NOAA, September 24, 2015.

ing the public to make use of the data has taken the form of webinars, in-person training sessions, conferences and more.

Impact

Like many other examples included in this report, the impact of GPS can be hard to capture. Unlike other examples, however, that difficulty does not stem from limited impact or a paucity of evidence and data. On the contrary, it is hard to capture the impact of GPS precisely for the opposite reason—because it has become so far-reaching and central to so many aspects of our lives. Nonetheless, in what follows, we attempt a broad overview of some of the most important ways in which GPS has transformed everyday life.

Economic Impact

According to Kim, it is a mistake to try to capture the impact of GPS primarily in economic terms. Pointing out that GPS was created for a range of purposes—including national security and public safety—he argues: “If there are economic benefits on top of that, well that’s great, but that’s not why we fielded the system. That’s not why we continue to field the system.”³⁵ Still, as with many technologies that have a wide variety of uses, examining the economic impact of GPS is a useful way to analyze the wider breadth of its impact. It allows us to better understand the sheer variety of fields that GPS has affected.

According to a 2011 NDP Consulting study, the U.S. government spent around \$19.6 billion to build and maintain GPS satellites and ground stations.³⁶ The same report estimated that a complete degradation of the GPS signal would cost the United States \$96 billion annually—with the vast majority of the impact being the result of lost productivity for commercial GPS users (\$67.6 billion).³⁷

Here, we examine a sampling of impacts across three industries:

³⁵ GovLab interview with Jason Kim, Senior Advisor, National Coordination Office for Space-Based Positioning, Navigation, and Timing, September 22, 2015.

³⁶ Pham, Nam D. “The Economic Benefits of Commercial GPS Use in the U.S. and The Costs of Potential Disruption.” NDP Consulting. June 2011. http://www.gpsalliance.org/docs/GPS_Report_June_21_2011.pdf

³⁷ Pham, Nam D. “The Economic Benefits of Commercial GPS Use in the U.S. and The Costs of Potential Disruption.” NDP Consulting. June 2011. http://www.gpsalliance.org/docs/GPS_Report_June_21_2011.pdf

Aviation

In 1984, defense contractor Rockwell International successfully flew a private jet from Cedar Rapids, Iowa to Paris—the first instance of a transcontinental flight relying exclusively on GPS for navigation. While the flight took four days due to the limited coverage of the GPS network at the time, the flight ultimately concluded within 20 feet of its targeted destination.³⁸ Pilots began using GPS equipment in 1993 as the primary means of en route navigation—though restrictions existed during landings. In March 1994, Continental Airlines testified before the House Subcommittee on Technology, Environment, and Aviation that GPS could help the industry save \$5 billion annually as a result of reducing delays and the better-targeted routing of flights. Moreover, Continental itself was estimated to save \$1.9 million each year as a result of fewer cancellations thanks to the ability to alter routes to avoid weather concerns.³⁹

In addition to millions in cost savings, GPS allows aircrafts to fly pre-established routes “from waypoint to waypoint, where waypoints do not depend on ground infrastructure,”⁴⁰ enabling greater flexibility, efficiency and accuracy in route planning and execution, particularly when travelling across areas without suitable ground surveillance or navigation equipment, such as across oceans. GPS has also helped improve approaches to airports, making landing during poor visibility conditions safer, or even feasible under what were previously prohibitive circumstances.⁴¹

Engineering and Construction

GPS also increases productivity in engineering and construction, primarily by providing highly accurate mapping, machine guiding and measurements for a variety of processes such as surveying, excavating, transportation management, urban planning and jobsite safety monitoring. Often, a single GPS device can achieve in just a few hours what used to take several people

³⁸ Bray, Hiawatha. *You Are Here: From the Compass to the GPS, the History and Future of How We Find Ourselves*. (New York: Basic Books), 2014.

³⁹ Easton, Richard D. and Eric F. Frazier. *GPS Declassified: From Smart Bombs to Smartphones*. (Lincoln: Potomac Books, University of Nebraska Press), 2013.

⁴⁰ “Aviation.” GPS.gov. <http://www.gps.gov/applications/aviation/>

⁴¹ “Aviation.” GPS.gov. <http://www.gps.gov/applications/aviation/>

many hours, using numerous instruments and labor-intensive processes. The resulting savings of this increased efficiency are estimated to be \$7.6 billion in construction labor and \$1.3 billion in capital machinery and equipment, according to the 2011 NDP Consulting study.⁴²

Transportation

In the early days of GPS, the Department of Transportation estimated that traffic congestion cost over \$73 billion in lost productivity annually.⁴³ GPS helps reduce this congestion by making traffic and mass transit management systems more accurate and efficient. Local and federal transportation agencies use GPS to help survey their road and highway networks and to locate, map and analyze road features such as service stations, emergency services and supplies, entry and exit ramps, and road damage. GPS is also used to provide more accurate traffic updates to drivers and the media, such as construction or an accident ahead, and additional research is being conducted to examine how GPS can help enable minimal vehicle control when there is a clear need for action, such as the predeployment of air bags.⁴⁴

Using GPS technology to better track and more accurately forecast the movement of freight has “made a logistical revolution” possible, for example through the application of what is known as “time-definite delivery.” In time-definite delivery, companies can use GPS to guarantee delivery and pickup at a set time, over both short or long distances.⁴⁵ Aside from providing customers the convenience of near real-time status of their package, delivery companies like FedEx, UPS and even the U.S. Postal Service use GPS to improve services and reduce costs in three critical ways: 1) reducing fuel use via accurate monitoring of driver routes and idling times; 2) pinpointing vehicle dispatching; 3) providing roadside service and updates if something happens to

42 Pham, Nam D. “The Economic Benefits of Commercial GPS Use in the U.S. and The Costs of Potential Disruption.” NDP Consulting. June 2011. http://www.gpsalliance.org/docs/GPS_Report_June_21_2011.pdf

43 Easton, Richard D. and Eric F. Frazier. *GPS Declassified: From Smart Bombs to Smart-phones*. (Lincoln: Potomac Books, University of Nebraska Press), 2013.

44 “Roads & Highways.” GPS.gov. <http://www.gps.gov/applications/roads/>

45 “Roads & Highways.” GPS.gov. <http://www.gps.gov/applications/roads/>

a vehicle en route.⁴⁶ The use and future of GPS in transportation systems is further detailed below.

Precise Timing

In addition to its location services, another benefit and major impact of GPS stems from the way it has allowed for space-based worldwide clock synchronization. Today, GPS satellites act as the global standard for precision timing.⁴⁷ The resulting split-second accuracy provides major benefits to several industries. Telecommunications companies, for instance use GPS to synchronize call hand-offs when a phone moves between areas serviced by different cellular stations.⁴⁸ Second, accurate timing helps financial companies reduce costs and improve algorithms, while improving market transparency and regulatory compliance.⁴⁹ As Kim notes: “Financial networks will have their requirements for timing accuracy going down to the microsecond within a couple of years. The only way you can get that is either run your own atomic clock, or use GPS.”⁵⁰ Finally, time synchronization across the power grid enables the efficient transmission and distribution of power and enables power companies and utilities to quickly identify the site of power line breaks, which assists in avoiding larger-scale blackouts.⁵¹

Smartphones / Mobile Navigation

Despite the relatively short period of time during which GPS has been available to the public, the system has enabled the creation of a massive consumer industry, and seen that industry’s disruption by an even more omnipresent technology. In many ways, the story of

46 Galloway, Ryan. 5 Ways GPS Tracking Gives Businesses a Competitive Edge. April 23, 2013. <http://www.forbes.com/sites/fedex/2013/04/23/5-ways-gps-tracking-gives-businesses-a-competitive-edge/>

47 Bray, Hiawatha. *You Are Here: From the Compass to the GPS, the History and Future of How We Find Ourselves*. (New York: Basic Books), 2014.

48 Jackson, William. “The serious side of GPS, where timing is everything.” GCN. November 12, 2013. <http://gcn.com/articles/2013/11/12/gps-timing-position.aspx>

49 Bach, Andrew F. “GPS & precision timing’s role in the financial services sector.” Juniper Networks. 2014. <http://www.gps.gov/governance/advisory/meetings/2014-12/bach.pdf>

50 GovLab interview with Jason Kim, Senior Advisor, National Coordination Office for Space-Based Positioning, Navigation, and Timing, September 22, 2015.

51 “Timing.” GPS.gov. <http://www.gps.gov/applications/timing/>

consumer GPS from its initial release through its continued maturity has been the move from stand-alone GPS receivers to GPS-enabled smartphone applications.

Mobile navigation became an official heading in the NPD Group's market share annual report in 2004, with the industry garnering \$72.8 million in revenues, chiefly among the top brands of Magellan, Garmin, Pioneer, Alpine and Kenwood.⁵² By the end of 2009, more than a third of households in the U.S. were believed to have at least one personal navigation device.⁵³

While the industry remained a market force over the next decade, especially market leader Garmin, smartphones began to eat into revenues in earnest in 2008 when Apple launched its first iPhone with GPS capabilities. By mid-2009, there were over 3,000 navigation applications in Apple's App Store.⁵⁴ As of 2015, a Pew Research Center survey found that nearly two-thirds of Americans own smartphones.⁵⁵ Back in 2012, when only 46 percent of Americans owned smartphones, Pew found that almost three-quarters of those smartphone owners used some type of GPS service on their phone.⁵⁶ If the prevalence of GPS usage among smartphone owners remained consistent, then in 2015, around 150 million Americans used GPS-enabled location services on their smartphones.⁵⁷

While turn-by-turn navigation remains the quintessential use of GPS on smartphones, the system enables a wide diversity of uses,

⁵² Easton, Richard D. and Eric F. Frazier. *GPS Declassified: From Smart Bombs to Smartphones*. (Lincoln: Potomac Books, University of Nebraska Press), 2013.

⁵³ Easton, Richard D. and Eric F. Frazier. *GPS Declassified: From Smart Bombs to Smartphones*. (Lincoln: Potomac Books, University of Nebraska Press), 2013.

⁵⁴ Easton, Richard D. and Eric F. Frazier. *GPS Declassified: From Smart Bombs to Smartphones*. (Lincoln: Potomac Books, University of Nebraska Press), 2013.

⁵⁵ "The Smartphone Difference." Pew Research Center. April 2015. http://www.pewinternet.org/files/2015/03/PI_Smartphones_0401151.pdf

⁵⁶ Zickuhr, Kathryn. "Three-quarters of smartphone owners use location-based services." Pew Research Center. May 11, 2012. <http://www.pewinternet.org/2012/05/11/three-quarters-of-smartphone-owners-use-location-based-services/>

⁵⁷ The 2015 Pew report found that over two-thirds of smartphone users at least occasionally use GPS for turn-by-turn navigation while driving, 25 percent to get public transit information and 11 percent to reserve a taxi or car service. The survey did not, however, ask respondents about their general use of GPS or location-enabled apps. "About the December Week 1 and Week 3 Omnibus Survey." Pew Research Center. <http://www.pewinternet.org/files/2015/03/MethodsAndSurveyQuestions.pdf>

including check-in services like Foursquare and location-aware recommendation engines like Yelp.⁵⁸ Some widely used GPS-enabled applications include:

- Google Maps: provides users with directions via car, transport or walking, integrating real-time updates to transit service changes or traffic delays⁵⁹
- Nike+ Running: uses smartphone GPS to track users' location and speed as they run,⁶⁰ similar to MapMyRun, which lets users record and share their workouts and goals⁶¹
- Waze: a social GPS navigation app that crowdsources traffic and road data, including accidents and speed traps, from users⁶²
- GasBuddy: finds and directs users to nearby gas stations sorted by price and distance, including details on other services available⁶³
- Life 360: creates private, invite-only maps showing where family members are—with without sending dozens of text messages⁶⁴
- Uber: hails taxies, private cars or ride shares and allows users to pay for their ride via phone⁶⁵

58 Easton, Richard D. and Eric F. Frazier. *GPS Declassified: From Smart Bombs to Smartphones*. (Lincoln: Potomac Books, University of Nebraska Press), 2013.

59 <http://www.gps-coordinates.net/>

60 Cassavoy, Liane. "21 Awesome GPS and Location-Aware Apps for Android." *PCWorld*. July 31, 2012. http://www.pcworld.com/article/260112/21_awesome_gps_and_location_aware_apps_for_android.html#slide22

61 "Top GPS-based apps for everything from traffic to travel." Qualcomm Snapdragon Blog. June 27, 2013. <https://www.qualcomm.com/news/snapdragon/2013/06/27/top-gps-based-apps-everything-traffic-travel>

62 Cassavoy, Liane. "21 Awesome GPS and Location-Aware Apps for Android." *PCWorld*. July 31, 2012. http://www.pcworld.com/article/260112/21_awesome_gps_and_location_aware_apps_for_android.html#slide22

63 Cassavoy, Liane. "21 Awesome GPS and Location-Aware Apps for Android." *PCWorld*. July 31, 2012. http://www.pcworld.com/article/260112/21_awesome_gps_and_location_aware_apps_for_android.html#slide22

64 "Top GPS-based apps for everything from traffic to travel." Qualcomm Snapdragon Blog. June 27, 2013. <https://www.qualcomm.com/news/snapdragon/2013/06/27/top-gps-based-apps-everything-traffic-travel>

65 <https://www.uber.com/features>

- Yelp: locates and provides user-submitted reviews of businesses and restaurants near users⁶⁶

Each of these mobile applications illustrates the growth of a whole new economic sector based on GPS, digital location-based services, the effects and impacts of which are still being discovered.

Public Safety and Emergencies

While the rise of cellular telephones is often credited with increasing public safety due to citizens' increased ability to reach authorities in the event of an emergency, the true game-changer for public safety came when a great number of cell phones became equipped with GPS. Rather than simply being able to report an emergency, often with little ability to pinpoint the exact location where it occurred, GPS capabilities enabled emergency response personnel to reach the exact location of the affected parties, whether the individual reporting the emergency was familiar with the location or not.⁶⁷

Reducing the time to locate an individual in need saves lives, particularly in disaster events. For example, teams used GPS, combined with other mapping technologies, to find victims, provide aid and improve relief effort response for global disasters such as the 2004 Indian Ocean tsunami, Hurricanes Katrina and Rita in 2005,⁶⁸ and the Ebola outbreak of 2014.⁶⁹ GPS is also used in tracking storms, hurricanes, floods, tornados and wildfires, mitigating both human and economic loss.⁷⁰

In addition, a host of new applications and benefits use GPS technology to enable new ways of increasing public safety. One good example is the popular application ChildChecker—a safety vest marketed by Purple Scout that constantly reports a child's exact location to schools and day care centers charged with monitoring their safety.

66 <http://www.yelp.com/about>

67 Bray, Hiawatha. *You Are Here: From the Compass to the GPS, the History and Future of How We Find Ourselves*. (New York: Basic Books), 2014.

68 "Public Safety & Disaster Relief." GPS.gov. <http://www.gps.gov/applications/safety/>

69 "2014 West Africa Ebola Response." OpenStreetMap. http://wiki.openstreetmap.org/w/index.php/2014_West_Africa_Ebola_Response

70 "Public Safety & Disaster Relief." GPS.gov. <http://www.gps.gov/applications/safety/>

⁷¹ Other examples include SPOT, a personal tracker for outdoor enthusiasts like hikers or skiers,⁷² and MOTOSafety, a tool allowing parents to monitor teen driving safety.⁷³

CORS Impacts

Looking exclusively at CORS, the system is being used across a diversity of industries, with uses including surveying, engineering, vehicle tracking, container tracking and more.⁷⁴ In 2008, CORS data was downloaded 10.6 million times.⁷⁵ A study prepared for the NGS in 2009 found that “the order of magnitude of CORS benefits is estimated as \$758 million per year. ... If benefits grew at a 15 percent annual rate, less than the recent growth rate of 22 percent, the order of magnitude of the present value of CORS benefits over the next 15 years would be \$18.5 billion.”⁷⁶

Industry is not the only sector making use of CORS. NOAA itself uses CORS “to monitor the ionosphere; to measure the troposphere, which helps with short-term weather modeling.” Neil Weston notes, “We can estimate tectonic plate motion and coastal subsidence; enhance tsunami warning systems.” Certain academic institutions also “download all of the CORS data that’s collected every night. They use the CORS data to perform large simulations.”⁷⁷

⁷¹ Easton, Richard D. and Eric F. Frazier. *GPS Declassified: From Smart Bombs to Smartphones*. (Lincoln: Potomac Books, University of Nebraska Press), 2013.

⁷² <http://www.outdoorsafetygps.com>

⁷³ <http://www.motosafety.com>

⁷⁴ GovLab interview with Dr. Neil D. Weston, Chief Scientist, National Geodetic Survey, National Ocean Service, NOAA, September 24, 2015.

⁷⁵ Leveson, Irving. “Socio-Economic Benefits Study: Scoping the Value of CORS and GRAV-D.” Final Report Prepared for the National Geodetic Survey. January 2009. http://www.ngs.noaa.gov/PUBS_LIB/Socio-EconomicBenefitsofCORSandGRAV-D.pdf

⁷⁶ Leveson, Irving. “Socio-Economic Benefits Study: Scoping the Value of CORS and GRAV-D.” Final Report Prepared for the National Geodetic Survey. January 2009. http://www.ngs.noaa.gov/PUBS_LIB/Socio-EconomicBenefitsofCORSandGRAV-D.pdf

⁷⁷ GovLab interview with Dr. Neil D. Weston, Chief Scientist, National Geodetic Survey, National Ocean Service, NOAA, September 24, 2015.

Among other aims, CORS enables the analysis of data related to plate tectonics, precipitable water vapor, free electrons in the atmosphere and much more, to the ends of:⁷⁸

- Defining legal marine and land boundaries
- Creating storm surge models
- Monitoring sea level rise
- Enabling restoration of coastal habitat
- Improving coastal resilience
- Predicting flooding vulnerability
- Mapping shorelines
- Predicting earthquakes
- Assessing hurricane damage
- Improving weather prediction
- Measuring space weather

Challenges

Although GPS has grown and spread rapidly, it has faced, and continues to face, a number of obstacles that either limit its spread or its potential to effect change. These include:

Governance Challenges

Effectively governing GPS poses significant challenges given the many agencies involved in the system and the diversity of needs that it satisfies. As Jason Kim of NCO puts it: “Wrangling all the different cats under one roof is difficult because there really aren’t that many governmental programs that touch so many different parts of the government—pretty much everyone uses GPS.”⁷⁹

⁷⁸ Leveson, Irving. “Socio-Economic Benefits Study: Scoping the Value of CORS and GRAV-D.” Final Report Prepared for the National Geodetic Survey. January 2009. http://www.ngs.noaa.gov/PUBS_LIB/Socio-EconomicBenefitsofCORSandGRAV-D.pdf

⁷⁹ GovLab interview with Jason Kim, Senior Advisor, National Coordination Office for Space-Based Positioning, Navigation, and Timing, September 22, 2015.

Not only are many entities involved in the decision-making processes regarding GPS, but different actors also take the lead on different issues. While the National Executive Committee for Space-Based Positioning, Navigation, and Timing (PNT) is co-chaired by the Department of Defense and Department of Transportation, the DoD is the main driver of GPS in general and the DoT acts as the lead agency for civil uses. The FCC, on the other hand, handles all issues related to the private sector and any other non-federal users of the GPS spectrum. Making matters even more complicated, the Department of Commerce handles federal use of the system, while the National Telecommunications and Information Administration (NTIA) has authority over the GPS spectrum.⁸⁰

All this overlapping authority can lead to turf battles and frequent confusion. Because a specific issue or problem may cover multiple dimensions, it is often unclear which agency or department has authority in a particular case. Kim adds, however, that one concern (and thus one agency) often trumps all others: “When it comes down to the nitty-gritty, hard decisions, who wins out? Well, it’s usually National Security because they’re paying the bills and we’re just along for the ride.”⁸¹

Funding and U.S. Policymaking

The initial investments in GPS have of course long since been recouped. But funding for any new challenges or investments remains challenging. This challenge is closely related to another challenge: GPS’ reliance on U.S. Government decision-making, particularly in these fiscally challenged times. As Kim notes, whenever the possibility of expansion or spreading GPS is considered, “that’s considered a new start and basically, any new start under this current fiscal environment is impossible. ... It’s kind of a sad situation, but it’s true.”⁸²

A deeper, underlying concern here relates to the U.S.’ central role in maintaining and running GPS. This role continues to elicit some

⁸⁰ GovLab interview with Jason Kim, Senior Advisor, National Coordination Office for Space-Based Positioning, Navigation, and Timing, September 22, 2015.

⁸¹ GovLab interview with Jason Kim, Senior Advisor, National Coordination Office for Space-Based Positioning, Navigation, and Timing, September 22, 2015.

⁸² GovLab interview with Jason Kim, Senior Advisor, National Coordination Office for Space-Based Positioning, Navigation, and Timing, September 22, 2015.

international concern, whether over the U.S.’ willingness to keep funding needed investments (for example, to ensure that updates to the system allow for backward compatibility)⁸³ or in ensuring that the system will remain open to international actors and will not return to the two-track system (one for civilian use, one for military) that used to exist. Such concerns were heightened in the wake of the September 11 terrorist attacks, when some observers feared the U.S. might deliberately degrade a system that could in theory be used by terrorists to target the U.S. itself. Today, such concerns are largely unfounded; the current generation of GPS satellites that began to be deployed in 2014 do not have the hardware capability to reintroduce selective availability.⁸⁴

Privacy Concerns

The increased adoption of GPS, driven in large part by a proliferation of GPS-enabled devices, has also led to growing concerns over privacy. While GPS allows individuals to know exactly where they are, it also often allows the makers of devices and applications to access the same information. In a 2010 report, the *Wall Street Journal* found that nearly half (47 out of 101) of the most popular smartphone applications share users’ location with third parties. It is important to note that many of these applications do not require access to users’ location in order to perform their core functions.⁸⁵ They may nonetheless share this data with companies (such as advertisers) or even government agencies. Similar concerns over privacy have swirled around car manufacturers and popular ride-sharing services, many of which collect potentially sensitive location-based information about individuals.⁸⁶

⁸³ Easton, Richard D. and Eric F. Frazier. *GPS Declassified: From Smart Bombs to Smartphones*. (Lincoln: Potomac Books, University of Nebraska Press), 2013.

⁸⁴ Bray, Hiawatha. *You Are Here: From the Compass to the GPS, the History and Future of How We Find Ourselves*. (New York: Basic Books), 2014.

⁸⁵ Thurm, Scott and Yukari Iwatani Kane. “Your Apps Are Watching You.” *The Wall Street Journal*. December 17, 2010. <http://www.wsj.com/articles/SB10001424052748704694004576020083703574602>

⁸⁶ Timberg, Craig. “Is Uber’s rider database a sitting duck for hackers?” *The Washington Post*. December 1, 2014. <https://www.washingtonpost.com/news/the-switch/wp/2014/12/01/is-ubers-rider-database-a-sitting-duck-for-hackers/>

In 2011, a bill was introduced in Congress titled The Geolocation Privacy and Surveillance Act (GPS Act). The bill, sponsored by Rep. Jason Chaffetz and Sen. Ron Wyden, sought to limit government surveillance using geolocation signals from mobile phones and GPS devices. The bill was reintroduced in 2015, but, as of early 2016, has not been passed.⁸⁷ Regardless of such efforts, however, concerns over privacy persist. If left unaddressed or unchecked, they have the potential to reduce trust and thus limit the spread (and potential benefits) of GPS technology.

GPS “Jamming” / Hacking / Security

As GPS grows in popularity and becomes increasingly intertwined with networked devices, the security challenges are also likely to grow. The potential impact of a successful hack into the system could be devastating. As Kim notes: “These satellites are up there. They’re pretty much sitting ducks. If anyone wants to try to mess with them, either through jamming or actually attacking the satellites—there’s just so many ways that you could massively disrupt our security, our economy, our safety.”⁸⁸

The potential impact of an attack that would degrade or otherwise affect GPS signals was felt in 2010, when the Newark Airport found that its new ground-based augmentation system (GBAS)—a system designed to improve the precision of GPS to aid in flight landings—was degrading intermittently. Upon investigation, the FAA found that GPS signals were being lost when truck drivers on the nearby New Jersey Turnpike drove past the site’s antenna while using GPS jammers in order to obscure their movements from managers and dispatchers.⁸⁹ While no major disasters resulted from this incident, it heightened concerns about the potential consequences of a security breach.

It is because of such longstanding concerns that Kim’s office is working to develop a ground-based system for providing redundant GPS

⁸⁷ “Geolocation Privacy Legislation.” GPS.gov. <http://www.gps.gov/policy/legislation/gps-act/>

⁸⁸ GovLab interview with Jason Kim, Senior Advisor, National Coordination Office for Space-Based Positioning, Navigation, and Timing, September 22, 2015.

⁸⁹ Easton, Richard D. and Eric F. Frazier. *GPS Declassified: From Smart Bombs to Smart-phones*. (Lincoln: Potomac Books, University of Nebraska Press), 2013.

signals in the event of any major issues with the satellites' signals. The U.S. government is "so concerned about fielding a backup," he notes, because "we recognize that GPS is a highly vulnerable system. The technology was developed in a different era when we didn't have to worry about hackers or malicious threats through jamming."⁹⁰

CORS Challenges

The central challenges for the CORS system are more in line with other open data initiatives included in this series of case studies. The first involves standardization. NOAA found that many of the people running its sites "wouldn't quite configure their receivers correctly, so it wouldn't output all of the information necessary." NOAA's ongoing outreach efforts and the continued maturity of the system should help to mitigate the standardization challenge.

Second, at the beginning, the sheer volume of data reaching the CORS handlers could be almost overwhelming. Weston recalls, "We were constantly buying hard drives and big systems to do the data management. ... Data comes in 24/7, round-the-clock, and we only had four or five individuals to do this." Weston himself played a central role in mitigating the challenge through automating much of the process: "My scripts were refined over the years to scale up from a handful to a hundred sites across the country to over 2,000."⁹¹

Looking Forward

Notwithstanding the debate over whether GPS can truly be considered an "open data" initiative or not, there is no doubt that GPS stands as a paradigmatic example of the transformative potential represented by accessible and freely available information. Going forward, it is certain that its influence is likely to grow, and be felt in a wider range of sectors and activities. We identify the following as some of the key areas in which GPS' impact is likely to be felt:

⁹⁰ GovLab interview with Jason Kim, Senior Advisor, National Coordination Office for Space-Based Positioning, Navigation, and Timing, September 22, 2015.

⁹¹ GovLab interview with Dr. Neil D. Weston, Chief Scientist, National Geodetic Survey, National Ocean Service, NOAA, September 24, 2015.

