A Field Guide to Open Data

Alicia Brown

2021-02-04

Contents

Chapter 1

Welcome!

Thank you for joining me on a journey to explore the use of Open Data in the civic world.

The book is organized into 3 parts -

- Part 1 provides an overview of what Open Data is, how it is used and who uses it.
- Part 2 questions the usefulness of Open Data to its producers, how use is measured and the challenges faced by producers and consumers of data.
- Part 3 imagines practices to increase the discovery and sharing of Open Data to deliver the promised insights and information for all.

I have had the great honor to work with many amazing organizations, agencies and governments to implement technology solutions that enabled the unleashing of data and analytics to power decision making and transparency.[1] The purpose for this book is to synthesize what I have learned working in the field with customers and present examples and ideas to grow the Open Data ecosystem. I also hope to provide context to the state of data and organizational roles so that there is an understanding of why data may be as it is and perspective on the concerns of the different groups involved in creating a successful program.

---

[1]Brown (2020)

Chapter 2

An Open Data Primer

2.1 What is Open Data?

"Open data is information or content made freely available to use and redistribute, subject only to the requirement to attribute it to the source."

— Gartner[1]

2.1.1 Topics

Popular data topics include emergency calls for service, budget, revenue, expenditures, property tax, permits, food inspections, government service calls, lost and found pets, crime incidents, jail bookings, voting registration and election results. When this data includes date, time and location, it can be valuable for performance management and informing staffing decisions. When it includes demographic information it makes it possible to analyze differences in data by gender, ethnicity and age and guide outreach and communication methods.

2.1.2 Databases

Data is commonly sourced from the transactional systems used to