Precision Farming

Agriculture is one of the key industries benefiting from GPS. As far back as 1996, around 5 percent of farmers used GPS to better target their efforts—down to fractions of an inch—in order to avoid overlapping the spread of “seeds, irrigation, fertilizers and herbicides.” A 2011 NPD Consulting study estimates that the aggregate annual benefits of GPS to crop farming totals \$19.9 billion per year.⁹² The next wave is so-called precision farming, which is expected to reach a 100 percent adoption rate by the early 2020s.⁹³ Beyond GPS-guided tractors and precision farming strategies, location-aware data collected from farmers is increasingly being aggregated to yield new insights into optimal strategies for agricultural development. Scott Shearer, chair of the Department of Food, Agricultural, and Biological Engineering at Ohio State University notes that this aggregated information can provide prescriptions to enable individual-level farms to, for instance, plant two different hybrid plants in different zones, allowing for improved returns—“the analytics are going to drive the development of those prescriptions.”⁹⁴

That said, major questions still remain about who owns the data generated by farms. Companies like Monsanto and John Deere are leading the precision farming big data revolution, and while providing data to companies that aggregate information from many farms can lead to better individual-level models and more strategic farming, some remain reluctant to surrender their data to large players in the field.⁹⁵

92 Pham, Nam D. “The Economic Benefits of Commercial GPS Use in the U.S. and The Costs of Potential Disruption.” NDP Consulting. June 2011. http://www.gpsalliance.org/docs/GPS_Report_June_21_2011.pdf

93 Easton, Richard D. and Eric F. Frazier. *GPS Declassified: From Smart Bombs to Smartphones*. (Lincoln: Potomac Books, University of Nebraska Press), 2013.

94 Bobkoff, Dan. “Seed by seed, acre by acre, big data is taking over the farm.” *Business Insider*. September 15, 2015. <http://www.businessinsider.com/big-data-and-farming-2015-8>

95 Bobkoff, Dan. “Seed by seed, acre by acre, big data is taking over the farm.” *Business Insider*. September 15, 2015. <http://www.businessinsider.com/big-data-and-farming-2015-8>

Indoor Navigation

One of the key weaknesses of GPS is the lack of indoor navigation capabilities. Taking lessons from GPS and the multi-billion-dollar industry it enables, private sector players are making moves to give consumers the ability to place themselves in large indoor areas. Companies big and small are increasingly providing detailed information on spaces like airports, museums and shopping malls to improve navigation capabilities.⁹⁶ Indeed, analysts predict that by 2018 the indoor location market will swell to \$2.6 billion worldwide from \$449 million in 2013.⁹⁷

A controversial application of indoor GPS can be seen in Amazon warehouses, where the retail giant has its employees wear GPS tags that, combined with a handheld scanner, direct them to the most efficient route to move products. However, critics argue that not only is monitoring an employee's every movement an invasion of privacy, but that such hyper-focus on efficiency is leading to oppressive employee surveillance, resulting in increasingly higher demands from employers and decreasing employee morale.⁹⁸

Intelligent Transportation Systems and Driverless Cars

GPS is in many ways at the heart of efforts by companies like Google, Apple, Tesla and others to introduce driverless cars. In a driverless car, GPS defines the “mission” of the autonomous vehicle by setting the starting and ending point of the drive. It considers all the road options and chooses the best path.⁹⁹ Of course, GPS cannot handle every variable on the road, and to account for the many factors and obstacles, technology and other companies are innovating

⁹⁶ Bray, Hiawatha. *You Are Here: From the Compass to the GPS, the History and Future of How We Find Ourselves*. (New York: Basic Books), 2014.

⁹⁷ Easton, Richard D. and Eric F. Frazier. *GPS Declassified: From Smart Bombs to Smartphones*. (Lincoln: Potomac Books, University of Nebraska Press), 2013.

⁹⁸ Solon, Olivia. “Wearable Technology Creeps into the Workplace.” *Bloomberg Business*. August 7, 2015. <http://www.bloomberg.com/news/articles/2015-08-07/wearable-technology-creeps-into-the-workplace>

⁹⁹ Pullen, John Patrick. “You Asked: How Do Driverless Cars Work?” *Time*. February 24, 2015. <http://time.com/3719270/you-asked-how-do-driverless-cars-work/>

with a form of “differential GPS” that would leverage radar and cameras to help vehicles deal with the shifting dynamics of a road.¹⁰⁰

While the widespread use of driverless cars in many ways still smacks of science fiction, in June 2015, Google announced that its experimental versions eclipsed the 1 million-mile mark—the equivalent of 75 years of typical adult driving. In a Google+ post, the Google Self-Driving Car Project team announced that, “along the way, we’ve navigated more than 200,000 stop signs, 600,000 traffic lights, and seen 180 million vehicles.”¹⁰¹ By 2040, the Institute of Electrical and Electronics Engineers expects that autonomous vehicles will represent three-quarters of cars on the road.¹⁰²

Terrestrial Backups

As mentioned above, the development of a terrestrial backup system to mitigate the impacts of a potential GPS outage—whether intentional or otherwise—is a key priority for the U.S. government. Jason Kim believes that, “we need it like yesterday because the vulnerability is the same it’s always been, but I think the recognition of it is becoming more recognized, is becoming more acute. The threat of exploiting that vulnerability is going way up.” The government is looking to explore potential scenarios—from disruptive solar activity to malicious attacks—and take steps to field a land-based alternative in the event of an interruption to GPS service.¹⁰³

Evolving the CORS System

The CORS network has two main next steps: expanding coverage to the entire U.S., and increasing the capacity of data passing between stations. On the first point, Weston notes that, “there are some areas in the country—North and South Dakota, Wyoming—where there aren’t that many CORS stations. There are enough up there to do a

¹⁰⁰ Pullen, John Patrick. “You Asked: How Do Driverless Cars Work?” *Time*. February 24, 2015. <http://time.com/3719270/you-asked-how-do-driverless-cars-work/>

¹⁰¹ <https://plus.google.com/+SelfDrivingCar/posts/iMHEMH9crJb>

¹⁰² Easton, Richard D. and Eric F. Frazier. *GPS Declassified: From Smart Bombs to Smartphones*. (Lincoln: Potomac Books, University of Nebraska Press), 2013.

¹⁰³ GovLab interview with Jason Kim, Senior Advisor, National Coordination Office for Space-Based Positioning, Navigation, and Timing, September 22, 2015.

certain type of positioning, but the people want the data faster.”¹⁰⁴ This leads into the second point, for NOAA, Weston notes, “our focus is to get the data from all the stations in real time. Once it’s available to NOAA, we can quickly turn it over and make it available to the public. Now, the sooner they have access to it, the sooner they can make use of it. Near real-time data access is critical.”¹⁰⁵ Gaining real-time data poses a few challenges, arising from both the government end and the outside partner segment. Weston explains, “we have to get all the site operators to modify their infrastructure to send the data to us as fast as they can. We, in turn, have to quality control that data in a fast time frame and make it available.”¹⁰⁶ Although increasing the turnaround of CORS data is a focus, Weston is quick to point out that “CORS data is not intended to do real-time positioning.”¹⁰⁷

As the world’s first and only global public utility, GPS is testament to the incredibly deep and broad impact that can arise from a government opening information. Today, GPS has become such an essential part of everyday life, so woven into the fabric of ordinary activities, that its relatively recent creation is often overlooked. While the debate over whether GPS should truly be labeled as “open data” persists, the success of GPS makes clear the type of innovation, economic development and broad social impacts that can follow from the free flow of information to the public.

¹⁰⁴ GovLab interview with Dr. Neil D. Weston, Chief Scientist, National Geodetic Survey, National Ocean Service, NOAA, September 24, 2015.

¹⁰⁵ GovLab interview with Dr. Neil D. Weston, Chief Scientist, National Geodetic Survey, National Ocean Service, NOAA, September 24, 2015.

¹⁰⁶ GovLab interview with Dr. Neil D. Weston, Chief Scientist, National Geodetic Survey, National Ocean Service, NOAA, September 24, 2015.

¹⁰⁷ GovLab interview with Dr. Neil D. Weston, Chief Scientist, National Geodetic Survey, National Ocean Service, NOAA, September 24, 2015.

PART V

Case Studies: Solving Public Problems

Open data is playing an increasingly important role in solving big public problems—from health crises to cases of institutional discrimination. Solutions often arise as a result of citizens and policy-makers accessing new forms of data-driven assessment of the problems at hand. Open data also enables data-driven engagement between stakeholders, producing more targeted interventions and enhanced collaboration.

The case studies in this section examine:

- The coordinated sharing of data among government and international actors during the response to the Ebola outbreak in Sierra Leone.
- The identification of discriminatory water access decisions in the case *Kennedy v the City of Zanesville*, resulting in a \$10.9 million settlement for the African-American plaintiffs.
- Increased awareness among Singaporeans about the location of hotspots for dengue fever transmission.
- Improved, data-driven emergency response following earthquakes in Christchurch, New Zealand.

- Troubling privacy violations on Eightmaps related to Californians' political donation activity.

CHAPTER 16

Battling Ebola in Sierra Leone

Data Sharing to Improve Crisis Response

By Andrew Young and Stefaan Verhulst

Summary: In 2014, the largest Ebola outbreak in history occurred in West Africa. Information on Ebola cases and response efforts was dispersed across a diversity of data collectors, and there was little ability to get relevant data into the hands of those who could leverage it. A number of data-driven initiatives sought to improve the quality of information available to humanitarians working to address the crisis. This case study examines three initiatives in particular—Sierra Leone’s National Ebola Response Centre (NERC), the United Nation’s Humanitarian Data Exchange (HDX) and the Ebola GeoNode—and shows both the potential and challenges of open data projects in combating Ebola specifically, and more generally in addressing humanitarian crises.

Dimension of Impact: Solving Public Problems—Data-Driven Engagement

Key Takeaways:

- Data and information have important roles to play in the battle not just against Ebola, but more generally against a variety of natural and man-made crises.
- However, in order to maximize that potential, it is essential to foster the supply side of open data initiatives—i.e., to ensure the availability of sufficient, high-quality information. This can be especially challenging when there is no clear policy backing to

push actors into compliance and to set clear standards for data quality and format.

- Particularly during a crisis, the early stages of open data efforts can be chaotic, and at times redundant. Improving coordination between multiple actors working toward similar ends—though difficult during a time of crisis—could help reduce redundancy and lead to efforts that are greater than the sum of their parts.

Context and Background

Open Data in Sierra Leone

In 2014, Sierra Leone was ranked 78th in the Open Data Barometer, a ranking which puts it near the bottom. It was nonetheless highlighted as a “country to watch,” due to the country’s Right to Access Information Law, passed in 2013, which requires proactive publication of a diversity of publicly held data and information.¹ The government has in recent years also made some efforts to make data more accessible to the public in digital and non-digital formats.² Also in 2013, Sierra Leone joined the Open Government Partnership (OGP), with open data projects representing a significant portion of its first OGP Action Plan commitments.³

One of the country’s more important open data efforts is Open Data for Sierra Leone,⁴ an initiative of the African Development Bank Group’s Open Data for Africa project. The site contains a number of data sets and infographics focused on economics, demographics, agriculture, energy, health care, education, food security and international trade. Understandably, as of late 2015, the focus of the site appeared to be largely on Ebola cases and fatalities.⁵

1 Abdulai, Saffa. “Connecting Open Data and the Right to Information in Sierra Leone.” Open Data Research Network. <http://www.opendataresearch.org/content/2014/642/connecting-open-data-and-right-information-sierra-leone>

2 <http://www.opendatabarometer.org/report/analysis/rankings.html>

3 http://www.opengovpartnership.org/sites/default/files/SIERRA%20LEONE_FINAL%20NATIONAL%20ACTION%20PLAN_6_15_2014_Final.pdf

4 <http://sierraleone.opendataforafrica.org>

5 <http://sierraleone.opendataforafrica.org>

The 2014 Ebola Outbreak

Ebola virus disease (EVD) was first discovered during two simultaneous outbreaks in 1976, one in Sudan and the other in the Democratic Republic of Congo. The disease takes its name from the Ebola River, where the Congolese outbreak first occurred,⁶ and its symptoms include fever, severe headache, muscle pain, weakness, fatigue, diarrhea, vomiting, abdominal pain and uncontrollable hemorrhaging.⁷

In 2014, the first ever Ebola outbreak to occur in West Africa was also the largest outbreak of the disease in history, and, according to the US Centers for Disease Control and Prevention (CDC), the “first Ebola epidemic the world has ever known.”⁸ From March 2014 to September 2015, 28,355 cases of Ebola were reported, over 66 times the number of cases reported in the previous (second-biggest) outbreak, which occurred in 2000–2001 in Uganda.⁹

The disease quickly spread across countries in West Africa—particularly Sierra Leone, Guinea and Liberia. The epidemic most likely spread so rapidly as a result of traditional burial practices and rituals that put family members of the deceased and those tasked with burying bodies at high risk for infection. Porous borders between the three countries also played an important role.¹⁰ In September 2014, the UN Security Council declared Ebola a “threat to interna-

6 “Ebola virus disease—Fact Sheet.” World Health Organization. Updated January 2016. <http://www.who.int/mediacentre/factsheets/fs103/en/>

7 “Ebola (Ebola Virus Disease)—Signs and Symptoms.” Centers for Disease Control and Prevention. <http://www.cdc.gov/vhf/ebola/symptoms/>

8 “Ebola (Ebola Virus Disease)—Outbreaks.” Centers for Disease Control and Prevention. <http://www.cdc.gov/vhf/ebola/outbreaks/>

9 “Ebola (Ebola Virus Disease)—Outbreaks Chronology: Ebola Virus Disease.” Centers for Disease Control and Prevention. <http://www.cdc.gov/vhf/ebola/outbreaks/history/chronology.html>

10 Garrett, Laurie. “Ebola’s Lessons: How the WHO Mishandled the Crisis.” *Foreign Affairs*. September/October 2015. <https://www.foreignaffairs.com/articles/west-africa/2015-08-18/ebolas-lessons>

tional peace and security”;¹¹ experts predicted that the disease could kill 100,000 people and require \$1 billion to successfully contain.¹²

On September 19, 2014, the United Nations Mission for Ebola Emergency Response (UNMEER) was established.¹³ UNMEER was “mandated as a temporary measure to harness the capabilities of all the relevant UN actors under a singular operational crisis management system to reinforce unity of purpose among responders and to ensure a rapid and effective response to the Ebola crisis.”¹⁴ UNMEER headed up the UN’s efforts to stop Ebola in West Africa until August 1, 2015, when the World Health Organization (WHO) took over a central coordinating role. Coordination between actors within Sierra Leone and across the international community proved essential to battling the Ebola crisis. Indeed, as the projects discussed below (and in several other case studies in this report) indicate, coordination, collaboration and information-sharing are central in the response to a variety of humanitarian and other crises.

Project Description and Inception

Using Data in Combating Ebola

Early on, it became apparent that information would be key in the battle against Ebola. Efforts to combat the epidemic were, for example, hampered by limited information sharing between national governments, aid organizations and front-line actors like the rural health clinics that often bore the brunt of the crisis. Even the most basic information—for example, the number of cases or dead—was hard to come by, making it difficult to assess the severity of the epi-

11 United Nations Security Council Resolution 2177 (2014). <http://www.ifrc.org/docs/IDRL/UN%20SC%20Res.pdf>

12 Spooner, Samantha. “Ebola could kill 100,000 and needs \$1 billion to defeat, but Nigeria shows the world how it’s done.” *Mail & Guardian Africa*. September 21, 2014. <http://mgafrica.com/article/2014-09-19-as-the-war-against-ebola-rages-nigeria-shows-the-world-how-its-done>

13 Garrett, Laurie. “Ebola’s Lessons: How the WHO Mishandled the Crisis.” *Foreign Affairs*. September/October 2015. <https://www.foreignaffairs.com/articles/west-africa/2015-08-18/ebolas-lessons>

14 “UNMEER-WHO Transition Message.” World Health Organization Regional Office for Africa. <http://www.afro.who.int/en/sierra-leone/press-materials/item/7897-unmeer-who-transition-message.html>

demic and target interventions. For all these reasons, the UNMEER chief, Anthony Banbury, promised in November 2014 that “We’re going to be really data- and information-driven and that will strengthen our response.”¹⁵

Recognizing the importance of information—governments, aid agencies and international organizations soon began designing and implementing a number of data tools. This case study focuses on three tools and initiatives that played a particularly important role in mitigating the public health crisis: the National Ebola Response Centre (NERC), the Humanitarian Data Exchange (HDX) and Ebola GeoNode.

National Ebola Response Centre (NERC)

In October 2014, Sierra Leone created the NERC, which led all facets of the on-the-ground Ebola response efforts in the country and coordinated the activities of District Ebola Response Centres (DERCs).¹⁶ NERC brought together the Presidency of the Republic of Sierra Leone with the ministries of Health and Sanitation; Local Government and Rural Development; Social Welfare, Gender and Children’s Affairs; Foreign Affairs; Finance and Economic Development; Defence; Internal Affairs; and Information and Communication. In addition, the NERC coordinated with external agencies and groups like the CDC, Red Cross, World Bank, UK Department for International Development (DFID), African Development Bank and the U.S. Embassy.

NERC was one of many efforts by international organizations and government actors to help domestic efforts to mitigate the effects of the disease and stop its spread. The African Governance Initiative (AGI), based in the United Kingdom, for instance, helped to set up a Situation Room within NERC, which focused on putting the most important information into the hands of the decision-makers in the best position to act on it. In many ways, the Situation Room—and

¹⁵ Westcott, Lucy. “Q&A with Anthony Banbury: UN’s Ebola Mission Can Still Do More.” *Newsweek*. November 24, 2014. <http://www.newsweek.com/qa-anthony-banbury-un-s-ebola-mission-can-still-do-more-286724>

¹⁶ Maxmen, Amy. “How Ebola Found Fertile Ground in Sierra Leone’s Chaotic Capital.” *National Geographic*. January 27, 2015. <http://news.nationalgeographic.com/news/2015/01/150127-ebola-virus-outbreak-epidemic-sierra-leone-freetown-photos-pictures/>

NERC more generally—were primarily efforts to create the infrastructure for collecting and disseminating important information related to the crisis. This infrastructure was sorely needed because, as OB Sisay, the director of the Situation Room, put it: “The core crux of the problem is not Ebola. It’s system failure.”¹⁷

Rupert Simons, CEO of Publish What You Fund and formerly an adviser to AGI, helped to set up the Situation Room and led the team as it conducted daily data collection efforts of conditions on the ground. A key piece of these efforts involved the design, implementation and maintenance of a reporting system for the 14 districts in Sierra Leone. Simons describes the reporting system as one where districts were asked to “call or email us once a day with information on the response. We didn’t ask them how many Ebola patients they had; the Ministry of Health knows that already. But we did ask how many safe beds they had, how many calls they got every day to investigate a case, and how many people had died, because everyone who dies needs to be given a safe medical burial.”¹⁸

Armed with such on-the-ground information, the Situation Room hosted twice-daily briefings for leadership at NERC.¹⁹ These briefings, focused on identifying and operationalizing key action points based on the day’s information, typically included officials from the government, UNMEER and other UN agencies, and governmental and nongovernmental aid agencies.²⁰

¹⁷ Maxmen, Amy. “How Ebola Found Fertile Ground in Sierra Leone’s Chaotic Capital.” *National Geographic*. January 27, 2015. <http://news.nationalgeographic.com/news/2015/01/150127-ebola-virus-outbreak-epidemic-sierra-leone-freetown-photos-pictures/>

¹⁸ Simons, Rupert. “Inside the National Ebola Response Centre.” Africa Governance Initiative. October 17, 2014. <http://www.africagovernance.org/article/inside-national-ebola-response-centre>

¹⁹ GovLab interview with Rupert Simons, Former Advisor to the African Governance Initiative; CEO Publish What You Fund, August 6, 2015.

²⁰ Example brief: http://nerc.sl/sites/default/files/docs/20151126_Nerc_Eve_Brief%20Count%20Down.pptx%201.pdf



Key Points:

- Nine cases were confirmed EVD+ during the reporting week. Three from Western Area Urban, two in Port Loko and four in Kambia district. Three of the cases were swab positives and 6 live blood cases
- The NERC is monitoring 9 EVD events across the country. One in Western Area Urban and 4 each in Kambia and Port Loko district
- The NERC principals and other partners on the 4th July 2015 visited Magazine Wharf, the Voluntary Quarantine Facility and the WA ERC. The purpose of the visit was to assess how the events at Magazine is being managed at the Wharf.
- A number of households were released from quarantine in Kaffu BulloM and Bureh Kasseh Maconteh(BKM) chiefdoms. This has led to the closure of five EVD events.
- Meeting for all Traditional Chiefs was held in Port Loko district. The purpose was to engage traditional chiefs on the revised by-laws on EVD regulations.
- There are a total of 1,514 contacts across the country. Six hundred and sixty seven in Kambia, 526 in Port Loko, 207 in the Western Area Urban and 114 at the Voluntary Quarantine Facility(VQF)

Figure 16-1. Portion of National Ebola Response Centre Weekly Newsletter

NERC also shared a daily evening briefing with the public (usually by radio), as well as a weekly newsletter and press conference.²¹ The public briefings included a wide variety of data on the current Ebola situation, including facts and figures on the status of the disease (often broken up by region), the location and accessibility of treatment centers, and other information on steps to address the crisis.²²

Humanitarian Data Exchange (HDX)

On July 15, 2014, at the Open Knowledge Festival in Berlin, the United Nations Office for the Coordination of Humanitarian Affairs (OCHA) announced the development of the Humanitarian Data Exchange (HDX). HDX was intended as “a new data sharing platform that encompasses the best standards in data collection, offering

²¹ GovLab interview with Rupert Simons, Former Advisor to the African Governance Initiative; CEO Publish What You Fund, August 6, 2015.

²² Example newsletter: <http://199.230.52.230/sites/default/files/docs/National%20%20Ebola%20Response%20%20Centre%20-%20%20Weekly%20%20%20Newsletter%20%2029th%20June%20-5th%20July%202015%20%282%29.pdf>

access to useful and accurate data.”²³ Designed to cover a wide range of humanitarian crises around the world, the platform was originally focused on two pilot locations in Kenya and Colombia. But as the Ebola crisis worsened, HDX’s potential in West Africa quickly became apparent and it became one of the most important data efforts in the fight against the epidemic.

HDX has been referred to as a “humanitarian data-centric Wikipedia.”²⁴ It allows users to track and follow specific data sets, create curated organization data hubs, and share data across previously siloed organizations working to improve humanitarian efforts around the world.²⁵ According to Javier Teran, an OCHA statistician who works on HDX, the platform focused from the start on the “baseline indicators that humanitarians are always using, like population statistics and mortality rates—the types of data related to a crisis, before, during and after.”²⁶ At its launch, HDX held around 1,600 files, covering a range of regions and humanitarian concerns—but none of those files were targeted to the use case that would take HDX from being a largely speculative platform to one being put to real practical application.²⁷

By late 2014, Teran recalls, “we [the team behind HDX] were appointed by the UN to be the platform for Ebola data exchange.”²⁸ At that point, the original 1,600 files on the platform were significantly increased in number, mostly with the addition of regional data—from Sierra Leone, Guinea and Liberia—drawn from WHO.

23 “Humanitarian Data: OCHA launches ground-breaking data exchange platform.” United Nations Office for the Coordination of Humanitarian Affairs. July 15, 2014. <http://www.unocha.org/top-stories/all-stories/humanitarian-data-ocha-launches-ground-breaking-data-exchange-platform>

24 Griliopoulos, Dan. “Data exchange helps humanitarians act fast and effectively.” *The Guardian*. December 29, 2014. <http://www.theguardian.com/global-development/2014/dec/29/humanitarian-data-exchange-ebola-refugees>

25 Griliopoulos, Dan. “Data exchange helps humanitarians act fast and effectively.” *The Guardian*. December 29, 2014. <http://www.theguardian.com/global-development/2014/dec/29/humanitarian-data-exchange-ebola-refugees>

26 GovLab interview with Javier Teran, Statistician, United Nations Office for the Coordination of Humanitarian Affairs, Humanitarian Data Exchange, September 18, 2015.

27 GovLab interview with Javier Teran, Statistician, United Nations Office for the Coordination of Humanitarian Affairs, Humanitarian Data Exchange, September 18, 2015.

28 GovLab interview with Javier Teran, Statistician, United Nations Office for the Coordination of Humanitarian Affairs, Humanitarian Data Exchange, September 18, 2015.

WHO fed data into the platform on the number of Ebola cases and fatalities, the locations of cases, the amount of money being spent on the crisis, as well as information on Ebola Treatment Centres (ETCs —e.g., how many were open at the time, how many were eventually planned to be opened, and their locations).²⁹ On-the-ground WHO representatives were tasked with collecting, validating and curating the information.³⁰

As of early 2016, HDX held 178 data sets on the Sierra Leone location page, including information on the number of health-care workers infected; status and location of ETCs; status and location of safe and dignified burial teams; location of education establishments; status and location of Ebola Community Care Centers, and much more.³¹ The expected opening date for new ETCs was one of the most important pieces of information in HDX.³² This information was immediately made open in computable formats to analysts and developers working on Ebola response.

In order to collect all this information, the HDX team combined WHO situation report data with information pulled from existing data sources, like OpenStreetMap’s geospatial data and OCHA’s common operational data sets.³³

A central task confronting the HDX team from the start was the need to “make connections with the different players in the field.”³⁴ Teran notes that, “it was a very difficult assignment, especially at the beginning of the crisis,” due to a lack of sophistication in existing information systems. For example, it took some time for WHO representatives “to consolidate the information, to standardize it, to make it comparable” with information from other sources and other countries. This challenge, like Ebola itself, was not limited to Sierra

29 GovLab interview with Javier Teran, Statistician, United Nations Office for the Coordination of Humanitarian Affairs, Humanitarian Data Exchange, September 18, 2015.

30 GovLab interview with Javier Teran, Statistician, United Nations Office for the Coordination of Humanitarian Affairs, Humanitarian Data Exchange, September 18, 2015.

31 <https://data.hdx.rwlabs.org/group/sle>

32 GovLab interview with Javier Teran, Statistician, United Nations Office for the Coordination of Humanitarian Affairs, Humanitarian Data Exchange, September 18, 2015.

33 GovLab interview with Javier Teran, Statistician, United Nations Office for the Coordination of Humanitarian Affairs, Humanitarian Data Exchange, September 18, 2015.

34 GovLab interview with Javier Teran, Statistician, United Nations Office for the Coordination of Humanitarian Affairs, Humanitarian Data Exchange, September 18, 2015.

Leone. “We were dealing with not only one country, but with three countries,” says Teran.³⁵ This need to work across borders added a significant degree of complexity to the HDX effort, and indeed to the other data projects discussed in this case study.

As the HDX platform matured, visualizations, many created with platforms like Tableau and CartoDB, were also added. The top of the Ebola Crisis page featured a link to each of the most important data sets on the platform: Cumulative Cases of Ebola, Cumulative Deaths from Ebola, Response Plan Coverage, People Receiving Food Assistance, and Open Ebola Treatment Centers. Each of these could be viewed as raw data or as user-friendly visualizations. According to Teran, the statisticians leveraging HDX data were “trying to convert all these spreadsheets, CSVs, and maps into information” that was more easy to digest and understand.³⁶ While the raw data was essential for policymakers and others fighting the disease, the visualizations played a particularly important role in disseminating information through intermediaries like the media.

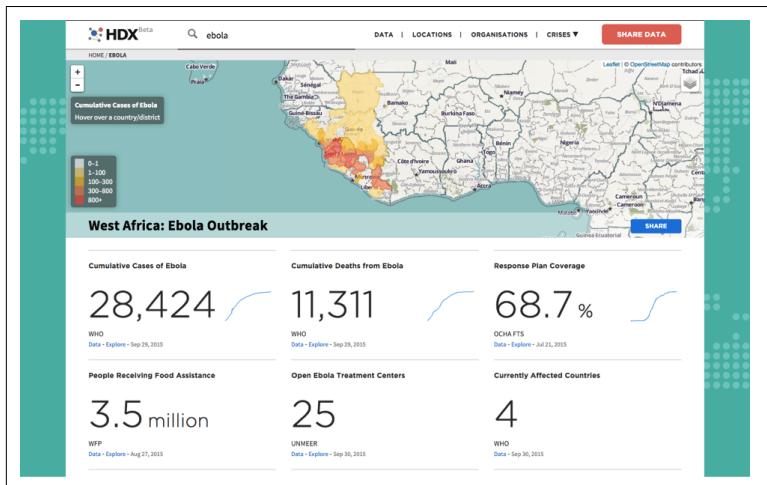


Figure 16-2. Humanitarian Data Exchange Ebola Crisis Page

³⁵ GovLab interview with Javier Teran, Statistician, United Nations Office for the Coordination of Humanitarian Affairs, Humanitarian Data Exchange, September 18, 2015.

³⁶ GovLab interview with Javier Teran, Statistician, United Nations Office for the Coordination of Humanitarian Affairs, Humanitarian Data Exchange, September 18, 2015.

The Ebola GeoNode

The result of a collaboration with the Red Cross, World Bank, Global Facility for Disaster Reduction and Recovery (GFDRR), UNMEER and the US Humanitarian Information Unit (HIU), Ebola GeoNode is an open source geospatial platform that lets users build maps and conduct geospatial analysis on Ebola's impacts in West Africa. According to Patrick Dufour, a former senior Web GIS developer at HIU, Ebola GeoNode is primarily "an open data platform" designed with the intent "to make as much [data] as you can, open."³⁷

As of early 2016, GeoNode consists of 61 data sets containing data on, among other things, administrative boundaries in affected countries, transportation and logistics information, and geo-tagged health crisis data. According to Dufour, logistics data—like the location of ETCs health facilities—occupy an especially important piece of the platform. While simple, such information helped to ensure that on-the-ground actors, like representatives of USAID and UNMEER, had a clear understanding of the location of important points—"sav[ing] a bunch of people time."³⁸

The GeoNode is more than a simple collection of points on a map, however. Information on the platform exists in three categories: Layers, Maps and Documents. In the Layers and Maps categories, users have the option of manipulating the information directly on the GeoNode website or downloading it for any other use. For example, when selecting the Sierra Leone Community Care Centres (CCC) layer, the user can download the geospatial data on the CCC's locations to her hard drive in a variety of formats, *or* create a new map on GeoNode using the CCC data as one layer, with the option of adding additional layers—like the location of global supply routes.³⁹ Similarly, the eight user-generated maps on the platform—Food Insecurity (Estimates), Health Facilities, Mali Admin Boundaries, Sierra Leone Admin Boundaries, Guinea Admin Boundaries,

³⁷ GovLab interview with Patrick Dufour, Former Senior Web GIS Developer, U.S. Department of State, Humanitarian Information Unit, December 17, 2015.

³⁸ GovLab interview with Patrick Dufour, Former Senior Web GIS Developer, U.S. Department of State, Humanitarian Information Unit, December 17, 2015.

³⁹ http://ebolageonode.org/layers/geonode%3Asle_heal_pt_unmeer_ccc

Liberia Admin Boundaries, and two Ebola Care Facilities maps⁴⁰—can be viewed, shared and printed as is, or duplicated and further built upon with additional geospatial data layers. The Documents category includes mapped trend analyses, such as “Evolution of confirmed Ebola cases over the period from 12 January to 22 February 2015.”⁴¹ These documents can be accessed directly, or the user can download the underlying metadata in a number of formats.

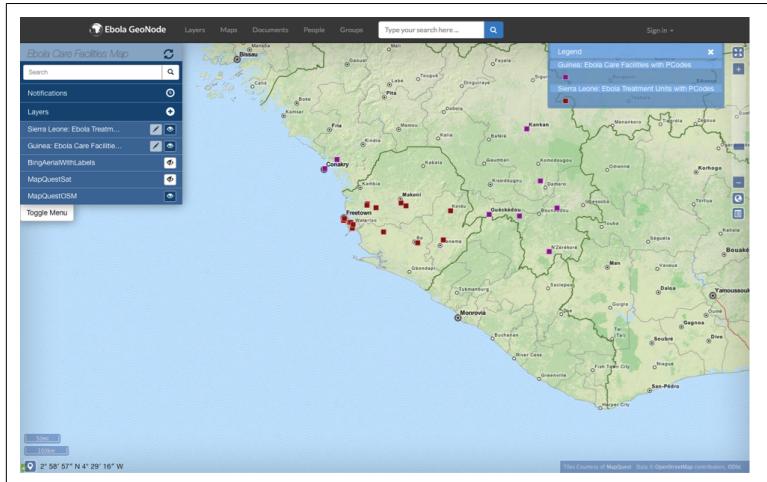


Figure 16-3. Ebola Care Facilities Map on Ebola GeoNode

The central purpose of the GeoNode platform was to alleviate information fragmentation that hampered the fight against Ebola. As the GFDRR website puts it: Data fragmentation meant that “field staff had to devote scarce time to finding and reassembling the data sets. In many cases, this task was far more difficult than it needed to be.”⁴² GeoNode played a particularly important role in enabling collaboration between employees working at international institutions and those on the ground in affected countries, particularly UNMEER representatives and employees of MapAction, a humanitarian mapping charity.⁴³ Vivien Deparday, a disaster risk management specialist at GFDRR, points out that much of the work for those

40 <http://ebolageonode.org/maps/#?limit=100&offset=0>

41 <http://ebolageonode.org/documents/541>

42 <https://www.gfdrr.org/ebola-geonode>

43 <http://www.mapaction.org>

maintaining the GeoNode platform outside of West Africa was “providing technical support on the platform” for those on the ground and “maintaining some authoritative data sets” that could be used and supplemented by the UNMEER and MapAction representatives.⁴⁴ Dufour agrees that the work of the people on the ground was critical for collecting and operationalizing the relevant data for the GeoNode—“otherwise, it’s just a bunch of people at headquarters talking to each other.”⁴⁵

Impact

As many of the case studies in this series show, information plays an important role in a variety of natural and man-made humanitarian disasters. Simons, who was on the forefront of the Ebola crisis, argues that, amid an epidemic, “data is absolutely crucial.” He adds that it is especially important in the early stages of an epidemic, when responders must decide “how to deploy the initially scarce resources.”⁴⁶ In this section, we examine some of the most important ways in which the data-driven efforts examined above had an impact on the Ebola crisis in Sierra Leone in particular, and West Africa in general.

Providing Evidence to Decision-makers

Across the projects included in this case study, one particularly important form of impact stands out: Data plays a critical role in giving policymakers a better understanding of conditions on the ground, and thus an evidence basis for their decisions. For example, HDX data was accessed (and supplied) by the UN World Food Programme (WFP), which turned out to be one of the main users of the platform. According to Teran, the WFP used HDX to understand how Ebola was affecting farmers and to forestall any potential food shortages. When Ebola hits agricultural communities, says Teran,

⁴⁴ GovLab interview with Vivien Deparday, Disaster Risk Management Specialist, Open Data for Resilience Initiative Operational Deployment Lead, Global Facility for Disaster Risk and Reduction, World Bank, December 17, 2015.

⁴⁵ GovLab interview with Vivien Deparday, Disaster Risk Management Specialist, Open Data for Resilience Initiative Operational Deployment Lead, Global Facility for Disaster Risk and Reduction, World Bank, December 17, 2015.

⁴⁶ GovLab interview with Rupert Simons, Former Advisor to the African Governance Initiative; CEO Publish What You Fund, August 6, 2015.

“nobody collects the rice from the fields from the paddy, and then there is no way [the farmer] can bring it to market, and then it’s a domino effect.”⁴⁷ WFP used data tracking and visualization tools on HDX to monitor these potential problems and to help the relevant authorities develop an appropriate response.

While NERC also provided information to decision-makers within international aid and governance organizations, as Simons notes, “the priority at the time was to make good data available to decision-makers within the government of Sierra Leone.”⁴⁸ In a post describing his time at NERC, Simons argues that, although much of the media focus during the crisis was trained on international organizations, “99% of those fighting [Ebola] in the field are Sierra Leoneans. Their government needs to lead that fight. The Situation Room provides them with the information to do so.”⁴⁹

The Impact of Maps

For the Ebola GeoNode in particular, and efforts to increase access to geospatial data more generally, assessing impact involves what Patrick Dufour of HIU calls a “perennial hard question”:⁵⁰ What is the impact of maps? The impacts are perhaps clearer in the case of Ebola response than in other uses. For instance, an article on OpenStreetMap (OSM), a free, crowdsourced mapping tool that provided information to NERC, HDX and GeoNode, notes that: “In the parts of West Africa affected by the Ebola epidemic, Google barely has the roads mapped out. More often than not, the names of villages are missing—and sometimes the village altogether.”⁵¹ Simons also recognizes the importance of the OSM data, noting that prior to its exis-

⁴⁷ GovLab interview with Javier Teran, Statistician, United Nations Office for the Coordination of Humanitarian Affairs, Humanitarian Data Exchange, September 18, 2015.

⁴⁸ GovLab interview with Rupert Simons, Former Advisor to the African Governance Initiative; CEO Publish What You Fund, August 6, 2015.

⁴⁹ Simons, Rupert. “Inside the National Ebola Response Centre.” Africa Governance Initiative. October 17, 2014. <http://www.africagovernance.org/article/inside-national-ebola-response-centre>

⁵⁰ GovLab interview with Patrick Dufour, Former Senior Web GIS Developer, U.S. Department of State, Humanitarian Information Unit, December 17, 2015.

⁵¹ Cassano, Jay. “Inside the Crowdsourced Map Project That Is Helping Contain the Ebola Epidemic.” *Fast Company*. October 22, 2014. <http://www.fastcolabs.com/3037350/elasticity/inside-the-crowdsourced-map-project-that-is-helping-contain-the-ebola-epidemic>

tence, “the [only] way to find out where to go was driving around and asking for directions.”⁵²

The geospatial information, often provided by OSM and available on the GeoNode and HDX, demonstrates how filling an information void—even with something as simple as road information—can play an important role in fighting an epidemic as complex as Ebola. And though difficult to quantify, Dufour is “comfortable knowing that [the GeoNode] was a success” as a result of “knowing that some guy, when he was making his daily situation report, he was using the Ebola GeoNode every day.”⁵³

Spurring Wider Use of Open Data in International Organizations

As mentioned, one of the unexpected benefits—“a big achievement” in the words of Teran—that came from HDX’s efforts in West Africa was inspiring greater uptake of open data in other international organizations. For example, as the WFP began publishing its food-price data as part of an effort to combat Ebola, it became more aware of the benefits of data and, in particular, open data portals. The organization has now increased its commitment to open data, publishing a range of information about food prices from around the world.⁵⁴ As of early 2016, the WFP organization page on HDX houses 19 data sets, including those pertaining to global food prices and food security in 70 locations. This is just one example of a phenomenon we have seen repeatedly across the projects examined in this report: how a single, localized use of open data can spur a much wider recognition of data’s value and potential.

Patrick Dufour from HIU sees a similar development of an open data “ecosystem” that he believes “is going to have an impact on long-term humanitarian information sharing.”⁵⁵ He points in partic-

⁵² GovLab interview with Rupert Simons, Former Advisor to the African Governance Initiative; CEO Publish What You Fund, August 6, 2015.

⁵³ GovLab interview with Patrick Dufour, Former Senior Web GIS Developer, U.S. Department of State, Humanitarian Information Unit, December 17, 2015.

⁵⁴ GovLab interview with Javier Teran, Statistician, United Nations Office for the Coordination of Humanitarian Affairs, Humanitarian Data Exchange, September 18, 2015.

⁵⁵ GovLab interview with Patrick Dufour, Former Senior Web GIS Developer, U.S. Department of State, Humanitarian Information Unit, December 17, 2015.

ular to the collaborative and data-sharing culture built as a result of the Ebola GeoNode and HDX working together. Vivien Deparday, from GeoNode, agrees, seeing a “complementary role” played by GeoNode and HDX in combating Ebola, and adding that “we will be trying to formalize a bit more of these roles in future scenarios.”⁵⁶

Internal Capacity Building and Identifying Best Practices

Finally, given that both GeoNode and HDX were launched by organizations playing a direct role in the Ebola response, the information provided through the platforms was also useful in improving their own response efforts, as well as establishing internal best practices. Dufour, for example, emphasizes that HIU used the GeoNode to create its own map products and target response efforts. “I think that phrase is to eat your dog food,” he says, emphasizing the valuable contribution made by the GeoNode platform to HIU’s efforts on the ground.⁵⁷ Taking the longer view, HDX applied the lessons learned during the Ebola response in its next major humanitarian initiative: the Nepal Earthquake of April 2015. Teran notes that when HDX began work for Nepal, the team knew to quickly build a partnership with the national government, and to immediately identify and liaise with the main organizational player on the ground—in this case the International Organization for Migration (IOM).⁵⁸

Further, many of the databases and data products created during the Ebola epidemic are likely to have a certain “asset value” that will extend beyond this immediate crisis. This is particularly true of the maps and cartographic products generated during the crisis, which are useful for a variety of circumstances beyond combating Ebola.

⁵⁶ GovLab interview with Vivien Deparday, Disaster Risk Management Specialist, Open Data for Resilience Initiative Operational Deployment Lead, Global Facility for Disaster Risk and Reduction, World Bank, December 17, 2015.

⁵⁷ GovLab interview with Patrick Dufour, Former Senior Web GIS Developer, U.S. Department of State, Humanitarian Information Unit, December 17, 2015.

⁵⁸ GovLab interview with Javier Teran, Statistician, United Nations Office for the Coordination of Humanitarian Affairs, Humanitarian Data Exchange, September 18, 2015.

Challenges

While the initiatives included in this case study had common areas of impact, they also faced common challenges. We have identified **three** main challenges that posed barriers to greater impact and effectiveness. Although we discuss these challenges with specific reference to the fight against Ebola, they also surface in many of the other case studies in this report, and have broader applicability to open data initiatives around the world.

Data Management

Perhaps the main challenge faced by many of the initiatives under discussion pertained to data quality. Some of the problems stemmed from quite simple—yet ultimately consequential—errors. For example, on the Ebola GeoNode, it was sometimes difficult to track when data was out of date. Dufour says that it was difficult to determine if a date referenced when data was published, collected or edited.⁵⁹ Likewise, simple misspellings of a patient’s name could cause problems on the HDX platform, leading—at least initially—to what Teran says was significant data duplication. For example, a patient’s name could be spelled differently at each stage of treatment (potential, suspected, confirmed), leading to that same patient being counted multiple times.⁶⁰

The way the data was handled at ETCs also posed challenges for HDX. Internal capacity at ETCs, and particularly the speed with which they could report data to HDX, was, understandably, not always as smooth as would be hoped. More generally, the sheer task of coordinating Ebola response efforts across three countries created tremendous challenges. Teran recalls that data would arrive at different times from different countries, or that on any given day one or more country’s numbers would simply never arrive—“it was always fluctuating.”⁶¹ He adds that these and other data problems were particularly acute in the early stages of the epidemic, that the

⁵⁹ GovLab interview with Patrick Dufour, Former Senior Web GIS Developer, U.S. Department of State, Humanitarian Information Unit, December 17, 2015.

⁶⁰ GovLab interview with Javier Teran, Statistician, United Nations Office for the Coordination of Humanitarian Affairs, Humanitarian Data Exchange, September 18, 2015.

⁶¹ GovLab interview with Javier Teran, Statistician, United Nations Office for the Coordination of Humanitarian Affairs, Humanitarian Data Exchange, September 18, 2015.

crisis was an ongoing learning process for all involved, and that many of the data challenges eased as time went on.⁶²

Collaborating with Government

Responding to an in-process crisis created major challenges of coordination and governance. For example, while on the ground in Sierra Leone working with NERC, Rupert Simons found that, “people spent a lot of time in coordination meetings instead of doing anything.”⁶³ Coordination with national governments could be particularly challenging, forcing humanitarian workers to master the intricacies of national and regional politics while in the midst of a crisis. Simons put it this way: “Even a crisis does not suspend the rule of Sierra Leonean government and politics, where information is power.” This was especially true of the relationship between the Ministry of Health and the National Ebola Response Centre, which—though nominally partners—often acted as rivals.⁶⁴

Others were more positive, but still pointed to challenges. Teran, for example, says that while the Sierra Leonean government in general collaborated well, the process for getting its data into HDX was not always *technically* straightforward.⁶⁵ Before data made its way onto HDX, it first had to pass from the government to WHO in the form of a report. WHO then captured the information, sent it to their headquarters in Geneva, and finally HDX gained access and was able to share the information on the platform. This extended process meant that while HDX had a direct line to data held by WHO, that data was not in fact real time.⁶⁶

62 GovLab interview with Javier Teran, Statistician, United Nations Office for the Coordination of Humanitarian Affairs, Humanitarian Data Exchange, September 18, 2015.

63 GovLab interview with Rupert Simons, Former Advisor to the African Governance Initiative; CEO Publish What You Fund, August 6, 2015.

64 GovLab interview with Rupert Simons, Former Advisor to the African Governance Initiative; CEO Publish What You Fund, August 6, 2015.

65 GovLab interview with Javier Teran, Statistician, United Nations Office for the Coordination of Humanitarian Affairs, Humanitarian Data Exchange, September 18, 2015.

66 GovLab interview with Javier Teran, Statistician, United Nations Office for the Coordination of Humanitarian Affairs, Humanitarian Data Exchange, September 18, 2015.

Lack of Existing Structures and Best Practice

Much of the effort to fight Ebola was done “on the fly” without the benefit of established procedures and practices. This made it substantially more difficult to mount an effective response. “[R]elationships and platforms should be established before a crisis happens so that you already have this relationship in place,” says Dufour, of HIU.⁶⁷ Deparday of GFDRR agrees, arguing that “it’s better to have a GeoNode deployed before crisis. It’s definitely hard to deploy it in the midst of a crisis.”

Perhaps surprisingly, the challenges arising from the “on the fly” GeoNode deployment were not primarily technical. Dufour and Deparday agree that setting up the platform and creating the technical infrastructure was in many ways the easy part. The difficult part was really in creating institutional frameworks, establishing relationships, and defining shared best practices and procedures that can be used by all responders.^{68,69}

In the wake of the crisis, a panel of independent experts comprising the Ebola Interim Assessment Panel concluded that the lack of infrastructure and established best practices played a significant role in hampering the response. Among the conclusions reached by the panel were:⁷⁰

- National surveillance activities need to be better integrated with existing components of health systems;
- Data aggregation and sharing was often limited, or nonexistent;
- Stronger collaboration was required, in particular between the private and public sectors.

⁶⁷ GovLab interview with Patrick Dufour, Former Senior Web GIS Developer, U.S. Department of State, Humanitarian Information Unit, December 17, 2015.

⁶⁸ GovLab interview with Patrick Dufour, Former Senior Web GIS Developer, U.S. Department of State, Humanitarian Information Unit, December 17, 2015.

⁶⁹ GovLab interview with Vivien Deparday, Disaster Risk Management Specialist, Open Data for Resilience Initiative Operational Deployment Lead, Global Facility for Disaster Risk and Reduction, World Bank, December 17, 2015.

⁷⁰ “Report of the Ebola Interim Assessment Panel.” World Health Organization. May 2015. <http://www.who.int/entity/csr/resources/publications/ebola/report-by-panel.pdf?ua=1>

The Ebola crisis had much to teach the international humanitarian community. These and other lessons are now being digested, processed, and hopefully will inform responses to future crises.

Looking Forward

In many ways, the Ebola crisis has established a set of best practices for future data usage in response to humanitarian crises. The teams behind HDX and GeoNode have many plans for the future—how to grow their platforms, how to use them in other contexts and geographies, and how to make them even more effective. Here, we highlight just a few of the plans for HDX and GeoNode, and describe the fate of NERC.

HDX: Enabling New Insights into the Spread of Ebola

The experience of HDX in West Africa during the Ebola crisis is teaching lessons to people around the world. Researchers at MIT and the University of Virginia, for example, are using the platform to analyze how the virus spread in affected countries. This analysis will focus especially on the cultural factors that enabled the rapid spread of the virus—for example, the tendency for families in the region to come together when someone is sick, rather than avoiding the infected person.⁷¹ Although no one is arguing for major cultural changes to occur—many people in Sierra Leone are unlikely to accept cremation, for example—researchers and HDX’s organizers believe that analysis of the data found on the platform can help uncover ways to stop the virus’ spread. One particularly important avenue of research involves the need for safer burial practices, which played a central role in spreading the virus.⁷²

In general, says Teran, HDX is seeking to “facilitate” the flow of information to parties that can yield potentially lifesaving insights from the platform. These parties include a wide range of actors—individual university researchers, and media groups like the *New*

⁷¹ GovLab interview with Javier Teran, Statistician, United Nations Office for the Coordination of Humanitarian Affairs, Humanitarian Data Exchange, September 18, 2015.

⁷² GovLab interview with Javier Teran, Statistician, United Nations Office for the Coordination of Humanitarian Affairs, Humanitarian Data Exchange, September 18, 2015.

*York Times*⁷³ and *National Geographic*.⁷⁴ The licensing system applied to the HDX data—users are allowed to profit from the use of the information as long as credit is given to the original data suppliers—lowers the potential barriers to wide use of the data even further.⁷⁵ It stands in many ways as a model for open sharing and collaboration around data, all in the interests of social change and improvement.

Next Steps for GeoNode

GeoNode's open nature results in “a very vibrant community with more and more people studying and contributing development.”⁷⁶ As a result, the organizers “can't really keep track” of where it's going. Indeed, Deparday, speaking about GeoNode in general, notes that, “This is, I think, a really great success story about open source.”⁷⁷ The GeoNode team, as of late 2015, is supporting around 20 active GeoNodes around the world. That number climbs to “a hundred or more if we are talking about people using it for their own use—people who have installed it for their university or their local governments.”⁷⁸

Perhaps the greatest impact of the Ebola GeoNode is still to come. Benson Wilder of HIU notes that, “There's a mind toward longer term management of the geospatial data and the ability to transfer the content and the governance to entities who will have an interest

⁷³ <http://www.nytimes.com/interactive/2014/07/31/world/africa/ebola-virus-outbreak-qa.html>

⁷⁴ GovLab interview with Javier Teran, Statistician, United Nations Office for the Coordination of Humanitarian Affairs, Humanitarian Data Exchange, September 18, 2015.

⁷⁵ GovLab interview with Javier Teran, Statistician, United Nations Office for the Coordination of Humanitarian Affairs, Humanitarian Data Exchange, September 18, 2015.

⁷⁶ GovLab interview with Vivien Deparday, Disaster Risk Management Specialist, Open Data for Resilience Initiative Operational Deployment Lead, Global Facility for Disaster Risk and Reduction, World Bank, December 17, 2015.

⁷⁷ GovLab interview with Vivien Deparday, Disaster Risk Management Specialist, Open Data for Resilience Initiative Operational Deployment Lead, Global Facility for Disaster Risk and Reduction, World Bank, December 17, 2015.

⁷⁸ GovLab interview with Vivien Deparday, Disaster Risk Management Specialist, Open Data for Resilience Initiative Operational Deployment Lead, Global Facility for Disaster Risk and Reduction, World Bank, December 17, 2015.

in using it and keeping it up to date.”⁷⁹ This focus on governance is key for future GeoNode efforts, particularly around open data preparedness since, as Deparday makes clear, “technology is the easy part.”⁸⁰

The Closure of NERC

In November 2015, WHO declared Sierra Leone officially Ebola-free after the passage of 42 days without any new cases reported.⁸¹ The next month, on December 31, 2015, NERC closed its doors. At that point, the Centre’s responsibilities were distributed among the Ministry of Health and Sanitation (MoHS), the Ministry of Social Welfare and the Office of National Security.⁸²

The closure of NERC was not welcomed by some in Sierra Leone. The Tanzanian newspaper *The Citizen* noted “skepticism” among the public as a result of “fear [that] the expertise within the institution can’t be replaced.”⁸³ The government, however, noted that the financial implications of maintaining NERC were too significant given the Centre’s preponderance of international staff.⁸⁴

Sadly, the body of a young woman tested positive for Ebola two months after the country was officially declared Ebola-free, and just

79 Quinn, Kristin. “An Unprecedented Response: The GEOINT Community’s Role in the Ebola Relief Effort.” *Trajectory Magazine*. 2015. <http://trajectorymagazine.com/government/item/1908-an-unprecedented-response.html>

80 GovLab interview with Vivien Deparday, Disaster Risk Management Specialist, Open Data for Resilience Initiative Operational Deployment Lead, Global Facility for Disaster Risk and Reduction, World Bank, December 17, 2015.

81 O’Carroll, Lisa and Umaru Fofana. “WHO officially declares Sierra Leone Ebola-free.” *The Guardian*. November 7, 2015. <http://www.theguardian.com/world/2015/nov/07/world-health-organisation-sierra-leone-ebola-free>

82 “Sierra Leone dissolves Ebola response centre.” *The Citizen*. January 6, 2016. <http://www.thecitizen.co.tz/News/Sierra-Leone-dissolves-Ebola-response-center/-/1840340/3022944/-/11h8g99z/-/index.html>

83 “Sierra Leone dissolves Ebola response centre.” *The Citizen*. January 6, 2016. <http://www.thecitizen.co.tz/News/Sierra-Leone-dissolves-Ebola-response-center/-/1840340/3022944/-/11h8g99z/-/index.html>

84 “Sierra Leone dissolves Ebola response centre.” *The Citizen*. January 6, 2016. <http://www.thecitizen.co.tz/News/Sierra-Leone-dissolves-Ebola-response-center/-/1840340/3022944/-/11h8g99z/-/index.html>

two weeks after NERC's closure.⁸⁵ It remains to be seen whether this case will become anything more than a tragic, isolated setback in the country's recovery from the crisis.

The data-driven response to the Ebola crisis in Sierra Leone in many ways took on the character of the crisis itself—a massive, chaotic task involving a broad range of well-meaning actors who were attempting to improve conditions while facing significant challenges. Although the direct impact of efforts like the initiatives described here can be difficult to quantify, it is clear that data played a key role in placing relevant information into the hands of those who needed it and in helping address the Ebola crisis. It is equally clear that these initiatives offer best practices and pathways to those at the forefront of disaster response around the world. In many ways, the experience in West Africa can be considered a valuable proof-of-concept for the positive role of data and information in a variety of crises, man-made or natural.

⁸⁵ Searcey, Dionne and Sheri Fink. "Day After a Victory Over Ebola, Sierra Leone Reports a Death." *The New York Times*. January 15, 2016.

CHAPTER 17

New Zealand's Christchurch Earthquake Clusters

Open Data for Improving Emergency Response

By Juliet McMurren, Stefaan Verhulst and Andrew Young

Summary: In February 2011, Christchurch was struck by a severe earthquake that killed 185 people and caused significant disruption and damage to large portions of a city already weakened by an earlier earthquake. In the response to the quake, volunteers and officials at the recovery agencies used open data, open source tools, trusted data sharing and crowdsourcing to develop a range of products and services required to respond successfully to emerging conditions. These included a crowdsourced emergency information Web app that generated 70,000 visits within the first 48 hours after the earthquake; a series of geographic information system (GIS) data sharing agreements between agencies that enabled the successful provision of mapping services throughout the response and recovery; websites using open property data that enabled citizens to check the status of their homes and land, and generated millions of hits within hours of release; a construction intention viewer built using open data and open source tools that saved NZ\$4 million in construction costs within its first year of use; and a crowdsourced competition for school children that generated over 18,000 new building footprints for open property databases at a cost of \$0.02 per footprint.

Dimension of Impact: Solving Public Problems—Data-Driven Engagement

Key Takeaways:

- Open maps and property data can be paired with trusted data sharing and open-source tools to craft quick, cost-effective solutions for crisis response.
- The capacity to recover and rebuild quickly can benefit tremendously from a good pre-existing data infrastructure, and especially on a robust and authoritative property data set.
- Crises can provide excellent preconditions for innovation, including the freedom and permission to innovate.
- Agile development and management techniques are particularly well suited to emergency response.
- Crowdsourcing emergency information can provide a way of engaging would-be volunteers safely in disaster relief efforts.

Context and Background

New Zealand is a high-income Pacific Island nation with a population of 4.6 million.¹ It was ranked seventh on the UN Human Development Index in 2013.² The country is located on the Pacific Ring of Fire, where the Australian and Pacific plates meet, and is seismically active. It experiences about 14,000 earthquakes every year, of which around 150 to 200 are strong enough to be felt.³ Most of these earthquakes occur down the Alpine fault, which runs down the center of the South Island, and along another fault that runs from southwest to northeast through the central North Island. In the last 200 years, New Zealand has experienced 12 major earthquakes resulting in loss of life.⁴

Because of this seismic history, New Zealand is an acknowledged world leader in earthquake engineering, having applied itself to

¹ http://www.stats.govt.nz/infoshare/default.aspx?RedirectReason=session_expired

² <http://hdr.undp.org/en/content/table-1-human-development-index-and-its-components>

³ <http://www.teara.govt.nz/en/earthquakes/page-1>

⁴ https://en.wikipedia.org/wiki/List_of_earthquakes_in_New_Zealand#Earthquakes_resulting_in_fatalities

learning the lessons from a series of deadly earthquakes in the late nineteenth and early twentieth centuries.⁵ The country has some of the most stringent building standards in the world, which set requirements for how buildings must perform in earthquakes.⁶ Current New Zealand building codes require structures with a 50-year use life to be able to withstand the predicted loads generated by earthquakes of a magnitude expected to occur every 500 years. New Zealand is also one of the only countries in the world to have national government earthquake insurance for homeowners through the Earthquake Commission (EQC).⁷

Open Data in New Zealand

New Zealand has an excellent record of press freedom and government transparency. The New Zealand press is considered free according to Freedom House,⁸ and the country is ranked sixth on the 2015 World Press Freedom Index by Reporters without Borders.⁹ New Zealand was ranked second on Transparency International's Corruption Perceptions Index in 2014, and first on the International Budget Partnership's two most recent Open Budget Surveys (2012 and 2015).¹⁰

New Zealand is ranked fourth on the Open Data Barometer,¹¹ and is considered a “high capacity” country, meaning that the country has existing open data policies, political backing and a general culture of data openness. It announced its intention to join the Open Government Partnership (OGP) in 2013, and, in October 2014, released its first action plan, completing the process of joining OGP.¹²

In August 2011, the Declaration on Open and Transparent Government was passed by the New Zealand government.¹³ Under that pol-

5 http://www.nzherald.co.nz/nz/news/article.cfm?c_id=1&objectid=10672097

6 <http://www.teara.govt.nz/en/earthquakes/page-4>

7 <http://www.teara.govt.nz/en/earthquakes/page-3>

8 <https://freedomhouse.org/country/new-zealand#.Vbo0eylViko>

9 <http://index.rsf.org/#!/index-details>

10 <http://survey.internationalbudget.org/#timeline>

11 <http://www.opendatabarometer.org/report/analysis/rankings.html>

12 <http://ssc.govt.nz/nz-ogg-action-plan>

13 <https://www.ict.govt.nz/guidance-and-resources/open-government/declaration-open-and-transparent-government/>

icy, central government agencies were directed, state-owned enterprises encouraged, and local government invited, to actively release high-value, non-personal data for reuse. During 2011, the New Zealand government also implemented New Zealand Government Open Access and Licensing Framework (NZGOAL),¹⁴ an open access and open licensing protocol for state agencies to use when releasing data for reuse. This protocol encourages the release of non-personal copyright data using the most open Creative Commons license, and the release of non-copyright data with no restrictions on its use, all in the interest of harnessing the economic and creative benefits of opening the data for reuse.¹⁵ As of December 2015, the New Zealand government data portal, data.govt.nz, offered a total of 3,813 data sets.¹⁶

Project Description and Inception

In 2010 and 2011, the city of Christchurch, New Zealand's third largest, with a population of 375,000, experienced a series of devastating earthquakes. On September 4, 2010, a magnitude 7.1 quake caused widespread property damage and minor injuries, but no deaths.¹⁷ Nearly six months later, on February 22, 2011, before the city had fully recovered from the first earthquake, it experienced a second severe quake. While weaker in magnitude than the first and a mere 12 seconds long, an unlucky combination of factors—shallow depth, steep angle, and an epicenter located a mere 10 kilometers from the city center¹⁸—meant that the second earthquake produced some of the most intense and violent shaking ever recorded in an urban area. Peak ground acceleration during the earth-

14 <https://www.ict.govt.nz/guidance-and-resources/open-government/new-zealand-government-open-access-and-licensing-nzgoal-framework/>

15 <http://www.ssc.govt.nz/nzgoal>

16 <https://data.govt.nz/>

17 <http://web.archive.org/web/20100905033623/http://earthquake.usgs.gov/earthquakes/recenteqswv/Quakes/us2010atbj.php>; “Why so few casualties in Canterbury quake?” Stuff.co.nz. September 4, 2010. <http://www.stuff.co.nz/national/4096813/Why-so-few-casualties-in-Canterbury-quake>

18 Office of the Prime Minister's Science Advisory Committee. “The Canterbury Earthquakes: Scientific answers to critical questions.” The Royal Society of New Zealand. 2011. <http://www.royalsociety.org.nz/media/Information-paper/The-Canterbury-Earthquakes.pdf>

quake approached 2g in parts of central Christchurch¹⁹ (as compared to 0.5g in the 2010 Haiti earthquake), and eyewitness accounts reported people literally being thrown into the air.²⁰

Julian Carver, former CIO of the Canterbury Earthquake Recovery Authority (CERA), describes the experience of the earthquake:

“February 22nd 2011, 12:51 p.m., I’m working from home, lying on my bed, reading email on my iPhone. Thirty seconds later, my city, my life, and my future had changed irrevocably. Anything not bolted down was on the floor and half of it was smashed. Computer monitors, TVs, bookshelves, food from the fridge. The power went off, then stayed off for five days. Mobile calls worked for a few minutes, then failed. Texts became patchy after an hour. The only thing that was semi-reliable was Twitter over 3G.”²¹

The quake caused significant structural damage to the city’s already weakened buildings. The country’s strict building codes and the mercifully short duration of the quake limited the damage,²² but 185 people were killed, half of them in a single building collapse, in what was New Zealand’s second-deadliest recorded natural disaster. As of April 2013, the cost of rebuilding stood at \$40 billion NZD.²³

The recovery was significantly aided by a number of projects making innovative use of open data, open source tools, crowdsourcing and trusted data sharing. These tools, which we describe below, were developed in a highly agile and iterative manner. They allowed the city to recover rapidly and cheaply; together, they suggest the tre-

19 Campbell, Hamish. “Technically it’s just an aftershock.” *New Zealand Herald*. February 24, 2011. http://www.nzherald.co.nz/opinion/news/article.cfm?c_id=466&objection_id=10708275

20 Lin II, Rong-Gong and Sam Allen. “New Zealand quake raises questions about L.A. buildings.” *Los Angeles Times*. February 26, 2011. <http://www.latimes.com/local/la-mequake-california-20110226-story.html>

21 Carver, Julian. “Using the Web in Earthquake Recovery.” New Zealand Government Web Toolkit. March 25, 2013. <https://webtoolkit.govt.nz/blog/2013/03/using-the-web-in-earthquake-recovery/>

22 Clifton, Charles. “Christchurch Feb 22nd Earthquake: A Personal Report by Charles Clifton—March 2011.” HERA. March 2011. http://www.hera.org.nz/Story?Action=View&Story_id=1398

23 GovLab interview with Julian Carver, former Chief Information Officer, Canterbury Earthquake Recovery Authority, August 3, 2015.

mendous potential of innovation and data-driven projects in the midst of a crisis and as responses to natural and other disasters.²⁴

Canterbury Recovery Map (Eq.org.nz)

Immediately after the 2011 earthquake, significant parts of the city were without water or sewerage for up to three weeks, since up to 80 percent of the city's below-ground infrastructure had been damaged.²⁵ Roads in some parts of the city were inaccessible because of damage or soil liquefaction, and normal channels of communication were significantly disrupted by power outages. One of the most immediate problems was a lack of information, as residents tried to work out where to go for essential goods and services. Official sources did exist: A cloud-hosted emergency information website (canterburyearthquake.govt.nz) had been set up immediately after the earthquake,²⁶ when it became apparent that the City Council servers were not up to the task of handling demand for emergency information because of power outages, building damage, and inadequate capacity. However, this website didn't have mapping capabilities, and depended on a small, overworked team—"four people sitting around a trestle table with laptops," according to Carver²⁷—physically located inside the emergency operations center in the heavily damaged city center.²⁸

Within hours of the quake, a group of skilled volunteers in New Zealand and overseas responded to these shortcomings by using Ushahidi, an open source disaster response platform²⁹ using open

²⁴ GovLab interview with Julian Carver, former Chief Information Officer, Canterbury Earthquake Recovery Authority, August 3, 2015.

²⁵ Clifton, Charles. "Christchurch Feb 22nd Earthquake: A Personal Report by Charles Clifton—March 2011." HERA. March 2011. http://www.hera.org.nz/Story?Action=View&Story_id=1398

²⁶ McNamara, Tim. "Eq.org.nz—The Power of Ushahidi." Institute of IT Professionals TechBlog. March 18, 2011. <http://techblog.nz/94-eqorgnzThePowerofUshahidi>

²⁷ GovLab interview with Julian Carver, former Chief Information Officer, Canterbury Earthquake Recovery Authority, August 3, 2015.

²⁸ McNamara, Tim. "Eq.org.nz—The Power of Ushahidi." Institute of IT Professionals TechBlog. March 18, 2011. <http://techblog.nz/94-eqorgnzThePowerofUshahidi>

²⁹ Carver, Julian. "Using the Web in Earthquake Recovery." New Zealand Government Web Toolkit. March 25, 2013. <https://webtoolkit.govt.nz/blog/2013/03/using-the-web-in-earthquake-recovery/>

map data that had been successfully deployed after the Haiti quake, to create Eq.org.nz, a crowdsourced disaster map.³⁰

The site helped residents navigate the post-earthquake urban environment by crowdsourcing information about damage, road closures, and the availability of essential supplies and services, and offers or requests for help.³¹ At the time, Tim McNamara, one of the Eq.org.nz leads, summed up the project: “We’re asking people to tell us where they are and what they see—if roads are blocked, which [store] is open, which [hardware store] is open, which medical center, where there are phones working and Internet access.”³² Contributors could enter information on a website form, or via email, SMS code, or Twitter with hashtags #eqnz or #helpme for emergency requests.³³ Human curators categorized every incoming message that contained both a fact and a location, and plotted it on a map. Eq.org.nz’s volunteers were able to verify, categorize and publish reports within five minutes of receipt of a new message. The data was published via an open Web application programming interface (API) that allowed third parties, including Environment Canterbury, to combine the site’s information with their own data.³⁴

By February 24, 2011, two days after the earthquake, the site had amassed 779 reports, 781 different locations, and almost 70,000 unique visitors.³⁵ It also helped inform the activities of local volunteers such as the Student Army and Farmy Army, which provided thousands of volunteers to help clear more than 360,000 tons of silt

30 Bell, Stephen. “Tech volunteers quick to help in quake aftermath.” *Computerworld*. February 24, 2011. http://www.computerworld.co.nz/article/498626/tech_volunteers_quick_help_quake_aftermath/

31 GovLab interview with Julian Carver, former Chief Information Officer, Canterbury Earthquake Recovery Authority, August 3, 2015.

32 Bell, Stephen. “Tech volunteers quick to help in quake aftermath.” *Computerworld*. February 24, 2011. http://www.computerworld.co.nz/article/498626/tech_volunteers_quick_help_quake_aftermath/

33 Bell, Stephen. “Tech volunteers quick to help in quake aftermath.” *Computerworld*. February 24, 2011. http://www.computerworld.co.nz/article/498626/tech_volunteers_quick_help_quake_aftermath/

34 McNamara, Tim. “Eq.org.nz—The Power of Ushahidi.” Institute of IT Professionals TechBlog. March 18, 2011. <http://techblog.nz/94-eqorgnzThePowerofUshahidi>

35 Meyer, Patrick. “Launching Eq.org.nz for the New Zealand Earthquake.” Ushahidi. February 24, 2011. <http://www.ushahidi.com/blog/2011/02/24/launching-eq-org-nz-for-the-new-zealand-earthquake/>

deposited by liquefaction from residential properties, over more than 80,000 volunteer working hours.³⁶ The Ushahidi volunteers reported feedback from users such as a Christchurch diabetic, who thanked them for telling her where she could get insulin.³⁷ The site was active for three weeks after the earthquake, until power and normal channels of communication were fully restored.

GIS and trusted data sharing

In the immediate aftermath of the second earthquake, officials responsible for GIS data at Christchurch City Council and Environment Canterbury found themselves overwhelmed with demands from Civil Defence and emergency services for maps to help with search and rescue and other emergency response.³⁸ Requests quickly outstripped their capacity, and the team knew that it needed to call in outside help. Ultimately, they found solutions in various data-sharing initiatives.

Volunteer teams from the Wellington City Council and Greater Wellington Regional Council, and Eagle Technology, a Wellington-based IT firm offering open systems and GIS platforms, had all offered help, but the traditional solution—mailing data on DVD or hard drive to the Wellington volunteers—was too slow. The Christchurch GIS data team obtained permission to open their data so that it could be freely used under a Creative Commons license, enabling the Wellington teams to produce maps to help with emergency response.³⁹ Static data, including aerial imagery of Christchurch taken within 48 hours of the quake, was uploaded to, and made

³⁶ “September 2010 Christchurch Earthquake.” Volunteer Army Foundation. <http://www.volunteerarmy.org/history/>

³⁷ Meyer, Patrick. “Launching Eq.org.nz for the New Zealand Earthquake.” Ushahidi. February 24, 2011. <http://www.ushahidi.com/blog/2011/02/24/launching-eq-org-nz-for-the-new-zealand-earthquake/>

³⁸ Carver, Julian. “Open_Data_Mini_Case_Studies.” Open New Zealand. February 6, 2013. https://wiki.open.org.nz/wiki/display/main/Open_Data_Mini_Case_Studies#Open_Data_Mini_Case_Studies-OpenDataSupportsEmergencyOperations

³⁹ Carver, Julian. “Open_Data_Mini_Case_Studies.” Open New Zealand. February 6, 2013. https://wiki.open.org.nz/wiki/display/main/Open_Data_Mini_Case_Studies#Open_Data_Mini_Case_Studies-OpenDataSupportsEmergencyOperations

freely available from the geodata distributor Koordinates.com,⁴⁰ while dynamic data was made available to the emergency mapping data teams via open geospatial standards.⁴¹

Similar data sharing occurred after the establishment of CERA. Having been set up on a relatively ad hoc basis six weeks after the second earthquake, CERA's office and IT solutions were basic, and had no mapping, GIS, or data functionality. Carver, the acting CIO, approached Land Information New Zealand (LINZ) for assistance with GIS data infrastructure: "They said, we've got a set of those services we could spin up for you, and then feed in all of the open data from LINZ data services, and that would get you started."⁴² Having established CERA's mapping capacity, Carver began to get requests for help from the Christchurch City Council, which was facing heavy demands generating maps for CERA demolition crews working in the central city. Christchurch City Council provided CERA with a list of all of its data sets that it could open up as data services, which was then prioritized by CERA. The data was then opened through open or secure data feeds, and the CERA GIS team was able to do analysis and make maps for the Christchurch City Council.⁴³

This began a pattern of GIS data sharing and opening between the earthquake recovery agencies. As Carver put it:

"It was a New Zealand-scale, Christchurch-scale [solution]. You could get the four or five people ... who understood the need, understood what users wanted, had the technical understanding and had the authority to make it happen ... in a room, every two weeks, [saying] OK, now we need to add this, or change this, or open this up ... It was very agile, very iterative."⁴⁴

40 <https://koordinates.com/>

41 Carver, Julian. "Open_Data_Mini_Case_Studies." Open New Zealand. February 6, 2013. https://wiki.open.org.nz/wiki/display/main/Open_Data_Mini_Case_Studies#Open_Data_Mini_Case_Studies-OpenDataSupportsEmergencyOperations

42 GovLab interview with Julian Carver, former Chief Information Officer, Canterbury Earthquake Recovery Authority, August 3, 2015.

43 GovLab interview with Julian Carver, former Chief Information Officer, Canterbury Earthquake Recovery Authority, August 3, 2015.

44 GovLab interview with Julian Carver, former Chief Information Officer, Canterbury Earthquake Recovery Authority, August 3, 2015.

Landcheck and MyProperty

By late June 2011, the Canterbury Earthquake Recovery Authority had completed its geotechnical assessment and zoning of residential property to indicate the level of risk for a given area in the event of an earthquake. Now it needed a way to communicate those zoning decisions to the people of Christchurch. As Carver recalled in a blog post: “Like everything in the recovery, time frames were tight. People want government decisions to be based on sound scientific and economic evidence. They also want to know where they stand (and can live), as soon as possible. CERA needed a way to let people see exactly which zone their house was in. That required an interactive website, capable of taking a massive initial load, which would be implemented in a very short time frame.”⁴⁵

The solution was a partnership between the engineering firm responsible for creating the earthquake zoning maps, and Trade Me, New Zealand’s largest online auction site. Carrying out the work pro bono, Trade Me built the Landcheck⁴⁶ site using open property and address data in four days, using their server farms in Auckland and Wellington. Carver reports that the site received 2 million page views in the first hour, and 5 million page views and 3.3 million individual property searches on the first day.⁴⁷

Three months later, in October 2011, Landcheck was replaced by MyProperty,⁴⁸ as CERA made public the results of citywide geotechnical studies. MyProperty allowed residents to check not only the zoning of their property, but also the technical category of the land, which defined how it was expected to perform in future earthquakes, and the foundation type required for new construction.⁴⁹ The technical category maps on MyProperty were built on the same GIS viewer and open data as had been used for Landcheck, incorpo-

45 Carver, Julian. “Using the Web in Earthquake Recovery.” New Zealand Government Web Toolkit. March 25, 2013. <https://webtoolkit.govt.nz/blog/2013/03/using-the-web-in-earthquake-recovery/>

46 <http://www.landcheck.org.nz/>

47 GovLab interview with Julian Carver, former Chief Information Officer, Canterbury Earthquake Recovery Authority, August 3, 2015.

48 <http://cera.govt.nz/my-property>

49 <http://cera.govt.nz/residential-green-zone-technical-categories/overview>

rating photographs, maps, zoning, and technical category data.⁵⁰ Although Carver notes it would be difficult to place a monetary value on the benefit provided by these sites, they were an essential public service that was widely accessed by the public, giving assurance to citizens about the safety of their property as aftershocks continued, and authoritative information about the legal context for repairs and rebuilds.⁵¹

Forward Works Viewer

During the repair and rebuild of the central city, which had been badly damaged by the February 2011 earthquake, CERA found itself tasked with the near-impossible: demolishing 1,200 commercial buildings, repairing all below-ground infrastructure (wastewater, stormwater, water supply, power, and broadband), and beginning the process of reconstructing new buildings, all within a small geographic area, at the same time. In an interview, Carver said that “the only way of viably doing that [was to enable] everyone to see everyone else’s forward construction intentions well enough in advance to avoid expensive clashes and delays.”⁵²

LINZ, CERA, and the other agencies coordinating the Canterbury rebuild responded with the Forward Works Viewer, a tool which gave those agencies and other public and private sector users a shared online view of horizontal infrastructure repair, planned buildings, and other construction.⁵³ The viewer allowed users to manage and view projects and their impacts spatially and over time, and detect potential clashes and opportunities for collaboration.⁵⁴

The development of Forward Works, which was carried out using the agile software development methodology Scrum, was subcontracted to companies with geospatial, engineering and Web develop-

⁵⁰ GovLab interview with Stephen Ferris, GIS and Data Manager, Canterbury Earthquake Recovery Authority, September 22, 2015.

⁵¹ GovLab interview with Julian Carver, former Chief Information Officer, Canterbury Earthquake Recovery Authority, August 3, 2015.

⁵² GovLab interview with Julian Carver, former Chief Information Officer, Canterbury Earthquake Recovery Authority, August 3, 2015.

⁵³ “Online tool enhances Canterbury rebuild.” Land Information New Zealand. July 29, 2014. <http://www.linz.govt.nz/news/2014-07/online-tool-enhances-canterbury-rebuild>

⁵⁴ “Forward Works Viewer.” Media Suite. <https://www.mediasuite.co.nz/forward-works.php>

ment expertise.⁵⁵ The site drew heavily on open property and road network data, open source tools, and open geospatial standards.⁵⁶ The open public road network data, however, had significant limitations, according to Carver.

"The road network data was simply road center-line ... and didn't tell you anything about lanes, directions, turns, and turn restrictions. We wanted to be able to build those things into Forward Works' viewer because we wanted to be able to assess impact on the road network of vertical construction or road closures due to road repair. Is it this lane, or this lane? Is it total closure or reduced capacity? That was quite important to know for traffic modelling, but we didn't have an open, freely reusable, routable roading network."⁵⁷

The solution was to contract four postgraduate GIS students for two weeks to bring OpenStreetMap for the relevant area fully up to date, at a cost of NZ\$10,000, and make it fully routable for Christchurch. They then integrated this data into the Forward Works viewer with an impact selector to enable planners to choose the best lane for closures.⁵⁸

The total cost of constructing the Forward Works viewer was NZ \$1.6 million. A LINZ assessment in 2014 showed that the Forward Works viewer had delivered NZ \$4 million in benefits since its launch in 2013, with a total of more than NZ\$20 million forecast.⁵⁹ These benefits were the result of cost savings in reduced clashes and delays, shared roadworks and trenching, reduced impact on the travelling public and traffic modelling that allowed twice as many closures within the central city while maintaining the same traffic flows.⁶⁰

55 "Forward Works Viewer." Media Suite. <https://www.mediasuite.co.nz/forward-works.php>

56 GovLab interview with Julian Carver, former Chief Information Officer, Canterbury Earthquake Recovery Authority, August 3, 2015.

57 GovLab interview with Julian Carver, former Chief Information Officer, Canterbury Earthquake Recovery Authority, August 3, 2015.

58 GovLab interview with Julian Carver, former Chief Information Officer, Canterbury Earthquake Recovery Authority, August 3, 2015.

59 "Online tool enhances Canterbury rebuild." Land Information New Zealand. July 29, 2014. <http://www.linz.govt.nz/news/2014-07/online-tool-enhances-canterbury-rebuild>

60 GovLab interview with Julian Carver, former Chief Information Officer, Canterbury Earthquake Recovery Authority, August 3, 2015.

Building Our Footprints

During the recovery, the agencies involved in rebuilding had identified deficits in some of the geospatial and property data sets. Residential building footprint databases did not exist for the satellite municipalities of Selwyn and Waimakariri, and the Christchurch data set was incomplete, with potential consequences for emergency response and rebuilding. In 2014, Jeremy Severinson, a LINZ employee who had been conducting postgraduate research assessing the trustworthiness of crowdsourced data at the University of Canterbury, approached Environment Canterbury with an idea.⁶¹

Severinson proposed crowdsourcing the creation of a database, in the form of a competition for school students. “Building Our Footprints” was run by Environment Canterbury in collaboration with LINZ and the University of Canterbury. Environment Canterbury created a Web app⁶² with instructions, registration, and login for participants, who digitized building outlines from open aerial photographs, attempting to achieve a trust metric above 75 percent. The first participant to achieve 75 percent or better was awarded the point for that building, and the participant with the greatest number of points won. LINZ provided a small amount of sponsorship for prizes, in the form of an iPad Mini for the eventual winner, cash, and movie tickets.⁶³ The competition ran for a month and generated 18,789 building outlines,⁶⁴ which were integrated into the relevant council databases and OpenStreetMap.⁶⁵

Carver admits that this competition was done “for fun, because we wanted to see what would work in terms of solving problems. ... Just as importantly, we got a bunch of kids, who might not have considered spatial or open data or technology ... in their careers, engaged with that. So it was ... just a little ‘Let’s see what happens!’ exemplar

⁶¹ GovLab interview with Iain Campion, former Application Team Leader, Environment Canterbury, July 28, 2015.

⁶² <http://www.canterburymaps.govt.nz/buildingourfootprints/>

⁶³ <http://www.canterburymaps.govt.nz/buildingourfootprints/Prizes>

⁶⁴ <http://www.canterburymaps.govt.nz/buildingourfootprints/Results>

⁶⁵ GovLab interview with Julian Carver, former Chief Information Officer, Canterbury Earthquake Recovery Authority, August 3, 2015.

—and it worked really well.”⁶⁶ Iain Campion is also enthusiastic about the competition: “We were quite keen on it, not just for the building footprints, but to give us an insight on how we could use ... the crowd for our other data sets, like water quality.”⁶⁷

Campion concedes, however, that not all of the objectives of the competition were met. Map areas were not randomly assigned, but were chosen by the participants, with the result that most chose the area in which they lived. The majority of participants came from a handful of schools within Christchurch itself, so the competition generated fewer building outlines from outlying towns, and duplicated some outlines already held by Christchurch City Council. The overall quality of the data, however, was good, and the cost per outline was a mere NZ\$0.02.⁶⁸

Impact

Those who lived and worked through the Christchurch earthquake response came away impressed by the potential of open data, open source tools, and data sharing to improve the efficiency and effectiveness of disaster response and recovery. Beyond the practical and financial impacts of the individual tools described above, there were several broader ways in which this impact was evident. Each holds valuable insights about the ways in which open data and open source tools can contribute to disaster response.

“Don’t be afraid to share [data]. If you can anonymize the data, share it, and people will make use of it.”

—Iain Campion, former Application Team Leader, Environment Canterbury

Higher-Quality Geospatial Data

Improvement of open data coverage and quality was an indirect benefit of at least two of the projects described above. The additional

⁶⁶ GovLab interview with Julian Carver, former Chief Information Officer, Canterbury Earthquake Recovery Authority, August 3, 2015.

⁶⁷ GovLab interview with Iain Campion, former Application Team Leader, Environment Canterbury, July 28, 2015.

⁶⁸ GovLab interview with Iain Campion, former Application Team Leader, Environment Canterbury, July 28, 2015.

data captured during the development of Forward Works and through Building Our Footprints greatly increased the accuracy and granularity of OpenStreetMap data for Christchurch, which was then made available for subsequent reuse by other users. Although intended as a simple solution to an emergent need, the distribution of GIS data through Koordinates to facilitate distributed mapping also opened new, previously unavailable geographic data sets.

Facilitating Collaboration

As with other examples in this report, a large part of the value of the data tools used in New Zealand came from the way in which they facilitated collaborative efforts and teamwork. In part, this was the result of an ability to work asynchronously and across geographies. McNamara summed up many of these benefits in a blog post from March 2011: “The open source model was critical for the success of [the Canterbury Recovery Map],” he wrote. “The open source community shares ... practices and norms for effective remote communication and collaboration. This meant that it was simple to manage a software project that had developers working in multiple time zones in a very constrained timeframe. [A large] part of the success was due to the ability for multiple people and organizations [to collaborate].”⁶⁹

Collaboration and teamwork were also facilitated by the neutrality of open data products and tools. McNamara points out that using open source tools allowed the Map to be vendor-neutral and ad-free, which made it easier for competitors to collaborate on the site. That neutrality also encouraged businesses to contribute data directly to the site, which shifted the burden of accuracy to those with the greatest interest in it: the businesses themselves.⁷⁰

Impact on Other Data Projects

Both Carver and Campion feel that the earthquake experience has advanced the cause of open data and accelerated data release in New Zealand. Carver notes that the earthquake provided a rapid and dra-

⁶⁹ McNamara, Tim. “Eq.org.nz—The Power of Ushahidi.” Institute of IT Professionals TechBlog. March 18, 2011. <http://techblog.nz/94-eqorgnzThePowerofUshahidi>

⁷⁰ McNamara, Tim. “Eq.org.nz—The Power of Ushahidi.” Institute of IT Professionals TechBlog. March 18, 2011. <http://techblog.nz/94-eqorgnzThePowerofUshahidi>

matic conversion to open data for the agencies involved in the recovery.

“Those … organizations, that previously didn’t have much in the way of policy or practice around open data or social media, went from ‘Oh, no, we couldn’t possibly do that!’ to it being the only thing that would work over the span of a week—and then just never looked back. It wasn’t that it would be a good idea and incrementally value-adding, but then you had to persuade a bunch of naysayers. It was the only thing that would work.”⁷¹

Campion agrees that the Christchurch experience has changed attitudes toward open data. “I think it has opened up all the agencies involved, and some of the peripheral ones. There’s not so much pushback any more. They know that that’s what should be happening.”⁷²

Carver also believes that the Christchurch experience has probably accelerated the pace of data release by local governments in New Zealand, by demonstrating both the benefits and the manageability of associated risk.⁷³ He observes that resistance to opening data has a similar pattern worldwide. “The fears … which are barriers to open data release, and therefore to realizing the value from open data release … are often just a slightly more explicit articulation of … ‘We can’t possibly do this because a bad thing will explode.’ [They] are almost always not especially evidence-based, and the risks are able quite easily to be managed. It’s primarily a change management process, not a risk management process.”⁷⁴ And “Crises give you an opportunity to go through that process of getting through those objections faster.”⁷⁵ Adds Campion, “Don’t be afraid to share [data].

71 GovLab interview with Julian Carver, former Chief Information Officer, Canterbury Earthquake Recovery Authority, August 3, 2015.

72 GovLab interview with Iain Campion, former Application Team Leader, Environment Canterbury, July 28, 2015.

73 GovLab interview with Julian Carver, former Chief Information Officer, Canterbury Earthquake Recovery Authority, August 3, 2015.

74 GovLab interview with Julian Carver, former Chief Information Officer, Canterbury Earthquake Recovery Authority, August 3, 2015.

75 GovLab interview with Julian Carver, former Chief Information Officer, Canterbury Earthquake Recovery Authority, August 3, 2015.

If you can anonymize the data, share it, and people will make use of it.”⁷⁶

- **Canterbury Recovery Map (Eq.org.nz):** In the two days after the earthquake, amassed 779 reports, 781 different locations, and almost 70,000 unique visitors; informed the activities of local volunteers who helped clear more than 360,000 tons of silt deposited by liquefaction from residential properties, over more than 80,000 volunteer working hours.
- **CERA:** Created maps for Christchurch City Council and performance analysis on those maps to inform the activities of demolition crews, among other uses.
- **Landcheck and MyProperty:** Provided citizens with information on safety of their property as aftershocks continued, and authoritative information about the legal context for repairs and rebuilds.
- **Forward Works Viewer:** In its first year, the Forward Works viewer delivered NZ\$4 million in benefits with a total of more than NZ\$20 million forecast, as a result of reduced clashes and delays, shared roadworks and trenching, reduced impact on the travelling public and improved traffic modelling.
- **Building Our Footprints:** The month-long competition generated 18,789 building outlines, which were integrated into the relevant council databases and OpenStreetMap.

Challenges

Data Infrastructure Challenges

Christchurch recovered remarkably fast from its earthquake, and many reasons have been given for its resilience: the existence of comprehensive insurance cover, including compulsory earthquake insurance; the geography of the city, which did not experience any single points of failure that could have devastating broader impacts if affected by the quake; and a rapid injection from the New Zealand government of NZ\$15 billion of the estimated NZ\$40 billion needed

⁷⁶ GovLab interview with Iain Campion, former Application Team Leader, Environment Canterbury, July 28, 2015.

to rebuild.⁷⁷ A further reason offered for the city's quick recovery was the existence of a robust data infrastructure that made it easy to resettle people in safe areas and enable the business community to continue to function.

Although good, the data infrastructure was not perfect, and those involved in the recovery effort have pointed out several flaws that could be remedied to improve the response to future disasters. In particular, both Carver and Campion point to the absence of a single, comprehensive and authoritative property data set. Carver notes that different pieces of property-related data—land parcels, records related to building footprints, addresses data and ratings—existed in “separate systems, across different organizations, and were often duplicated or differently updated and inaccurate.”⁷⁸

Lack of a Property Framework

Campion points to a related problem: the lack of a “good property framework” that would allow responders to confidently and comprehensively track, check and record response locations. “Relying on addresses that are not a standard or unique data set ... was really problematic,” he said. “For instance, [search and rescue teams] would go to a building in the CBD and they would say, yes, I’ve checked 34 Chester St West—but there is no actual address of 34 Chester St West, although there is a building there. There may be two Smith Streets in Christchurch: Well, which one did you check? I don’t know!”⁷⁹ A single comprehensive property data set might have reduced the cost of claims administration, allowing more money to go toward the rebuild. Such a standardized, public address data set exists in Denmark, resulting in improved public services, including emergency response.⁸⁰

⁷⁷ GovLab interview with Julian Carver, former Chief Information Officer, Canterbury Earthquake Recovery Authority, August 3, 2015.

⁷⁸ GovLab interview with Julian Carver, former Chief Information Officer, Canterbury Earthquake Recovery Authority, August 3, 2015.

⁷⁹ GovLab interview with Iain Campion, former Application Team Leader, Environment Canterbury, July 28, 2015.

⁸⁰ McMurren, Juliet, Stefaan Verhulst and Andrew Young. “Denmark’s Open Address Data Set: Consolidating and Freeing Up Address Data.” *Open Data’s Impact*. January 2016. <http://odimpact.org/case-denmarks-open-address-data-set.html>

Collaborations and Sharing Challenges

Finally, although the recovery effort was in many ways a testament to strong teamwork, there were some problems related to collaboration. In particular, Carver reports that the team behind Eq.org.nz developed certain tensions with some parts of the official emergency response teams. He says:

“The people doing the Web and social media in the emergency operations center were much more communicative and friendly with the crisis mapping teams [of Eq.org.nz] than the ... civil defense and emergency management folks, who got quite concerned that there was all this information being published and it wasn’t authoritative. [The] citizen response was, well, you don’t have any maps or authoritative information, so even if this isn’t perfect, it’s a lot better than nothing! That interaction and conflict meant fairly substantial post-earthquake upgrading of the emergency services’ understanding of how to engage with online communities during a natural disaster.”⁸¹

Carver contrasts the Canterbury example with the more successful response of the Queensland police to the Brisbane floods in 2010 and 2011, who made extensive and successful use of social media.⁸²

Looking Forward

Advancing Data-Driven Innovation in New Zealand

The experience of the Christchurch earthquake, which demonstrated that crises demanded innovative, highly cost-effective and rapid solutions, has pushed New Zealand to explore new uses and sources for data. Having embraced open data, the New Zealand government is now looking beyond it, to the possibilities of data-driven innovation. In August 2015, Statistics NZ, in partnership with the New Zealand Treasury and a group of expert stakeholders, announced the Data Futures Partnership,⁸³ a cross-sector collaboration of influential individuals working to drive change in NZ’s data

⁸¹ GovLab interview with Julian Carver, former Chief Information Officer, Canterbury Earthquake Recovery Authority, August 3, 2015.

⁸² GovLab interview with Julian Carver, former Chief Information Officer, Canterbury Earthquake Recovery Authority, August 3, 2015.

⁸³ http://m.stats.govt.nz/about_us/what-we-do/our-publications/cabinet-papers/data-futures-partnership-cabinet-paper

ecosystem. The partnership is intended to develop catalyst data-use projects and encourage increased trusted data sharing use of data that cannot be opened for reasons of privacy or commercial sensitivity between government agencies, and potentially between public and private sectors, to help promote data-driven innovation.⁸⁴

Carver believes the expansion from open data to data-driven innovation and data sharing was accelerated by the Christchurch experience. “If anything, this has happened faster because we’ve had lots of successes with open data in public, because we’ve had this earthquake context in which we had permission to innovate.”⁸⁵ Carver sees significant opportunity for innovation in crises. “The necessary preconditions for innovation are starvation, pressure, and perspective shift,” he says, quoting complexity theorist and management expert Dave Snowden. “Under significant time pressure, with fewer than normal resources, in a situation that’s really important to get right, [you’re] much more likely to come up with innovative solutions than ... when everything is comfortable.”⁸⁶

⁸⁴ GovLab interview with Julian Carver, former Chief Information Officer, Canterbury Earthquake Recovery Authority, August 3, 2015.

⁸⁵ GovLab interview with Julian Carver, former Chief Information Officer, Canterbury Earthquake Recovery Authority, August 3, 2015.

⁸⁶ Carver, Julian. “Using the Web in Earthquake Recovery.” New Zealand Government Web Toolkit. March 25, 2013. <https://webtoolkit.govt.nz/blog/2013/03/using-the-web-in-earthquake-recovery/>

CHAPTER 18

Singapore's Dengue Cluster Map

Open Data for Public Health

*By Andrew Young, David Sangokoya and
Stefaan Verhulst*

Summary: Faced with a dengue fever epidemic, the government of Singapore and the city's National Environmental Agency (NEA) turned to open data to control the spread of the disease. In 2005, the NEA started sharing information on the location of dengue clusters as well as disease information and preventive measures online, through a website now commonly known as the "Dengue Website." Since then, the NEA's data-driven cluster map has evolved, and it became an integral part of the campaign against a dengue epidemic in 2013. The campaign had two key components: an awareness-raising campaign, and a system of alerting the community to the severity of the dengue situation and the corresponding preventive measures to take. The data was also opened to app developers, who have used it to create a more nuanced and rich picture of the spread of dengue fever.

Dimension of Impact: Solving Public Problems—Data-Driven Engagement

Key Takeaways:

- In many cases, open data initiatives can become more impactful when they integrate an element of citizen engagement—in the interest of conducting user-centered design, filling gaps in existing data sets or driving use of the data and/or platform.

- Open data projects that are problem-focused have great potential to expand across borders to other areas facing similar challenges.
- Particularly in problem-focused efforts, there is a need for ongoing iteration around open data projects to address shifting conditions on the ground or better target the needs of the affected user community.

Context and Background

Dengue Fever

Dengue fever is a viral disease that causes symptoms such as high fever, severe headaches with pain behind the eyes, joint and muscle pain, skin rashes, nausea, vomiting, bruising, and bleeding from the nose or gums. Humans are the natural reservoirs of the dengue virus, of which there are four serotypes, and the disease is spread by the female *Aedes* mosquitoes, who pick up the virus from infected humans and then further transmit the virus to other humans through its bites.¹ The severe form of the disease—dengue haemorrhagic fever—is potentially fatal, particularly for older children and adults with weakened immune systems. This variation of dengue damages the lymph and blood vessels, and can lead to an enlargement of the liver.²

Over 2.5 billion people living in tropical areas are at risk for dengue, and the disease is particularly widespread in the Asia Pacific region, where some 70 percent of the population is considered at risk.³ However, the Centers for Disease Control and Prevention (CDC) found that the reported incidence of dengue is growing across the world, in developed as well as developing countries. The CDC posits that this increase is likely due to a combination of a number of factors, such as population growth, urbanization, lack of sanitation, increased long-distance travel, ineffective mosquito control, and cli-

¹ “Dengue.” World Health Organization. <http://www.who.int/topics/dengue/en/>.

² Associated Press. “Singapore on alert after dengue fever cases rise.” *The Independent*. June 11, 2013. <http://www.independent.co.uk/life-style/health-and-families/health-news/singapore-on-alert-after-dengue-fever-cases-rise-8654478.html>.

³ “Dengue.” World Health Organization. <http://www.who.int/topics/dengue/en/>.

mate change. Improvements in reporting capacity could also play a role in the apparently increasing numbers.⁴

Dengue's four different serotypes⁵ create an added challenge in efforts to address the spread of the disease. Once a patient suffers from dengue, they develop lifelong immunity to that particular serotype of the disease. However, immunity to the other serotype variations is only short-lived, and the risk of developing a more severe version of the disease increases with each subsequent infection by other serotypes.⁶

There is currently no vaccine or drug to combat dengue, and the principal form of treatment involves intravenous rehydration. The World Health Organization (WHO) reports that about 500,000 people with the severe form of dengue are hospitalized each year, with around 2.5 percent of those people dying. Highlighting the importance of initiatives like the one described here, WHO believes that early detection and medical attention could bring dengue's fatality rate below 1 percent.⁷ Individuals armed with an understanding of their level of risk for contracting dengue fever, it would seem, would be far more likely to seek early medical attention—the key factor in improving their chance of survival.

Dengue Fever in Singapore

Dengue has posed a public health challenge to the residents of Singapore since the mid-1960s, when it supplanted malaria as the number one mosquito-borne disease affecting the country. Recent years have witnessed two of the country's worst epidemics. In 2005, authorities in Singapore confirmed 14,209 cases of dengue, with 25

4 "Dengue and Climate." Centers for Disease Control and Prevention. <http://www.cdc.gov/Dengue/entomologyEcology/climate.html>.

5 A serotype is a "group of intimately related microorganisms distinguished by a common set of antigens." Merriam-Webster Medical Dictionary. <http://www.merriam-webster.com/medical/serotype>.

6 "Host Response to the Dengue Virus." Scitable by Nature Education. <http://www.nature.com/scitable/topicpage/host-response-to-the-dengue-virus-22402106>.

7 Wong, Chun Han. "Dengue Fever on the Rise in Singapore." *The Wall Street Journal*. April 29, 2013. <http://www.wsj.com/articles/SB10001424127887324743704578446162122159222>.

deaths.⁸ In 2013, Singapore saw a second major epidemic, with cases rising by more than 50 percent over the 2005 outbreak⁹ to 22,170.¹⁰ These increases are most likely being driven by a combination of greater population density,¹¹ increased international travel, and urbanization. A lower level of immunity throughout the population—the result of long periods of relatively low dengue incidence—has further elevated the risk of transmission. The availability of better and faster diagnostic kits also likely helped to increase the number of reported cases.¹²

Over time, the Singaporean government has developed a dynamic, evolving dengue control program that seeks to respond to changes in the disease's epidemiology and to conditions on the ground.¹³ Authorities have taken an evidence-based integrated approach, coupling virus, mosquito and case surveillance with source reduction—through, among other efforts, the removal of potential mosquito breeding grounds (in the form of standing water in, for instance, flower pot plates and pails) by residents; increasing insecticide fumigation; and penalizing people with a \$200 fine if their homes are found to be non-reported mosquito breeding zones within dengue cluster and transmission areas.¹⁴

Awareness-building has also played a key role. In April 2013, the National Environment Agency (NEA) launched the “Do the Mozzie Wipeout” campaign, focused on improving public awareness of den-

8 Wong, Chun Han. “Dengue Fever on the Rise in Singapore.” *The Wall Street Journal*. April 29, 2013. <http://www.wsj.com/articles/SB10001424127887324743704578446162122159222>.

9 GovLab Interview with the National Environment Agency of Singapore, July 23, 2015.

10 “Fact Sheet on Managing Singapore’s Dengue Outbreak.” National Environment Agency of Singapore. March 11, 2014. <http://www.nea.gov.sg/docs/default-source/corporate/COS-2014/managing-singapore's-dengue-outbreak.pdf>.

11 According to the World Bank, Singapore had a population density of 7,814 people per square kilometer in 2014, making it the third-densest country in the world. Source: “Population density (people per sq. km of land area).” World Bank Data. http://data.worldbank.org/indicator/EN.POP.DNST?order=wbapi_data_value_2014+wbapi_data_value+wbapi_data_value-last&sort=desc.

12 GovLab Interview with the National Environment Agency of Singapore, January 22, 2016.

13 GovLab Interview with the National Environment Agency of Singapore, July 23, 2015.

14 Tiong, Sabrina. “NEA officers fanning out in dengue hot spots.” *AsiaOne*. March 31, 2013.

gue, inspiring citizens to take preventive actions, and encouraging advocacy over social media and in person.¹⁵ The campaign included a variety of innovative communications and outreach components, including home-visits by volunteers to raise awareness; road shows aimed at educating the public; a campaign website and a regularly updated Facebook page.¹⁶

Open Data in Singapore

Singapore has an increasingly tech-driven and innovative public and government culture, perhaps in opposition to concerns over civil liberties issues in the nation state—Freedom House labels Singapore as “partly free” with a rating of 4 in civil liberties and political rights.¹⁷ In 2009, for instance, the prime minister’s office established the Centre for Strategic Futures, “a think tank within government, with the freedom to focus on issues of strategic importance even if they are not perceived as immediately urgent.”¹⁸ The think tank’s initiatives and research have focused on enabling broadly focused dialogue between citizens and the government,¹⁹ and arming government employees with the skills and competencies needed to lead in the 21st century.²⁰

As of 2014, the nation’s Internet penetration rate was 78.5 percent according to Internet World Statistics.²¹ It is considered a “high capacity” country on the Open Data Barometer, albeit at the very bottom of that cluster, with limitations in the areas of political, social and economic impacts, data accessibility, accountability, and citizen and civil society readiness.²² Part of the governmental effec-

15 GovLab Interview with the National Environment Agency of Singapore, July 23, 2015.

16 <https://www.facebook.com/Stop.Dengue.Now?ref=ts>

17 <https://freedomhouse.org/report/freedom-world/2013/singapore>.

18 Centre for Strategic Futures. “Foresight.” 2014. <http://www.csf.gov.sg/docs/default-source/default-document-library/csf-report-2014.pdf>.

19 Kuah, Adrian W.J. and Lim Seok Hui. “After Our Singapore Conversation: The Futures of Governance.” Singapore Civil Service College. June 2014. <https://www.cscollege.gov.sg/Knowledge/ethos/Ethos%20-%20Issue%202013,%20June%202014/Pages/After%20Our%20Singapore%20Conversation%20The%20Futures%20of%20Governance.aspx>.

20 Wei, Leong Ming. “Developing Talents in Futures.” in *Foresight*. 2014. <http://www.csf.gov.sg/docs/default-source/default-document-library/csf-report-2014.pdf>.

21 <http://www.internetworldstats.com/asia.htm#sg>.

22 <http://www.opendatabarometer.org/report/summary/>.

tiveness for which Singapore is known, is likely due in part to its relatively small size and keen focus on government-driven efficiency across sectors.

As of late 2015, the government's open data portal (data.gov.sg/) houses data sets from 71 government ministries and agencies, including over 350 data sets related to health originating from the Health Promotion Board, Health Sciences Authority and Ministry of Health. The portal also features a diversity of applications, including many focused on improving public health, such as apps focused on improving citizens' nutritional intake (Mealplan), sharing government-sponsored sport and fitness activities (Health Friend), and encouraging citizens to stop smoking (Smoke Free @ SG).

Singapore has also used the data portal as a means for catalyzing citizen engagement around available data, through, for example, the Data in the City Visualization Challenge, which asked citizens to tell "your Singapore story using data."²³ The winning entrant, The Singapore Story as told by the Straits Times, visualizes a 170-year-old Singaporean newspaper archive to capture the mood of the era in which the news stories appeared.²⁴

As Professor Ee-Peng Lim, co-director of Living Analytics Research Centre, Singapore Management University, describes, "Singapore is heading toward an effort to use both public and private enterprise data to innovate her economy as well as her society. This is currently the mindset of the government, as well as the nation."²⁵

23 <https://ideas.ecitizen.gov.sg/a/pages/visualisationchallenge-home>

24 <http://singaporennews.github.io/>

25 GovLab interview with Professor Ee-Peng Lim, Ph.D., Co-Director, Living Analytics Research Centre, Singapore Management University, September 28, 2015.

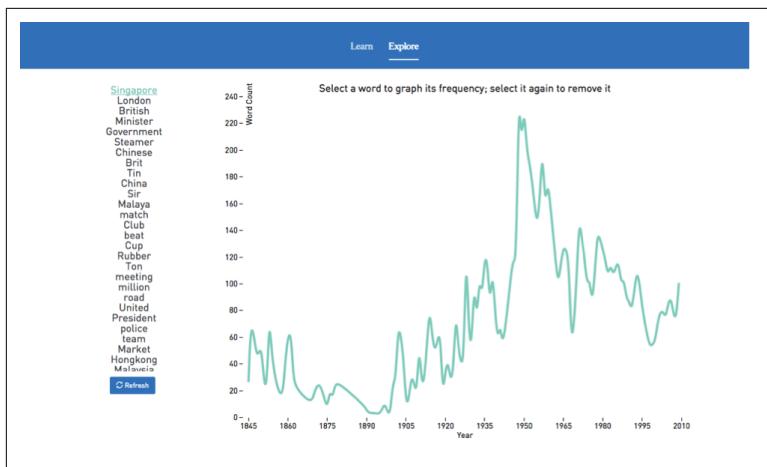


Figure 18-1. The Singapore Story, winner of the Singaporean Data Visualization Challenge

Project Description and Inception

Among Singapore's various efforts to combat dengue, the cluster map stands out for its combination of technology with on-the-ground public efforts. The cluster map grew originally out of the 2005 dengue epidemic. In September of that year, the weekly incidence of dengue climbed to over 700, a record high for Singapore.²⁶ As a result, the government began to take more proactive measures to provide information to the public, notably through updates on dengue cases and dengue clusters that were uploaded on the NEA and the Ministry of Health (MOH) websites every day.²⁷ This eventually developed into a website (the "Dengue Website," as it came to be known, dengue.gov.sg). That website brought together maps and tables from the NEA and MOH sites (along with other information) to build cluster maps that help authorities and the public geographically visualize locations where dengue is a problem. Among other things, these maps help plan vector control operations, inform the

²⁶ GovLab Interview with the National Environment Agency of Singapore, July 23, 2015.

²⁷ The NEA defines a dengue cluster as "a locality with active transmission where intervention is targeted. It is formed when two or more cases have onset within 14 days and are located within 150m of each other (based on residential and workplace addresses as well as movement history)." GovLab Interview with the National Environment Agency of Singapore, July 23, 2015.

public where vector control operations will be carried out, and enable the public in affected areas to take necessary precautions to protect themselves from mosquito bites and help to remove or prevent mosquito breeding areas.

The data used to build the cluster map is gathered from a variety of sources, including medical practitioners who attend or treat dengue cases²⁸ and laboratories that diagnose the disease. Because much of the data is gathered at the individual level, it is anonymized before being uploaded to the site (see note on privacy concerns below). The data is fed into a network that connects the NEA's headquarters with five satellite offices, as well as the Environmental Health Institute, an NEA-staffed laboratory where investigators conduct research on dengue and GIS modeling as a tool to combat dengue. The dengue cluster map is probably the first instance where the NEA made available to the public on a continuous basis such comprehensive information that was previously only used for internal operational planning²⁹

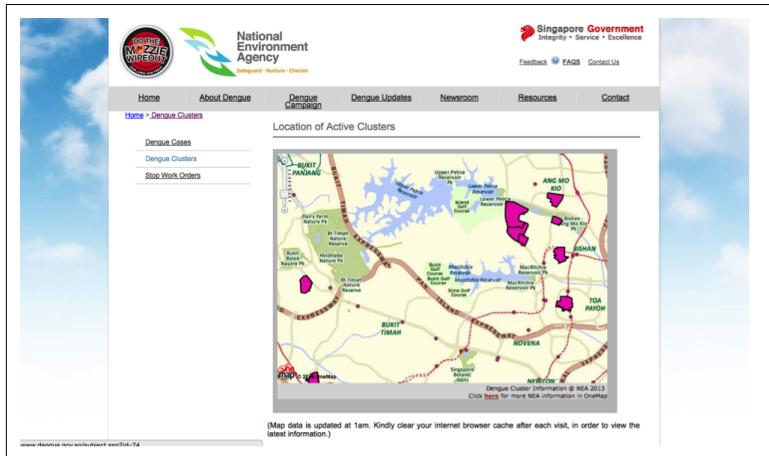


Figure 18-2. Map of Active Dengue Clusters on the “Dengue Website”

28 Under the Infectious Diseases Act, Cap. 137, every medical practitioner is required to notify the Director of Medical Services (Ministry of Health) not later than 24 hours from the time of diagnosis of dengue fever/dengue haemorrhagic fever.

29 GovLab Interview with the National Environment Agency of Singapore, August 31, 2015.

Over the years, the cluster map has changed—and improved—considerably. For example, soon after the website’s launch, it became apparent that a “point map”—a simple map with dots representing affected areas—did not provide a useful level of information to the public. In particular, it provided an inadequate indication of the exact boundaries of dengue clusters, and left some citizens confused about whether or not they resided within a cluster area. As a result, the point map was replaced with a map showing the boundary of each cluster. To complement this, the key breeding habitats detected were also published alongside the dengue clusters.³⁰

In addition, in 2013, following the epidemic mentioned above, a color-coded system was introduced to indicate the severity of an epidemic in a particular cluster. The map was also enhanced to include more granular data, including specific blocks in public housing and road names for private houses. All these enhancements have helped to give the public more specific, actionable information on how to protect themselves from mosquito bites in dengue clusters, and where a high level of vigilance is required in order to combat mosquito breeding.

Community outreach plays a key role in improving the cluster map. Direct outreach to citizens helps the NEA team to identify functionality gaps for the map, ensuring that the tools on offer match the needs of the user base. Perhaps more importantly, direct citizen communication continues to play a large role in supplementing the data housed in the cluster map. Upon receiving dengue case information, the NEA epidemiological officers authorized by the Ministry of Health interview cases to verify case details and learn more about citizens’ usual routines and recent travel history. This information is key for identifying where the mosquito bite leading to infection is likely to have occurred.

Simultaneously, a second group of officers use a predefined algorithm to incorporate the new cases and information obtained from epidemiological investigation into the cluster map and update existing cluster boundary information for daily publishing on the Dengue Website.

³⁰ GovLab Interview with the National Environment Agency of Singapore, July 23, 2015.

Impact

Since its inception, the Dengue Website has seen significant, sustained use, particularly during the peak dengue season. As indicated by the graph below, the site typically sees between 25,000 and 45,000 monthly visits, with that number rising to almost 100,000 during peak season.

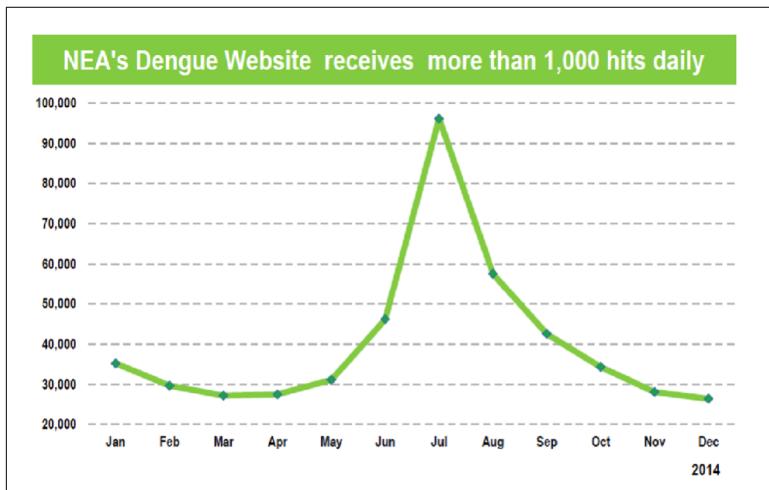


Figure 18-3. Traffic to the Dengue Website in 2014. Copyright Singapore NEA. Used with permission.

It would be a mistake, however, to assess the impact of the site and its maps solely through direct Web traffic. Singapore's efforts to combat dengue with open data have changed behaviors and attitudes across a variety of different social and demographic groups. It is an important initiative to make information more readily available to the public during this period to urge the public to play a part in dengue prevention and to help curb the dengue transmission.³¹ The reach of the cluster map has been extended by making it available through NEA's myENV app for smartphones on various operating platforms. Beyond just a passive aggregator of information, the app has opened up new functionalities for the data behind the cluster map, such as automated user alerts based on user-defined set-

³¹ GovLab Interview with the National Environment Agency of Singapore, August 31, 2015.

tings and location-based alerts. We consider three impact groups in particular: average citizens, media and developers.

Average Citizens

The largest group of users for the Dengue Website and cluster map is the “Web-savvy public.”³² Given dengue’s potential for rapid transmission, arming the public with information on the areas of active transmission allows the public to take preventive measures against dengue fever, including removal of potential mosquito breeding and wearing insect repellent.³³ Prior to the NEA’s concerted effort to give the public this information, there was no surefire way for the public to know the areas of potential active dengue transmission, likely leading to states of either constant vigilance or gradual, learned complacency.

Media

Since the Dengue Website’s inception, journalists have increasingly turned to it for reliable information. According to the NEA, for example, “the mainstream media regularly uses the case count and case breakdown in clusters that are published on the website to report on the prevailing dengue situation, particularly during the peak season.”³⁴ With the importance of open dengue information to the media in mind, since 2014 the NEA has published a brief assessment of the dengue situation each week to go along with the hard data on dengue cases from the previous week, to inform media reports and, subsequently, the public of the situation on the ground.

³⁵

Developer Community

One of the more impressive spin-off benefits of the Dengue Website is evident in the galvanization and mobilization of Singapore’s developer community. In recent years, local developers have built a number of Web and mobile applications that build on the data and maps

³² GovLab Interview with the National Environment Agency of Singapore, July 23, 2015.

³³ GovLab Interview with the National Environment Agency of Singapore, August 31, 2015.

³⁴ GovLab Interview with the National Environment Agency of Singapore, July 23, 2015.

³⁵ GovLab Interview with the National Environment Agency of Singapore, July 23, 2015.

included on the website. Several of these applications improve on the functionality of the Dengue Website and illustrate open data's capacity to spur innovation and social and economic development.

Built from raw data sets available through an application programming interface (API) on the Singapore government open data portal, developers have created applications built using this API, in addition to the myENV app created by the NEA itself. Two of the most popular are:

- DengueLah, created by Buuuk, a mobile app development agency, tells the user where dengue active clusters are and, using location awareness, alerts the user if he or she approaches within a 2km radius of a known dengue cluster. The app also has a page with tips for prevention against dengue infection, illustrated with cartoons.
- X-Dengue, created through a collaboration between the NEA and Smart Communities Pte Ltd, a local software development company, alerts users on dengue clusters through location-based SMS and email alerts. Users can also check if their current location is considered to be at risk. The accompanying Web portal allows users to update danger zones that are then broadcast through the app.³⁶ X-Dengue has since been discontinued, with many users lamenting the loss of the platform on its Facebook page. “All the best!” one user writes, “You’ve helped lots of people across SG with your community spirit!”³⁷

Challenges

Over the course of the Dengue Website’s first decade of existence, it has increased in sophistication and become one of the measures in Singapore’s dengue control program. Yet, if Singapore’s open data efforts are to remain effective, particularly at a time when dengue

³⁶ “Dengue on the Rise—Some Apps to Stay Informed on Dengue Clusters.” *Iheartapple*. <http://iheartapple.com/2013/01/dengue-on-the-rise-some-apps-to-stay-informed-on-dengue-clusters/>.

³⁷ <https://www.facebook.com/xdengue>

itself has shown great resilience and even a resurgence around the world, it will need to overcome several challenges. These include:

- **Timeliness:** Dengue information is only useful if it is current. Especially during peak dengue months, when the dengue case-load is high, maintaining the currency of dengue data requires good coordination among the team assembling the data.
- **Privacy:** As described above, information on dengue cases must be anonymized in order to maintain the privacy of cases.³⁸ Given the fact that dengue is vector transmitted, there is a lessened concern of stigma being attached to carriers of the disease, but, of course, the need for maintaining the privacy of citizens remains essential. While speaking about dengue in particular—but pointing to challenges relevant for a broader range of open data initiatives—Ee-Peng Lim of the Living Analytics Research Centre describes a sometimes difficult balance “between protecting the data, the privacy of the data,” and allowing “certain room for people to innovate.”³⁹
- **Complacency:** The boundaries displayed in the dengue cluster map play a key role in making it known to the public where increased vigilance is required. These approximated boundaries, however, also had the unintended and unforeseen effect of fostering undue complacency in some segments of the population. In some cases, the NEA found that some members of the public felt that “they should only act (to undertake dengue prevention measures) when their residence falls within the cluster boundary.”⁴⁰ Upon recognizing this tendency, the NEA began releasing targeted communications to the public on the importance of sustaining their vigilance even during “peacetime” so as to inculcate a regular habit against mosquito breeding.
- **User-Friendly Competition:** In 2014, Xie Rufeng, a developer based in Singapore, created outbreak.sgcharts.com, a website for mapping cases of dengue in the country. Rufeng was aware of the NEA’s similar cluster map, but noted that he is a “big fan of Google Maps, which is highly customizable, so I decided to roll

³⁸ GovLab Interview with the National Environment Agency of Singapore, July 23, 2015.

³⁹ GovLab interview with Professor Ee-Peng Lim, Ph.D., Co-Director, Living Analytics Research Centre, Singapore Management University, September 28, 2015.

⁴⁰ GovLab Interview with the National Environment Agency of Singapore, July 23, 2015.

up my sleeves and do my own map.”⁴¹ Moreover, Rufeng’s site includes historical data on dengue cases, which is not found on the NEA cluster map. He notes, “I thought it will be interesting to provide an interactive timeline where users can view past dengue clusters. This information will help us to understand where dengue has re-emerged, or how long does it take for dengue clusters to reach a certain size.”⁴² With the existence of a feature-rich, user-friendly competitor, NEA will likely need to add new functionalities to its cluster map in order to keep pace. Additionally, as some users begin to rely on [outbreak.sgcharts.com](#), concerns are likely to arise over whether the information found on the third-party site is fully accurate, or if it is spreading any level of misinformation to the public.

Looking Forward

For over a decade, Singapore has used open, digital tools to give the public the information and tools necessary to stop the spread of dengue. Looking ahead, the NEA will likely need to make a choice about the future of its open data efforts around data. The first option would be to add additional functionalities to the site—including the use of the overwhelmingly popular Google Maps engine—to keep pace with similar non-governmental competitors. Adding new functionalities, however, will likely be resource-intensive, making focused efforts to improve the site difficult, especially considering that keeping the information housed on the site current (itself a resource-intensive task) will always be of paramount importance. The second option would be to move away from platform efforts and instead focus on simply collecting and supplying up-to-date, high-quality dengue data, and letting the developer community take the lead in producing public-facing platforms to mitigate the spread of the disease.

41 Xu, Terry. “App site tracks dengue cases in Singapore using Google Maps.” *The Online Citizen*. July 30, 2014. <http://www.theonlinecitizen.com/2014/07/app-site-tracks-dengue-cases-in-singapore-using-google-maps/>

42 Xu, Terry. “App site tracks dengue cases in Singapore using Google Maps.” *The Online Citizen*. July 30, 2014. <http://www.theonlinecitizen.com/2014/07/app-site-tracks-dengue-cases-in-singapore-using-google-maps/>

Beyond updates to the site, the NEA is conducting outreach to public health institutions and practitioners in other dengue-endemic regions to explore the potential for future collaborations and/or knowledge-sharing. This type of knowledge-sharing has occurred, for instance, at the regular Asia-Pacific Dengue Workshop, organized by the NEA and WHO, where participants from across dengue-endemic countries are trained in how to leverage geographic information system (GIS) data to develop situation awareness.⁴³ UNITEDDengue, a Web-based platform for data-sharing among countries in the region has also been initiated.⁴⁴

While there is new hope regarding a potential vaccine to prevent the spread of dengue,⁴⁵ the Singaporean government will likely continue to face the threat of outbreaks of the disease for at least the near future. By recognizing that the best way to engender safe, healthy decision-making regarding the spread of the disease is to give citizens timely access to information on the location of active dengue clusters, Singapore is taking steps toward turning the potential victims of dengue—whether application developers or citizens armed with information to help avoid enabling new cases—into the factors stopping its spread.

43 GovLab Interview with the National Environment Agency of Singapore, July 23, 2015.

44 <https://www.uniteddengue.org/index.html>

45 “Potential dengue vaccine protects those aged 9 and above: Study.” *Your Health Asia One*. August 4, 2015. <http://yourhealth.asiaone.com/content/potential-dengue-vaccine-protects-those-aged-9-and-above-study>

CHAPTER 19

United States' Eightmaps

The Unintended Negative Consequences of Open Data

By Auralice Graft, Stefaan Verhulst and Andrew Young

Summary: In 2008, voters in California were presented with a choice regarding the legal right of same-sex couples to marry. Proposition 8, which passed successfully on November 4, 2008, made same-sex marriage illegal in the state. Given the strong opinions on both sides of the issue, it is not surprising that debate did not conclude after the final ballot was counted. In 2009, the anonymous creator(s) of Eightmaps.com launched a tool providing detailed information on supporters of Prop 8. The site collected information made public through state campaign finance disclosure laws and overlaid that information onto a Google map of the state. Through Eightmaps, users could find the names, approximate locations, amount donated, and, where available, employers of individuals who donated money to support Prop 8 and stop same-sex marriage in the state. While the site is no longer live, and a federal court eventually found Prop 8 unconstitutional in 2010, Eightmaps demonstrates how the increased computability and reusability of open data—especially data related to contentious issues like same-sex marriage—might be used in unexpected ways that not only create major privacy concerns for citizens, but could also lead to harassment and threats based on political disagreements.

Impact: Solving Public Problems—Data-Driven Engagement

Key Takeaways:

- Open data efforts, especially those that involve individual-level information, have the potential for unintended consequences in the form of negative impacts to citizen privacy.
- Legislation focused on transparency and accountability passed prior to the rise of the Internet could have insufficient safeguards for privacy, given that today's data analysis and computation capabilities could not have been predicted to reach the current level even throughout the late 1990s and early 2000s.
- Accessible public data can be mashed up with other data sources using free tools to create new insights and public impacts. In some cases, those impacts are positive and create new value to users. In others, the results can be more troubling.

Context and Background

In the interest of creating transparency and accountability in the political process and around political messaging, among other reasons, the U.S. government today publicly discloses personal information about voters through two different data sets: those for voter data and those for campaign finance. Voter data sets are created when citizens register to vote. The type of data required varies to some extent by state. The state of California, where the case under consideration occurred, requires that voters provide their name, address, party affiliation, previous residence and signature, among other information. Aside from a citizen's signature, this data is readily accessible online. Some other states require the submission of a formal request to access such data. Campaign finance data sets on the other hand, are compiled and disclosed by campaigns that are required to file data about donors who give over a certain amount in support of that candidate or referenda. As with voter data, the type of data required to be disclosed also varies by state. In California, which is known to have likely the strongest campaign finance disclosure laws in the country,¹ a donor's name, address, occupation and employer are required. As of 2015, the donation threshold over which disclosure is mandated is \$100 in California, with campaign

¹ Alexander, Kim. "Initiative Disclosure Reform: Overview and Recommendations." Greenlining Institute. June 16, 2011. <http://www.calvoter.org/issues/disclosure/pub/greenliningpaper.pdf>

finance data available on the California Secretary of State's website.² Anyone with access to the Internet can view it online or download it as a Microsoft Excel file.³ A more detailed report including donors' complete addresses (which is redacted from the online display) can also be accessed by submitting a signed request to the California Secretary of State's office, at which point it is delivered on a CD. Large data users, like news organizations, can also access a live online feed following the completion of an agreement with the Secretary of State's office.

Legal Foundation of Campaign Finance Data Disclosure Laws in California

Written in the 1970s as an outgrowth of the Watergate scandal, California's campaign finance disclosure laws built upon earlier requirements to provide just the names of donors openly to the public. Then a movement gained steam to ensure that the public remain fully informed, and that improper practices (like those relating to the Watergate scandal) be prohibited by the enactment of new disclosure laws. California's current Governor Jerry Brown who was secretary of state at the time, and Bob Stern of the Center for Government Studies, among others, wrote California's Political Reform Act of 1974, which required that all candidates disclose the name, address, occupation and employer for every donation over the threshold—which at the time was \$50. Information must be filed for donations toward candidates (to limit the potential for corruption) as well as donations toward issues and referenda (to help voters know who supports what, which in turn helps them make informed political decisions).⁴ The law stipulates that disclosure must be made in an itemized, timely manner, and include all contributions and expenditures. It also prohibits anonymous contributions and contains a regulatory framework imposing penalties for failure to comply. Kim Alexander of the California Voter Foundation recalls, “Writing and passing the law was done through the initiative pro-

² cal-access.sos.ca.gov/

³ Messner, Thomas M. “The Price of Prop 8.” Heritage Foundation. October 22, 2009. <http://www.heritage.org/research/reports/2009/10/the-price-of-prop-8>

⁴ GovLab interview with Kim Alexander, President and Founder of the California Voter Foundation, September 16, 2015.

cess, so the language was very voter focused in terms of people having the right to be informed.”⁵

Electronic Filing

Initially, after disclosure laws were enacted, campaign finance data was filed on paper and housed in filing cabinets and accessible at the Secretary of State’s office. However, in the decades that followed, it became increasingly clear to many who were familiar with this data that it should be digitized. Alexander, who worked to change laws to require that reports be filed electronically, says, “It was clear there was so much more to be learned about what was going on in campaign finance if campaign finance records were computerized.”⁶ The first bill was drafted in 1995, and the final bill enacted in 1997. The 1997 bill states that candidates’ disclosure records have to be filed online: “This bill would require the secretary of state to develop a process whereby reports and statements that are required under the act to be filed with the secretary of state could be filed electronically and viewed by the public at no cost on an online disclosure system by way of the largest nonproprietary, cooperative public computer network.”⁷ California was among the first states to pass these electronic filing laws.⁸ Alexander recalls, “I was very aware of the fact that if California did pass this law, it would be seen as a ‘bellwether’ kind of act that other states would likely copy. Because we’re California, we have a lot of influence in the rest of the country.”⁹ San Francisco was the first jurisdiction to enact and implement an electronic filing law, in 1993 and 1995, respectively.¹⁰

⁵ Alexander, Kim. “Initiative Disclosure Reform: Overview and Recommendations.” Greenlining Institute. June 16, 2011. <http://www.calvoter.org/issues/disclosure/pub/greenliningpaper.pdf>

⁶ GovLab interview with Kim Alexander, President and Founder of the California Voter Foundation, September 16, 2015.

⁷ Senate Bill No. 49, Chapter 866. http://www.leginfo.ca.gov/pub/97-98/bill/sen/sb_0001-0050/sb_49_bill_19971011_chaptered.pdf

⁸ Alexander, Kim. “Initiative Disclosure Reform: Overview and Recommendations.” Greenlining Institute. June 16, 2011. <http://www.calvoter.org/issues/disclosure/pub/greenliningpaper.pdf>

⁹ GovLab interview with Kim Alexander, President and Founder of the California Voter Foundation, September 16, 2015.

¹⁰ GovLab interview with Kim Alexander, President and Founder of the California Voter Foundation, September 16, 2015.

Privacy Concerns of Electronic Filing

People like Kim Alexander of the California Voter Foundation were adamant advocates of electronic filing, but were also aware that digitization could have negative consequences in terms of privacy. She recalls, “We didn’t want to create a process where people who had the right to speak with their dollars—as the U.S. Supreme Court has affirmed they can—felt that right was being denied because their personal information was disclosed to a wide audience; that there shouldn’t be this price they would pay.”¹¹ This was in the mid-1990s, before the meteoric rise of the Internet. At the time, campaign finance data reports were provided on CDs or floppy disks. Electronic filing on the Internet as per campaign disclosure laws only began in 2000. From the beginning, stakeholders had a sense that there were potential security threats related to the electronic filing, but not to the extent that became possible after the rise of the Internet. To calm security concerns, the final version of 1997 bill required that donors’ addresses be redacted from the online display. Candidates still had to file addresses as part of their reports—so they were included in the formal record—but they were not viewable on secretary of state’s website. Donor addresses were still publicly available but digital access to them required a special, signed agreement with secretary of state’s office. The paper records, however, continue to be filed with the secretary of state and are made publicly accessible free of any redaction.¹²

California’s Proposition 8

Proposition 8 was a California state constitutional amendment created by opponents of same sex marriage forbidding the licensing or recognition of same sex marriages in California. The proposition was titled the “California Marriage Protection Act,” and read: “Only marriage between a man and a woman is valid or recognized in the State of California.”¹³ Proposition 8 was preceded by Proposition 22 which was adopted in 2000 as an ordinary statute with exactly the

¹¹ GovLab interview with Kim Alexander, President and Founder of the California Voter Foundation, September 16, 2015.

¹² GovLab interview with Kim Alexander, President and Founder of the California Voter Foundation, September 16, 2015.

¹³ http://www.ag.ca.gov/cms_pdfs/initiatives/i737_07-0068_Initiative.pdf

same wording as Proposition 8. Proposition 22 was invalidated on constitutional grounds at the U.S. Supreme Court level in 2008. Proposition 8 replaced it as a state constitutional amendment and was passed in 2008 California state elections with a vote of 52-48 percent. It was subsequently ruled as unconstitutional by a federal court in 2010.

The Proposition 8 campaign was extremely polarized. Supporters and opponents alike felt passionate about its outcome and significance. While supporters argued vehemently that marriage should be reserved for men and women, opponents felt equally strongly that California's constitution should, as the legal document argues, "guarantee the same freedoms and rights to everyone," and that "no one group should be singled out to be treated differently."¹⁴ In many cases, voters iterated their beliefs and sentiments with donations. Kim Alexander notes that "there was *a lot* of money spent on both sides of this initiative. It was one of the most expensive initiatives we've ever seen [in California]."¹⁵ The LA Times described it as the "state's costliest initiative," reporting that a total of more than \$83 million was raised on both sides.¹⁶

Donors to the Proposition 8 campaign experienced widespread intimidation and public shaming, led in large part by a group called Californians Against Hate. This group's stated mission was to "fight back" against those supporting Proposition 8 by letting "the world know who donors [to the proposition] are."¹⁷ Their website listed people and organizations who had contributed \$5,000 or more to the campaign. That list included individuals' addresses, which were publicly available as a result of campaign finance laws, as well as phone numbers and website addresses—information that is not included in the publicly accessible data, but rather collected manually through research on identified individuals. The group, and others affiliated to it—including [BoycottManchesterHotels.com](#) and [BoycottAISelfStorage.com](#)—also organized boycotts against very

¹⁴ California General Election—Official Voter Information Guide. November 4, 2008. <http://vigarchive.sos.ca.gov/2008/general/argu-rebut/argu-rebutt8.htm>

¹⁵ GovLab interview with Kim Alexander, President and Founder of the California Voter Foundation, September 16, 2015.

¹⁶ "Proposition 8: Who gave in the gay marriage battle?" *Los Angeles Times*. <http://projects.latimes.com/prop8/>

¹⁷ <http://www.californiansagainsthate.com/>

large donors, including a hotel owner who contributed \$125,000 to Yes on Proposition 8, and a storage company whose owners donated \$700,000. As a result of these efforts, in some cases people lost their jobs due to their support for the proposition.¹⁸

Kim Alexander echoes a widely held opinion in California and beyond, “Voters should have never been asked to decide Proposition 8 in the first place. Now we know in hindsight that it was an unconstitutional proposition. People were in a fight over constitutional rights, and that’s part of the reason why there was so much passion. That’s why it was so contentious. In hindsight, we can say it should never have gone before the voters in the first place.”¹⁹

Project Description and Inception

Californians Against Hate targeted larger donors to Proposition 8. But several websites showed less restraint, including Eightmaps.com. Launched in early 2009 by opponents of Proposition 8, the creators of Eightmaps—also known as “Prop. 8 Maps”—took the publicly available names and address information of anyone who had given more than \$100 to the campaign and “mashed it” (i.e., overlaid it) with Google maps to provide an electronic map to the residences and workplaces of those donors. Gawker—a media outlet that at times traffics in uncomfortable levels of transparency²⁰—called it an “icon for the extremes to which political transparency can be taken.”²¹

Eightmaps.com is no longer available online, but an archive search brings up some of its pages (See [Figure 19-1](#) below). There, a headline announces “Prop 8 Maps: A mash-up on Google Maps and Prop 8 Donors,” and “Proposition 8 changed the California state

18 Richardson, Valerie. “Pestered Prop 8 donors file suit.” *The Washington Times*. March 23, 2009. <http://www.washingtontimes.com/news/2009/mar/23/pestered-prop-8-donors-file-suit/?page=all>; “Boycotts related to California Proposition 8.” *Ballotpedia*. http://ballotpedia.org/Boycotts_related_to_California_Proposition_8

19 GovLab interview with Kim Alexander, President and Founder of the California Voter Foundation, September 16, 2015.

20 Roy, Jessica. “Gawker Slammed for Story Outing Condé Nast Exec.” *New York Magazine*. July 17, 2015. <http://nymag.com/thecut/2015/07/gawker-slammed-for-story-outing-conde-nast-exec.html>

21 Tate, Ryan. “Map of Anti-Gay Donors Created by Big Chicken.” *Gawker*. February 8, 2009. <http://gawker.com/5149276/map-of-anti-gay-donors-created-by-big-chicken>

constitution to prohibit same-sex marriage. These are the people who donated in order to pass it.” The site displays a Google map that apparently used to depict markers indicating contributors’ names, approximate location, amount contributed, and, if available, their employer (these markers are no longer available in the archived version). The site provides the option to view maps in San Francisco, Salt Lake City and Orange County.²²

It is not clear when exactly Eightmaps was launched or when or how the wider public first began realizing it existed. But beginning the week of February 9, 2009, it was covered widely in the U.S. media. Articles about it appeared in many major newspapers, including *The New York Times*,²³ the *Los Angeles Times* and the *San Francisco Chronicle*, as well as in blogs and other news sources. Reactions varied from concern about negative outcomes to praise for the idea. Referencing Eightmaps, one headline asked, “Google Map: Intimidation or Conversation Starter?”²⁴

The New York Times wrote that Eightmaps revealed a darker side of transparency laws: “Eightmaps.com is the latest, most striking example of how information collected through disclosure laws intended to increase the transparency of the political process, magnified by the powerful lens of the Web, may be undermining the same democratic values that the regulations were to promote.” This could lead to citizens becoming discouraged from participating in the political process, the article pointed out.²⁵

A blogger for the *Dallas Morning News* saw Eightmaps as “a terrible development,” and worried about the threat to Proposition 8 donors’ privacy. “Given that there has been harassment by gay radicals of people who supported Prop 8, it is potentially dangerous that it’s

22 <http://web.archive.org/web/20090302143108/http://www.eightmaps.com/>

23 Stone, Brad. “Prop 8 Donor Web Site Shows Disclosure Law Is 2-Edged Sword.” *The New York Times*. February 7, 2009. http://www.nytimes.com/2009/02/08/business/08stream.html?_r=0

24 “Google Map: Intimidation or Conversation Starter?” WND. February 11, 2009. <http://www wnd com/2009/02/88616/>

25 Stone, Brad. “Prop 8 Donor Web Site Shows Disclosure Law Is 2-Edged Sword.” *The New York Times*. February 7, 2009. http://www nytimes com/2009/02/08/business/08stream.html?_r=0

now easy to find your way to the homes of these donors—even small donors,” they opined.²⁶

Others saw the effort as justified, and suggested that it gave Proposition 8 donors what they deserve. The SFist.com blogged:

“Is this a witch hunt? No. Is it cruel? A bit, maybe. But aren’t the 36,000 people whose marriages risk nullification entitled to know which of their neighbors paid to force them into divorce? Criminals are entitled to face their accusers in court. why can’t married couples know who’s funding their forced divorces? And do the yes-on-8 people REALLY want to play the ‘who’s a bigger victim’ game with the gays? No.”²⁷

Another blogger described the site as “an interesting convergence ... between information technology and election donor transparency,” arguing that: “The laws in California are clear: donate enough and your name goes public. If the filth who donated to support institutionalized bigotry are truly proud of themselves, this shouldn’t be a problem for them. Keep in mind Eightmaps doesn’t expose everyone who voted for Prop 8, only those who donated enough that they tripped California’s public disclosure laws. Well, you get what you pay for.”²⁸

The creators of eightmaps.com remain anonymous and made no statements about the site. They also refused all requests to be interviewed. This makes it impossible to know about project inception. No major efforts to unmask the creators were mentioned in the media, though many have observed the irony of the creators’ choice to remain anonymous, given the full-disclosure nature of the website.²⁹ Gawker referred to them at the time as “Big Chicken[s],” stating that “when the people behind a public shaming tool insist on

26 “Eightmaps.com and too much information.” *The Dallas Morning News*. January 14, 2009. <http://dallasmorningviewsblog.dallasnews.com/2009/01/eightmapscom-an.html/>

27 Keeling, Brock. “Map of Prop 8 Donors.” *SFist*. January 9, 2009. http://sfist.com/2009/01/09/mash-up_map_of_google_maps_and_prop.php

28 “Exposing bigots: eightmaps.com.” *The Planetologist*. February 14, 2009. <https://planetologist.wordpress.com/2009/02/14/exposing-bigots-eightmapscom/>

29 Lincoln, Ross A. “Eightmaps.com: Hypocritical Privacy Violation, or Reverse-Super Judo?” *LAist*. February 19, 2009. http://laist.com/2009/02/19_for_people_concerned_that_the.php

lurking in the shadows, they cover their particular tactical effort in a veneer of slime.”³⁰



Figure 19-1. Screenshot of eightmaps.com. Image copied from the New York Times³¹

Impact

The impact of the Eightmaps project was significant. It began with widespread media coverage and, importantly, a public debate on transparency in political donations—both arguably positive outcomes. Not surprisingly, other impacts were not as positive.

Harassment of Proposition 8 Donors

Many donors to Proposition 8 began experiencing threats, vandalism, intimidation and property destruction in the weeks and months after Eightmaps was launched—though whether or not their information was obtained from Eightmaps or other similar sources cannot be confirmed. Some reported receiving envelopes containing white powder. *The Washington Times* quoted Charles LiMandri, a Proposition 8 supporter who lived in San Diego as having begun to receive unexpected correspondence after Eightmaps. “I got about

³⁰ Tate, Ryan. “Map of Anti-Gay Donors Created by Big Chicken.” Gawker. February 8, 2009. <http://gawker.com/5149276/map-of-anti-gay-donors-created-by-big-chicken>

³¹ Stone, Brad. “Prop 8 Donor Web Site Shows Disclosure Law Is 2-Edged Sword.” *The New York Times*. February 7, 2009. http://www.nytimes.com/2009/02/08/business/08stream.html?_r=0

two dozen emails and hate phone calls,” Mr. LiMandri was quoted as saying.³² Others received messages like “Burn in Hell,” “Consider yourself lucky,” and “If I had a gun I would have gunned you down along with each and every other supporter.”³³ Other Proposition 8 donors reported being pushed out of their jobs following the release of their donation activity.³⁴

The result was a potential chilling effect on some donors’ sense of safety. While the Californians Against Hate website had also provided donors’ names and addresses, these were for a smaller number of people (because of the higher threshold) and Eightmaps took things one step further by actually depicting addresses, along with other information about the donor, on a map. Alexander of the California Voter Foundation said, “This idea of giving people a map to someone’s home is a way of inviting widespread attack on that person. When you put that on the Internet, you are literally providing an invitation or suggestion that people go to this person’s home and confront them. This was threatening on an exponential level.”³⁵

This chilling effect was largely the result of Eightmaps mashing up of publicly accessible and manually collected data. While individual-level voter and political donor data has been available to the public for years, mashups were far from the norm even late into the 00s. “In 2008 [just before Eightmaps launched], people weren’t doing that with the Internet,” said Kim Alexander.³⁶

In this way, Eightmaps was an important, thought-provoking lesson for open data advocacy. Advocates had argued tirelessly about the benefits of disclosure, about the possibility for limiting corruption and of informing voters. But they had not perhaps considered the

³² Richardson, Valerie. “Pestered Prop 8 donors file suit.” *The Washington Times*. March 23, 2009. <http://www.washingtontimes.com/news/2009/mar/23/pestered-prop-8-donors-file-suit/?page=all>

³³ “Google Map: Intimidation or Conversation Starter?” WND. February 11, 2009. <http://www wnd com/2009/02/88616/>

³⁴ Richardson, Valerie. “Pestered Prop 8 donors file suit.” *The Washington Times*. March 23, 2009. <http://www.washingtontimes.com/news/2009/mar/23/pestered-prop-8-donors-file-suit/?page=all>

³⁵ GovLab interview with Kim Alexander, President and Founder of the California Voter Foundation, September 16, 2015.

³⁶ GovLab interview with Kim Alexander, President and Founder of the California Voter Foundation, September 16, 2015.

extent to which outcomes could be negative as well. Alexander notes, “It’s really hard when you’re an open data advocate to consider all the potentially negative things that can be done.”³⁷ These risks are, of course, magnified dramatically by the Internet. “We realized that when you make data available to everyone [on the Internet], that means *everyone*.”³⁸

Threats to Free Speech

Projects like Eightmaps can also threaten free speech by undermining disclosure laws. On January 8, 2008, prior to the believed date of Eightmaps’ launch, a federal lawsuit was filed by James Bopp that argued California disclosure laws should be changed so candidates would no longer be required to report information on Proposition 8 donors. Much of the case was built on testimony of those who had felt threatened in the wake of Proposition 8, and subsequently demanded that the disclosure law be thrown out or donors’ names shielded as a result of retaliation against some contributors. “Certainly the state has an interest in requiring disclosure,” said Richard Coleson, an attorney representing the plaintiffs, Protect Marriage, “but there has to be an exception when there is a reasonable probability of reprisal.”³⁹ The case went before U.S. District Judge Morrison England in November 2008, and was denied. “If there ever needs to be sunshine on a particular issue, it’s a ballot measure,” he concluded, emphasizing the continued need for transparency on issues such as the Proposition 8 measure. He also said that illegal reprisals resulting from online databases—like those available on Eightmaps the following year—while “repugnant and despicable,” could be dealt with by law enforcement, without giving donors additional legal protection.⁴⁰

³⁷ GovLab interview with Kim Alexander, President and Founder of the California Voter Foundation, September 16, 2015.

³⁸ GovLab interview with Kim Alexander, President and Founder of the California Voter Foundation, September 16, 2015.

³⁹ “Prop 8 Donors: Find Out Who Backed California’s Anti-Gay Marriage Amendment.” *The Huffington Post*. May 25, 2011. http://www.huffingtonpost.com/2009/02/02/prop-8-donors-find-out-wh_n_163234.html?ir=India&adsSiteOverride=in

⁴⁰ Egelko, Bob. “Prop. 8 campaign can’t hide donors’ names.” *SFGate*. January 30, 2009. <http://www.sfgate.com/bayarea/article/Prop-8-campaign-can-t-hide-donors-names-3174252.php>

Activism Around Gay Rights

Not everyone focused on the negative impacts of Eightmaps. Some saw its impact to be positive, particularly in terms of activism around gay rights. Opponents of Proposition 8 were thrilled to see what they saw to be supporters of discrimination (by opposing gays' right to marriage) targeted and boycotted.⁴¹ Erica Anderson, of the Erica-America.com blog was quoted in the MTV newsroom as saying "With Eightmaps, the gays are fighting back, and I say good for them, good for us. This tool is a perfect example of democracy meeting Web 2.0—in a thoughtful, productive way."⁴²

Californians Against Hate founder Fred Karger was propelled into gay rights activism by his fight against Proposition 8: "Our opponents want to send a message to all of us that we are second-class citizens, who are not entitled to the same rights as our mothers and fathers, brothers and sisters, aunts, uncles, co-workers, neighbors and our friends who happen to be straight. This will not deter me. I will continue to bring attention to those companies and individuals who spend millions and millions of dollars to stop us from attaining full and equal civil rights," he wrote in a September, 2009 *Huffington Post* piece.⁴³

This impact on gay rights was not limited to the immediate period around the launch of Eightmaps, but has snowballed into more permanent changes in national cultural attitude—particularly in terms of Silicon Valley and the world of information technology. Five years after Eightmaps, in 2014, Mozilla CEO Brendan Eich was forced to resign days after assuming the position when he was outed as having donated \$1,000 to support Proposition 8 back in 2008. His company's Firefox Web Browser faced boycott threats as a result of that support and he resigned saying, "Under the present circumstances, I cannot be an effective leader." The *SFGate* newspaper quoted Frank Schubert, a political consultant who ran the Prop. 8 campaign, as

⁴¹ GovLab interview with Daniel Kreiss, Assistant Professor, University of North Carolina School of Journalism and Media, September 18, 2015.

⁴² Stoltz, Kim. "Are Prop 8 Opponents Using EightMaps.com for the Right Reasons?" *MTV News*. February 10, 2009. <http://newsroom.mtv.com/2009/02/10/are-prop-8-opponents-using-eightmapscom-for-the-right-reasons/>

⁴³ Karger, Fred. "Fighting for Civil Rights Has Consequences." *The Huffington Post*. November 21, 2009. http://www.huffingtonpost.com/fred-karger/fighting-for-civil-rights_b_294273.html?ir=India&adsSiteOverride=in

saying “There’s now no place in current society for holding a view that people have held for thousands of years.”⁴⁴

Not everyone was pleased with Eich’s predicament. Gay activist and commentator Andrew Sullivan blogged that he was disgusted by the episode, saying, “If this is the gay rights movement today—hounding our opponents with a fanaticism more like the religious right than anyone else—then count me out.”⁴⁵

Transparency vs Privacy

Not surprisingly, the voices in support of Eightmaps are less prevalent than those who are troubled by the potential for other negative outcomes, regardless of their political leanings or feelings about gay marriage in particular. Concern about a preponderance of threats and intimidation and particularly about destruction of privacy govern much of the discussion in the wake of the Eightmaps case. In fact, this issue is really the crux of the case. How to mitigate the tension between the need for transparency in voter and campaign finance data with the need to protect citizens’ privacy? At what point does one need become more important than the other? We explore these questions in the next section on Challenges.

Challenges

A primary outcome of the case of Eightmaps is the challenge of how to balance the sometimes competing goals of transparency and privacy. This can also be framed as the right to speak with one’s dollars versus the right to be informed about who is funding campaigns—rights the U.S. Supreme Court has said are vital to the decision-making process in the states that have it.⁴⁶ At what point does one right become more important than the other?

⁴⁴ Wildermuth, John. “Mozilla’s Prop. 8 uproar reveals much about tech, gay rights.” *SFGate*. April 11, 2014. <http://www.sfgate.com/politics/article/Mozilla-s-Prop-8-uproar-reveals-much-about-tech-5393875.php>

⁴⁵ Wildermuth, John. “Mozilla’s Prop. 8 uproar reveals much about tech, gay rights.” *SFGate*. April 11, 2014. <http://www.sfgate.com/politics/article/Mozilla-s-Prop-8-uproar-reveals-much-about-tech-5393875.php>

⁴⁶ GovLab interview with Kim Alexander, President and Founder of the California Voter Foundation, September 16, 2015.

Transparency and privacy compete against each other in several domains of the law. For example, in domains that relate to court records or community notification laws for sex offenders. The Eightmaps case shows how campaign finance laws are another domain. On the one hand, individual privacy is considered important because, as one journal article puts it, “the secret ballot and associational privacy are at stake.”⁴⁷ On the other hand, public interests such as the need to limit corruption and keep citizens informed can outweigh the importance of privacy.⁴⁸ Also to be considered are the dangers of threatening individuals’ sense of privacy and security, one of which is that it might inhibit them from participating in political life.

These tensions are an unintended outcome of transparency laws. “I don’t think the law was designed to identify people for direct feedback to them from others on the other side,” Joseph Clare, a San Francisco accountant, was quoted as saying in *The New York Times*, “I think it’s been misused.”⁴⁹ They are certainly questions that have been brought to the fore by the advent of the Internet. Disclosure laws were written before the Internet, when donors’ information resided in government filing cabinets and would largely be inaccessible to many people. “Nobody anticipated when the laws were initially written that reports were going to be widely accessible to everybody in the world,” says Alexander of the California Voter Foundation, who helped update disclosure laws. “Identity theft was not an issue at the time. It didn’t matter so much if people knew your address—there was not much they could do with it. Now you can do things like mash it up with other data—which is what happened with Eightmaps,” she said.⁵⁰

47 Johnson, Deborah G., Priscilla M. Regan and Kent Wayland. “Campaign Disclosure, Privacy and Transparency.” *William & Mary Bill of Rights Journal*, Vol. 19, No. 4. 2011. <http://scholarship.law.wm.edu/cgi/viewcontent.cgi?article=1585&context=wmborj>

48 Johnson, Deborah G., Priscilla M. Regan and Kent Wayland. “Campaign Disclosure, Privacy and Transparency.” *William & Mary Bill of Rights Journal*, Vol. 19, No. 4. 2011. <http://scholarship.law.wm.edu/cgi/viewcontent.cgi?article=1585&context=wmborj>

49 Stone, Brad. “Prop 8 Donor Web Site Shows Disclosure Law Is 2-Edged Sword.” *The New York Times*. February 7, 2009. http://www.nytimes.com/2009/02/08/business/08stream.html?_r=0

50 GovLab interview with Kim Alexander, President and Founder of the California Voter Foundation, September 16, 2015.

These concerns also do not resound with everyone. For example, Daniel Kreiss, who studies the impact of technological change on the public sphere and political practice, and was aware of Eightmaps at the time of launching, did not share concerns about its privacy outcomes. While condemning any speech or activity intended to silence or harm another person, Kreiss also indicates that there is no empirical evidence of people voting less or being less politically active because of laws requiring that their political donation data being made public—or as a result of privacy concerns. Kreiss is far more concerned about transparency in terms of how political data is used by third-party data processing entities—for example, by Super PACs. He feels it is ironic that there is more talk about transparency in terms of the information the government collects than about such third-party efforts to use that data. To Kreiss, projects like Eightmaps are both “a valuable tool,” and also par for the course in a democratic system. “Outcomes like Eightmaps are part and parcel of participating in political life. If we’re going to say that money is speech and there is going to be transparency and accountability for what we say, then we also have to accept the consequences of our actions. If you are going to donate to a cause, you should expect there may be consequences like this,” he said.⁵¹

Ira Rubinstein, who writes about privacy and big data, agrees. Rubinstein doesn’t believe voter privacy is so important once a citizen gets to the stage of their political life where they are voting or donating to causes. He does think citizens need privacy in the earlier, more formative stages of their political lives. “At that stage they need space of quiet reflection for formatting their own positions,” he said. “But once a person takes that step toward supporting or opposing something that has a clear public face, that person has entered the realm of politics” and has to recognize what that means in terms of possible outcomes, he said.⁵²

Rubinstein does, however, think the U.S. tilts too far in favor of transparency and away from privacy, at least in the context of small-donor donations. Rubinstein also reiterates Kreiss’ concerns about a lack of transparency around larger political donations. “The prob-

⁵¹ GovLab interview with Daniel Kreiss, Assistant Professor, University of North Carolina School of Journalism and Media, September 18, 2015.

⁵² GovLab interview with Ira Rubinstein, Research Fellow and Adjunct Professor of Law at New York University, September 14, 2015.

lem is that big money gets to shield who donors are—through Super PACs and other similar entities—whereas small donors have to say who they are.”⁵³

Looking Forward

To help mitigate the potentially negative impacts of political data disclosure, a few strategies could be considered.

Raising the Itemized Disclosure Threshold

By raising the threshold after which in-depth disclosure must occur, people could take an active role in certain pieces of political life without being concerned about potentially threatening activity happening as a result. Rubinstein and Alexander, for example, argue that there is no strong reason not to raise the minimum given the high potential to improve privacy with a very small likelihood of negative impacts on anticorruption efforts. Others, like Kreiss, see little justification in restricting the collection and transparency of political donation data, given an individual’s conscious decision to play a part in political life.⁵⁴

Thresholds do create additional challenges, however. Alexander and the California Voter Foundation found that many campaigns itemize contributions at \$99—just below the threshold for public disclosure.⁵⁵ As a result, the foundation recommended that the Fair Political Practices Commission enforcement agency more aggressively monitor and fine campaigns intentionally obscuring donor information that, by law, should be made public.⁵⁶

⁵³ GovLab interview with Ira Rubinstein, Research Fellow and Adjunct Professor of Law at New York University, September 14, 2015.

⁵⁴ GovLab interview with Daniel Kreiss, Assistant Professor, University of North Carolina School of Journalism and Media, September 18, 2015.

⁵⁵ Alexander, Kim. “Testimony at Public Hearing of the Fair Political Practices Commission’s Subcommittee on the Political Reform Act & Internet Political Activity.” March 17, 2010. <http://www.calvoter.org/issues/disclosure/pub/031710KAremarks.html>

⁵⁶ GovLab interview with Kim Alexander, President and Founder of the California Voter Foundation, September 16, 2015.

Placing New Safeguards on Downloads of Donor Information

While maintaining transparency, states could consider putting new safeguards onto their campaign activity and disclosure databases to help avoid threats to privacy. By asking for, for example, a valid form of identification before allowing for the bulk download of individual-level donor data, users with malicious intent could reconsider and/or the proper authorities could more easily address any illegal activities enabled by access to the data.

Better Transparency About How Data Is Used by Political Campaigns and Other Actors

The question of transparency versus privacy is perhaps of greatest concern not so much when it comes to cases like Eightmaps that use government-collected citizen data, but when it comes to what Ira Rubinstein calls “enhanced political data.” This data is what Rubinstein calls “more granular”: It is founded on government-collected voter data and then enhanced by a combination of data from commercial brokers as well as historical participation data that a party or candidate gathers over time. Political campaigns and parties have created such mashups for decades, but the ability to create them has expanded greatly in recent years. Such data offers more insight into citizens’ profiles and preferences, and is not available to the public at all. It is used and swapped by political campaigns and other actors to target the electorate. Daniel Kreiss feels there is “an alarming lack of transparency” here.⁵⁷ Alexander, Rubinstein and Kreiss all emphasize a need to require that entities trading in this data to be more accountable and explicit about how they’re using it.⁵⁸

⁵⁷ GovLab interview with Daniel Kreiss, Assistant Professor, University of North Carolina School of Journalism and Media, September 18, 2015.

⁵⁸ “Voter Privacy in the Digital Age: Recommendations.” California Voter Foundation. June 9, 2004. <http://www.calvoter.org/issues/votprivacy/pub/voterprivacy/recommendations.html>

After Proposition 8 was found unconstitutional by the U.S. Supreme Court in 2010, much of the public memory has moved past the fractious debates of 2008 and 2009. And while marriage equality is now the law of the land, and Eightmaps is no longer accessible to the public, the questions and debates it raised about balancing transparency and privacy in the realm of individual-level political activity remain.

CHAPTER 20

Kennedy v. City of Zanesville, United States

Open Data as Evidence

By Christina Rogawski, Stefaan Verhulst and Andrew Young

Summary: For over 50 years, while access to clean water from the City of Zanesville water line spread throughout the rest of Muskingum County, residents of the predominantly African-American area of Zanesville, Ohio were only able to use contaminated rainwater or drive to the nearest water tower to truck water back to their homes. After years of legal battles, one of the key pieces of evidence used during Kennedy v the City of Zanesville was a map derived from open data from the water company displaying houses connected to the water line and data showing town demographics. The insights from the map showed significant correlation between the houses occupied by the white residents of Zanesville and the houses hooked up to the city water line, and the case went in favor of the African-American plaintiffs, awarding them a \$10.9 million settlement.

Dimension of Impact: Solving Public Problems—Data-Driven Assessment

Key Takeaways:

- Access to open data, combined with other forms of data, can lead to important insights and evidence of conditions on the ground and how they are impacting different communities—in this case, highlighting systemic inequalities.

- Awareness of open data is an important first step that may be overlooked, particularly if stakeholders are not particularly data- or tech-savvy.
- Owners of data sets may add new (or more rigorously enforce existing) barriers to access if the data can negatively impact them.
- The usefulness and relevance of open data can be amplified when data sets are supplemented with data collected through other means—like crowdsourcing or surveys.

Context and Background

For decades, residents of the Coal Run neighborhood in Zanesville Ohio, a predominately African-American neighborhood, were denied public water service despite living within one mile of public water lines.¹ The situation went back to 1956, when a now-defunct water board refused to extend service to parts of Coal Run. As some residents described in a 2008 *New York Times* article, the water stopped “where the black folks started.”²

¹ “Kennedy v. City of Zanesville.” Relman, Dane & Colfax PLLC. <http://www.relmanlaw.com/civil-rights-litigation/cases/zanesville.php>

² Johnson, Dirk. “For a Recently Plumbed Neighborhood, Validation in a Verdict.” *The New York Times*, August 11, 2008. http://www.nytimes.com/2008/08/12/us/12ohio.html?_r=0.

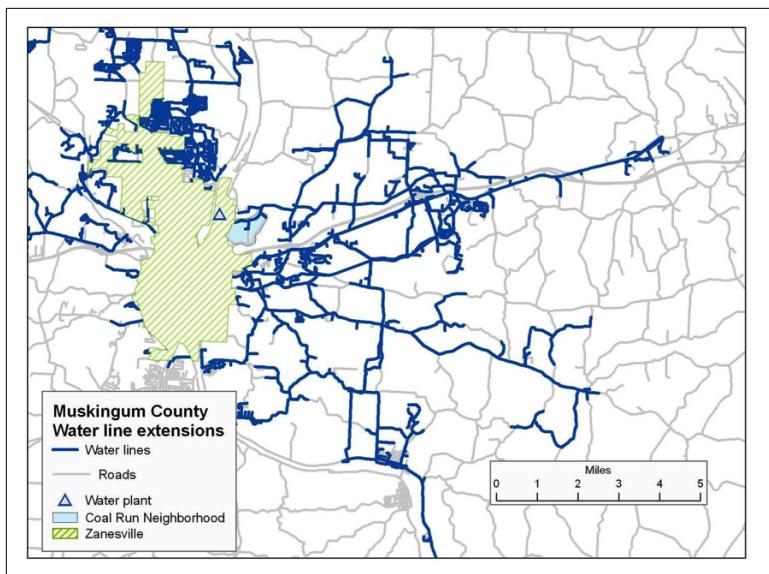


Figure 20-1. Map of Water Line Extensions in Muskingum County³. Copyright Allan Parnell. Used with permission.

Many residents were forced to rely on extreme measures to source water. For example, they had to use electric pumps to retrieve water from a cistern that was fouled with animals and residue from old coal deposits. Due to contamination, many residents couldn't even use the water and spent time and money instead trucking water in.⁴ Others collected rain via buckets and gathered snow in the winter.⁵

Not only did the situation impose a daily burden on residents, it was also demeaning and humiliating. One lawyer cited the following example, which reveals the racial disparities in water distribution:

³ Parnell, Allan M. "Maps Used in Support of the Plaintiff's Argument in Kennedy et al. v. City of Zanesville, et al." Legal Services of Northern California Race Equity Project. August 6, 2008. <http://equity.lsn.org/2008/08/maps-used-in-support-of-the-plaintiff-%E2%80%99s-argument-in-kennedy-et-al-v-city-of-zanesville-et-al/>

⁴ GovLab interview with Tara Ramchandani, Attorney, Relman, Dane, & Colfax PLLC, August 3, 2015.

⁵ Johnson, Dirk. "For a Recently Plumbed Neighborhood, Validation in a Verdict." *The New York Times*, August 11, 2008. http://www.nytimes.com/2008/08/12/us/12ohio.html?_r=0.

“One man ... spent the whole morning trying to get water or deal with the water shortage. Meanwhile, he could see his white neighbor caddy corner to him sprinkling his lawn. It became clear if you were white and living outside Zanesville you would get water, but if you were black, you wouldn’t.”⁶

In 2002, some two dozen black residents of Coal Run filed a complaint with the Ohio Civil Rights Commission, saying they had been denied service because of race. The next year, the commission found “probable cause” of discrimination and a month after that, Muskingum County officials announced they would extend water to Coal Run, to be completed in 2004.⁷

The decision to extend the water lines did not mark the end of the battle, however. In 2005, after construction on the new water lines was completed, 67 residents of the Coal Run neighborhood filed a lawsuit, alleging that the City of Zanesville and the East Muskingum Water Authority had refused to provide them public water service for over 50 years simply because they lived in the one predominately African-American neighborhood in a virtually all-white county⁸—in the 2000 Census, Muskingum County was found to have a 93.9 percent white population, with the black community making up only 4 percent of the county.⁹

The case was eventually taken up by the civil rights law firm Relman, Dane & Colfax, based in Washington, D.C. In 2008, after a three-year trial, a federal jury returned verdicts totaling nearly \$11 million against the City of Zanesville. This case study examines the innovative use of public data that went into building the successful case and, in the process, addressing a decades-old civil rights violation.

⁶ GovLab interview with Tara Ramchandani, Attorney, Relman, Dane, & Colfax PLLC, August 3, 2015.

⁷ Johnson, Dirk. “For a Recently Plumbed Neighborhood, Validation in a Verdict.” *The New York Times*, August 11, 2008. http://www.nytimes.com/2008/08/12/us/12ohio.html?_r=0.

⁸ “Kennedy v. City of Zanesville.” Relman, Dane & Colfax PLLC. <http://www.relmanlaw.com/civil-rights-litigation/cases/zanesville.php>

⁹ <http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk>

Case Description and Inception

To determine whether there was an association between race and access to public water services in the Coal Run neighborhood, the plaintiff's lead attorneys, John Relman and Reed Colfax, obtained the demography services of Dr. Allan Parnell of the Cedar Grove Institute for Sustainable Communities. The Cedar Grove Institute is a nonprofit in Mebane, North Carolina that provides technical assistance, analysis and training to help community groups promote equitable community development.¹⁰ It grew out of the for-profit company, McMillan and Moss Research, Inc., whose "research and analyses were being called for in cases involving civil rights, predatory lending, segregation in schools, institutionalized discrimination and community economic development."¹¹ Mr. Colfax reached out to Dr. Parnell based on the recommendation of Jennifer Klar, now a partner at Relman, Dane & Colfax who met Dr. Parnell during a conference, and based on the Cedar Grove Institute's reputation for work in civil rights cases and community development. Ms. Klar, as well as pro bono attorneys from the law firm Jones Day and other organizations, also worked with Mr. Relman and Mr. Colfax on the case.¹²

As civil rights attorneys, Relman and Colfax were well aware of how public data can help provide critical evidence in cases, but lacked the technical know-how to analyze the data themselves. Parnell, a well-known public data specialist, regularly serves as an expert witness in civil rights cases using open data. For example, he was one of the plaintiff experts in *Texas Department of Housing and Community Affairs v. Inclusive Communities Project*, the 2015 Supreme Court decision that affirmed the validity of disparate impact cases where public data are key.¹³ Parnell therefore led the data research and analysis for Relman and Colfax, eventually settling on a strategy of combining data from multiple sources (public GIS data, water bill-

10 <http://www.cedargroveinst.org/mission.php>

11 <http://www.cedargroveinst.org/mission.php>

12 GovLab interview with Allan Parnell, Ph.D., Vice President, Cedar Grove Institute for Sustainable Communities, October 2, 2015.

13 GovLab interview with Allan Parnell, Ph.D., Vice President, Cedar Grove Institute for Sustainable Communities, July 21, 2015.

ing data and demographic data) to create maps that established a clear pattern of racial discrimination.¹⁴,¹⁵

Relman and Colfax agreed with Parnell's strategy in using public data. Once the decision was taken, however, Parnell soon realized that census data would not be effective due to the small size of the neighborhood in question and the distribution of residents within the blocks. For example, within each block, the northern part was typically predominantly white and the southern predominantly non-white.

Instead of using census data, Parnell proposed using publicly available geographic information systems (GIS) data from Muskingum County to perform a house-by-house analysis within the neighborhood.¹⁶ Although not available via an open data portal, GIS data is typically available by request, and Parnell and Reed successfully obtained the needed GIS data through a direct request to Muskingum County officials, and it was provided in a standard, machine-readable format. GIS data uses spatio-temporal location as the key index variable. Parnell explained that for most municipalities, one needs to first fill out a form requesting access to such GIS data, but that there is a "wealth of data available if you know how or where to ask for it."¹⁷ Essentially, the nature of GIS data allows users to analyze and interpret data in ways that make it easier to identify, manipulate and understand relationships, patterns and trends, and then visualize that data in forms that are accessible for anyone (data expert or not) to understand and share (e.g., maps, globes, reports and charts).¹⁸ Parnell, being experienced with GIS data, recognizes the opportunities GIS data presents more readily than those with

¹⁴ GovLab interview with Allan Parnell, Ph.D., Vice President, Cedar Grove Institute for Sustainable Communities, July 21, 2015.

¹⁵ Parnell, Allan M. "Maps Used in Support of the Plaintiff's Argument in Kennedy et al. v. City of Zanesville, et al." Legal Services of Northern California Race Equity Project. August 6, 2008. <http://equity.lsnc.net/2008/08/maps-used-in-support-of-the-plaintiff-%E2%80%99s-arguement-in-kennedy-et-al-v-city-of-zanesville-et-al/>

¹⁶ Parnell, Allan M. "Maps Used in Support of the Plaintiff's Argument in Kennedy et al. v. City of Zanesville, et al." Legal Services of Northern California Race Equity Project. August 6, 2008. <http://equity.lsnc.net/2008/08/maps-used-in-support-of-the-plaintiff-%E2%80%99s-arguement-in-kennedy-et-al-v-city-of-zanesville-et-al/>

¹⁷ GovLab interview with Allan Parnell, Ph.D., Vice President, Cedar Grove Institute for Sustainable Communities, July 21, 2015.

¹⁸ "What is GIS?" ESRI. <http://www.esri.com/what-is-gis/howgisworks>

less experience, such as attorneys. However, should the open data movement continue to grow, more people of all backgrounds could recognize and use GIS data.

Parcel data—which identified all occupied houses in the study areas, the location of water lines with dates of construction, Zanesville’s city limits and the street locations—provided the backbone of the case. Additionally, Relman, Dane & Colfax obtained water-billing data, which provided the addresses of all houses with public water service. With this data in hand, Parnell’s team of paralegals undertook a door-to-door effort to: a) confirm that property identified in the parcel data was an occupied house; and b) determine racial composition and how long each resident had lived in that location in order to determine that “there was no difference between the people with and without water other than race.”¹⁹

“The easier it is to access data, and the more people that can access data without having to pay for it, the more egalitarian society we will have.”

—Tara Ramchandani, Relman, Dane, & Colfax PLLC

Using the public GIS data, the household survey information, the plaintiff information and the addresses of houses with billed water service, Parnell’s colleague, Ben Marsh, Ph.D., Professor of Geography and Environmental Studies at Bucknell University, built the GIS layers for the maps showing a clear pattern of racial discrimination. Parnell wrote the expert report used in the case based on the maps, survey information and additional information taken from the plaintiffs. During the trial, Relman and Colfax walked the jury through the information contained in the maps by rebuilding Parnell’s maps, “layer by layer,” while explaining how each piece of information was obtained and what it uncovered about water access discrimination.²⁰

¹⁹ GovLab interview with Allan Parnell, Ph.D., Vice President, Cedar Grove Institute for Sustainable Communities, July 21, 2015.

²⁰ Parnell, Allan M. “Maps Used in Support of the Plaintiff’s Argument in Kennedy et al. v. City of Zanesville, et al.” Legal Services of Northern California Race Equity Project. August 6, 2008. <http://equity.lsnc.net/2008/08/maps-used-in-support-of-the-plaintiff-%E2%80%99s-arguement-in-kennedy-et-al-v-city-of-zanesville-et-al/>

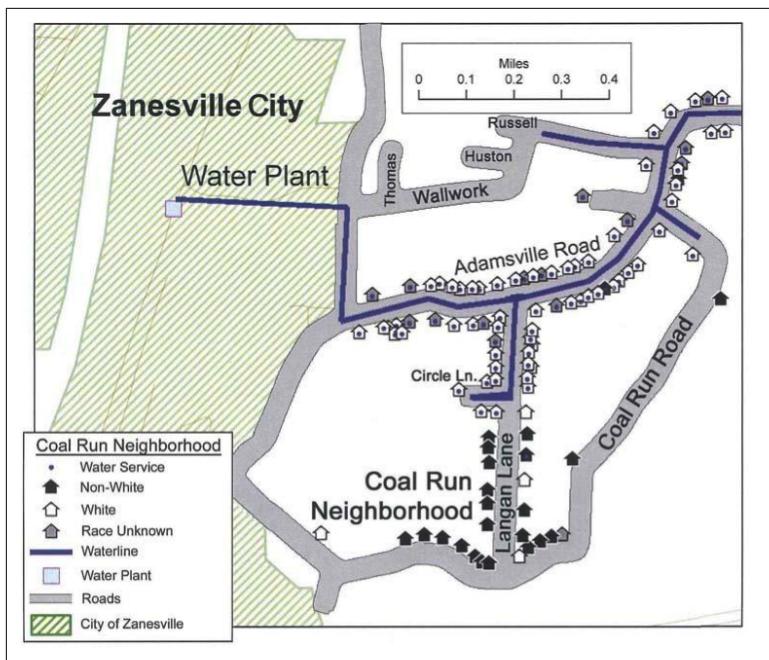


Figure 20-2. Map of Water Line Extensions in Muskingum County with GIS layers²¹. Copyright Allan Parnell. Used with permission.

Meanwhile, the expert testifying on behalf of the defense attorneys representing the City of Zanesville, Muskingum County and the East Muskingum Water Authority also attempted to use data-driven maps to support the opposition's case. The defense's expert, however, did not effectively manipulate the data or maps, leading to a mismatch between the claims made by the attorneys and the information being displayed. The defense used the GIS and census data to try and argue that race did not affect who had water service, claiming that all residents of a certain census block had water if a water line intersected any part of that census block. This is demonstrably false, and the defense was unable to prove the claim. Parnell countered the assertion using the same census data, showing that in 2000, the water lines for the block in question only served 34 African-

²¹ Parnell, Allan M. "Maps Used in Support of the Plaintiff's Argument in Kennedy et al. v. City of Zanesville, et al." Legal Services of Northern California Race Equity Project. August 6, 2008. <http://equity.lsnc.net/2008/08/maps-used-in-support-of-the-plaintiff-%E2%80%99s-arguement-in-kennedy-et-al-v-city-of-zanesville-et-al/>

American residents—all of whom lived in a nursing home whose population was 88 percent white.²²

The use of open data, in some cases drawn from the same source, to make contradictory points from two sides of the same court case demonstrates the danger of selective, perhaps manipulative use of data to lead people to make incorrect conclusions. The defense was unable to use the data to create a compelling case, however, and, as a result, appeared “clueless.”²³

Impact

The impact of using open data in the Kennedy v. Zanesville case can broadly be split into two categories: immediate, and long term.

The *immediate impact* was clear, tangible and in many respects highly positive. The use of open data (along with maps) was at the heart of a legal strategy that identified and remedied a long-standing civil rights violation. As Parnell put it: Some black residents were unaware that “for 50 years their [white] neighbor had a hot tub [while] they couldn’t turn the tap on.”²⁴ Open data helped visualize and irrefutably identify a systematic form of discrimination that had long been woven into the fabric of daily life in Zanesville.

Significant monetary damages were also assessed:

- The federal jury awarded \$11 million against the City of Zanesville and the East Muskingum Water Authority for illegally denying water on the basis of race.
- The jury also awarded \$80,000 in damages to Fair Housing Advocates Association, the agency that initially assisted the

²² Parnell, Allan M. “Maps Used in Support of the Plaintiff’s Argument in Kennedy et al. v. City of Zanesville, et al.” Legal Services of Northern California Race Equity Project. August 6, 2008. <http://equity.lsncc.org/2008/08/maps-used-in-support-of-the-plaintiff-v-kennedy-et-al-v-city-of-zanesville-et-al/>

²³ GovLab interview with Allan Parnell, Ph.D., Vice President, Cedar Grove Institute for Sustainable Communities, July 21, 2015.

²⁴ GovLab interview with Allan Parnell, Ph.D., Vice President, Cedar Grove Institute for Sustainable Communities, July 21, 2015.

plaintiffs with their administrative complaints before the Ohio Civil Rights Commission.²⁵

- All told, the plaintiffs were eligible for payments of between \$15,000 and \$300,000.²⁶

The **medium and longer-term impact**, while in many ways positive, is somewhat less clear. On the one hand, the use of data and maps quantified and solidified the case made by the plaintiffs that their neighborhood had long suffered discrimination. Although the city of Zanesville had completed laying water pipes before the jury reached its verdict, the data-driven case established by the legal team validated the original complaint, and potentially made it harder for municipal authorities to scale back their expanded water distribution or deny water services to other neighborhoods, as communities or attorneys could reference this case in the future.

In addition, in July 2015, The U.S. Department of Housing and Urban Development (HUD) announced new rules to the Fair Housing Act of 1968, which barred outright racial discrimination, then routine, and required active desegregation in housing.²⁷ The new rules require “cities and towns all over the country to scrutinize their housing patterns for racial bias and to publicly report, every three to five years, the results. Communities will also have to set goals, which will be tracked over time, for how they will further reduce segregation.”²⁸ While not a direct result of this case, the new HUD rules will create even more open data that can be applied to civil rights and fair housing cases. Law firms engaging in civil rights litigation often rely on GIS or similar open data sets for evidence in a variety of cases, including fair housing, school segregation, redevelopment and relocation, code enforcement abuse and unequal service provision.²⁹

25 “Kennedy v. City of Zanesville.” Relman, Dane & Colfax PLLC. <http://www.relmanlaw.com/civil-rights-litigation/cases/zanesville.php>

26 “Kennedy v. City of Zanesville.” Relman, Dane & Colfax PLLC. <http://www.relmanlaw.com/civil-rights-litigation/cases/zanesville.php>

27 http://www.huduser.org/portal/affht_pt.html#final-rule

28 Badger, Emily. “Obama administration to unveil major new rules targeting segregation across U.S.” *The Washington Post*. July 8, 2015. <http://www.washingtonpost.com/news/wonkblog/wp/2015/07/08/obama-administration-to-unveil-major-new-rules-targeting-segregation-across-u-s/>

29 “Expert Witness Work.” Cedar Grove Institute for Sustainable Communities. <http://www.cedargroveinst.org/partners.php>

Data helps illuminate the local political geography, providing “a link to a diversity of policy decision information, such as the relationship of race to the proximity of superfund sites.”³⁰ In illegal discriminatory lending cases, open data enables attorneys to determine where loans from the bank are going, and whether loans are given to minority neighborhoods or not.³¹ Regarding the new HUD rules, civil rights attorneys representing fair housing cases will have even more data to identify and prove patterns of discrimination, with civil-rights groups lauding the decision as “an important advancement on what’s been one of the most fraught frontiers of racial progress.”³²

Yet despite such advances, research conducted by Parnell (and others) has, somewhat paradoxically, concluded that the overall cause of open data has been little advanced in the community and among Muskingum County officials. In fact, perhaps because the case cost the county \$11 million and significant negative publicity, its aftermath has actually been marked in many cases to a tightening of data availability and supply. For example, Parnell found that public GIS data from Muskingum County, which although not online, had previously been available through a simple phone or form request, became more challenging to locate and access, requiring more phone calls and forms, and, often, the skills of an attorney or someone familiar with navigating the bureaucracy.³³ Parnell has also experienced similar difficulties in obtaining water and sewer information in certain California and North Carolina municipalities, with officials mandating that only necessary third parties, such as engineering companies, may access such data. The process for obtaining GIS data greatly varies across governments—by location and by level—although Parnell observed that this increase in steps to access GIS data became even more apparent after the attacks on

³⁰ GovLab interview with Allan Parnell, Ph.D., Vice President, Cedar Grove Institute for Sustainable Communities, July 21, 2015.

³¹ GovLab interview with Tara Ramchandani, Attorney, Relman, Dane, & Colfax PLLC, August 3, 2015.

³² Badger, Emily. “Obama administration to unveil major new rules targeting segregation across U.S.” *The Washington Post*. July 8, 2015. <http://www.washingtonpost.com/news/wonkblog/wp/2015/07/08/obama-administration-to-unveil-major-new-rules-targeting-segregation-across-u-s/>

³³ GovLab interview with Allan Parnell, Ph.D., Vice President, Cedar Grove Institute for Sustainable Communities, October 2, 2015.

September 11, 2001, as data custodians had more leeway in citing security concerns to delay access.³⁴ Unfortunately, Parnell believes that uncovering the true motives behind these actions, if they are indeed to discourage lawsuits and/or hide potential wrong-doing, would prove extremely time- and resource-intensive.³⁵

In several cases, these restrictions on data were justified in the name of security. For example, county officials said that making infrastructure data more fully open, for example by publishing it online in a downloadable format, posed a potential security threat that would allow terrorists to locate targets such as water or energy plants. However, Parnell argues that since that type of location information is not particularly difficult to find without access to GIS data, should one be motivated, these policy changes may need to be re-evaluated if the result is less transparency and decreased access to public data.³⁶

Challenges

These less-than-optimal outcomes of the Kennedy v. Zanesville case point to some recurring obstacles faced by open data advocates. In particular, the deployment of security as a justification for restricting access to information is fairly common, as several other case studies in this series indicate.

Broadly, the Zanesville case suggests three challenges to the wider release and dissemination of open data:

Security Concerns

After the incidents of September 11, 2001, concerns over security are frequently raised by custodians of data. These can express themselves as anxieties over national security, data security or other forms of security. These concerns are, of course, often masks for other reasons (for example, a desire to restrict negative publicity or

³⁴ GovLab interview with Allan Parnell, Ph.D., Vice President, Cedar Grove Institute for Sustainable Communities, October 2, 2015.

³⁵ GovLab interview with Allan Parnell, Ph.D., Vice President, Cedar Grove Institute for Sustainable Communities, July 21, 2015.

³⁶ GovLab interview with Allan Parnell, Ph.D., Vice President, Cedar Grove Institute for Sustainable Communities, July 21, 2015.

avoid lawsuits). As Parnell, who has dealt with officials across the spectrum of transparency, in counties across the country, explains, “You’re either going to hide your data or you’re going to fix things.”

Nonetheless, putting aside the validity of security as a justification for restricting access, there are steps open data advocates can take to mitigate such concerns.

Awareness and Usability

Open data offers a powerful way to combat discrimination (and various other injustices), but like all technologies, it is only a tool, its potential defined by the extent to which it is usable and actually used. Repeatedly, we see instances where data is made available but lies under-utilized due to a lack of awareness or barriers to usability. A similar pattern was very much evident in Zanesville, where the data finally deployed to such great success in the lawsuit had in fact been available for years but residents were unable to use it.

As Tara Ramchandani, an attorney at Relman, Dane & Colfax, explains:

How do you know if you’re being denied water service if you have to know to get the data, and then hire a lawyer to actually get and use the data? That makes it unfair. The more information that’s easily accessible [in simplified formats], the more easily you can understand what’s happening to you, and you can put your experience in context of the population around you.³⁷

She points out that even attorneys, who may use this type of data more regularly than the average citizen, often have to rely on experts in order to identify which data sets are useful, and how to access them. In addition, open data is often most useful when combined with other forms of closed or proprietary data (in this case, for example, the free and open public GIS data was combined with a Zanesville-specific, door-to-door survey conducted by Parnell’s team to confirm the race of residents in the neighborhood). The sophisticated technical and other skills required to access and combine data are quite often out of reach for ordinary citizens. As Ramchandani puts it: “The easier it is to access data, and the more

³⁷ GovLab interview with Tara Ramchandani, Attorney, Relman, Dane, & Colfax PLLC, August 3, 2015.

people that can access data without having to pay for it, the more egalitarian society we will have.”³⁸

Advocates of open data therefore need to first increase awareness of open data, as “it never crosses most people’s minds that this information is out there.”³⁹ Civil rights groups and law firms in particular should be targeted in outreach campaigns, and provided resources to learn how to access, use or find experts in open data to help support their cases.

Data—the Human Factor

Open data is a tool.⁴⁰ Its true potential derives from the way it is used by humans. Data unquestionably played a key role in winning the case, but one of the plaintiffs’ attorneys from Relman, Dane & Colfax noted that testimony given at trial was an equally, if not more, important factor in winning the case than the data and mapping evidence. For example, plaintiffs describing in their own words the experiences of discrimination, the hardships of not having access to water and the disappointment in repeatedly asking for water and being denied, painted a very moving, emotional story for the jury to consider. Meanwhile, the testimony of the defendants and all the people involved in deciding to deny people water, and the investigative process of discovering discrimination patterns in decision-makers’ behavior, also proved to be a powerful narrative in the courtroom, helping the jury to more fully understand and relate to the experiences of the community.⁴¹

Therefore, while the data and maps proved to be critical, hard evidence in the plaintiffs’ case, the more visceral aspects of the case help describe the real-world impact.⁴² For example, many Coal Run residents, including Doretta Hale, 74, wept the first time clean water

³⁸ GovLab interview with Tara Ramchandani, Attorney, Relman, Dane, & Colfax PLLC, August 3, 2015.

³⁹ GovLab interview with Allan Parnell, Ph.D., Vice President, Cedar Grove Institute for Sustainable Communities, July 21, 2015.

⁴⁰ GovLab interview with Allan Parnell, Ph.D., Vice President, Cedar Grove Institute for Sustainable Communities, July 21, 2015.

⁴¹ GovLab interview with Tara Ramchandani, Attorney, Relman, Dane, & Colfax PLLC, August 3, 2015.

⁴² GovLab interview with Tara Ramchandani, Attorney, Relman, Dane, & Colfax PLLC, August 3, 2015.

came through her pipes, describing, “I could wash clothes whenever I wanted ... I could go out and water the flowers.”⁴³

While not a challenge experienced by the plaintiffs in this case, it is clear that, especially in situations involving advocacy or persuasion, cold, hard data can only go so far. Personal experiences can help exemplify and ground the insights uncovered through open data, perhaps making takeaways easier to understand and relate to.

Looking Forward

Although Muskingum County seems to have pushed back against open data in some manner at the local level as a result of this case, water services are now supplied to residents equally. More broadly, the open data movement in Ohio on the state level does seem to be growing. The Office of the Ohio Treasurer, for example, has launched OhioCheckbook.com, an interactive online tool allowing users to search and access state spending data, as part of its transparency program. As of late 2015, the state House of Representatives was also developing a bill to launch DataOhio, an initiative that promotes open data standards and transparency in a number of ways. If enacted, state and local agencies in Ohio will be required to adhere to an open data standard. Similar to the federal open data portal and portals established in other states, DataOhio would establish an online catalog (data.ohio.gov) to provide descriptions of data sets, tutorials and tools. To help provide the financial backing for spurring open data activity, DataOhio also calls for the disbursal of \$10,000 in grants to local governments as an incentive to provide budgetary, staffing and compensation information online in an open data format using uniform accounting.⁴⁴ The city of Cincinnati, meanwhile, has its own open data portal to “provide access to government data, improve services, increase accountability and stimulate economic activity.”⁴⁵

⁴³ Johnson, Dirk. “For a Recently Plumbed Neighborhood, Validation in a Verdict.” *The New York Times*, August 11, 2008. http://www.nytimes.com/2008/08/12/us/12ohio.html?_r=0.

⁴⁴ “Society boosts DataOhio legislation.” The Ohio Society of CPAs. May 20, 2015. <http://www.ohiocpa.com/utilities/displaynewsitem/2015/05/21/society-boosts-dataohio-legislation>

⁴⁵ <https://data.cincinnati-oh.gov/>

As more state-wide and large-city open data initiatives are implemented across Ohio, there is potential for trickle-down effects. Government officials in smaller localities like Zanesville could be encouraged to embrace open data and help boost awareness among residents of the broad potential impacts of making government data more publicly accessible.

APPENDIX A

Appendix

Open Data Definitions—What's in a Name?

Open Definition (referenced by Open Data Handbook, ODI, Open Data Census, and OECD Open Data Analytical Framework)

“Open data is data that can be freely used, reused, and redistributed by anyone—subject only, at most, to the requirement to attribute and share alike.”

The White House, 2013 OMB Memorandum

“Open data refers to publicly available data structured in a way that enables the data to be fully discoverable and usable by end users.”

Data.Gov.UK

“Open data is data that is published in an open format, is machine readable, and is published under a license that allows for free reuse.”

Dbpedia: A nucleus for a web of open data

“Open data is the idea that certain data should be freely available to everyone to use and republish as they wish, without restrictions from copyright, patents, or other mechanisms of control.”

Open Data Institute

“Open data is information that is available for anyone to use, for any purpose, at no cost. Open data has to have a license that says it is open data. Without a license, the data can't be reused.

These principles for open data are described in detail in the Open Definition.”

LinkedGov

“Open data is non-personally identifiable data produced in the course of an organization’s ordinary business, which has been released under an unrestricted license. Open public data is underpinned by the philosophy that data generated or collected by organizations in the public sector should belong to the tax-payers, wherever financially feasible and where releasing it won’t violate any laws or rights to privacy (either for citizens or government staff).”

McKinsey Global Institute

“Machine-readable information, particularly government data, that’s made available to others. These open data sets share the following four characteristics: 1) Accessibility: a wide range of users is permitted to access the data. 2) Machine readability: the data can be processed automatically. 3) Cost: data can be accessed free or at negligible cost. 4) Rights: limitations on the use, transformation, and distribution of data are minimal.”

Open Data Now

“Open data is accessible public data we can use to launch new ventures, analyze trends, make decisions, and solve problems.”

Open Data Barometer

Excerpt from report indicates that researchers assessed data sets based on the “full Open Definition requirements of being machine readable, accessible in bulk, and openly licensed.”

The World Bank

“Data is open if it satisfies both of the following conditions:

- Technically open: available in a machine-readable standard format, which means it can be retrieved and meaningfully processed by a computer application
- Legally open: explicitly licensed in a way that permits commercial and non-commercial use and re-use without restrictions.”

Figure A-1. Open data definitions: matrix of analysis

	Free	Negligible Cost	Publicly Available	Re-usable	Can be Redistributed	Non-exclusive [No Restrictions from copyright, patents, etc.]	Structured for Usability	Requires "Open" License	Non-Personally Identifiable	Produced during business operation	Belongs to the Taxpayer [when not in violation of laws/privacy]	Accessible in Bulk
Open Definition <i>Open Knowledge Foundation</i>	<											
OMB Memo, 2013 <i>The White House Sylvia Burwell et al.</i>	<											
Data.Gov.UK <i>Antonio Acuña</i>												
"DBpedia: A Nucleus for a Web of Open Data" <i>Sören Auer et al.</i>	<											
Open Data Institute (ODI) <i>Open Data Institute</i>												
LinkedGov <i>LinkedGov</i>												
McKinsey <i>James Manyika et al.</i>												
Open Data Now <i>Joel Gurin</i>												
Open Data Barometer <i>Tim Davies</i>												
The World Bank <i>The World Bank</i>												

About the Authors

Andrew Young is the Associate Director of Research at [The GovLab](#), where he leads a number of grant-funded research efforts focusing on the impact of technology on public institutions. He has written extended work on how public sector institutions use new technology to coordinate work and provides research and writing support to all members of GovLab's team and to its extended network of participants in GovLab's training programs. He is also the Network Coordinator of the GovLab-chaired MacArthur Foundation Research Network on Opening Governance. In this role, he plans and organizes collaborative research projects and events with the Network's members, post-docs, and advisory group who span a dozen disciplines and institutions. Andrew earned his Master's degree in the Media, Culture and Communication department of NYU's Steinhardt School of Culture, Education and Human Development, with a focus on Technology and Society.

Stefaan G. Verhulst is the cofounder and Chief R&D of the [GovLab at Tandon-New York University](#) where he is responsible for experimentation and evidence gathering on how to transform governance using advances in science and technology. Before joining NYU full time, Verhulst spent more than a decade as Chief of Research for the Markle Foundation, where he continues to serve as Senior Advisor. He is also affiliated with the Department of Culture and Communications at New York University, the Central European University in Budapest; and the Annenberg School for Communications (Penn University). Previously at Oxford University he cofounded and was the head of the Programme in Comparative Media Law and Policy at the Centre for Socio Legal Studies, and also served as Senior Research Fellow of Wolfson College. He also taught several years at the London School of Economics. He has published widely and is a regular speaker on issues at the intersection of innovation, policy and governance.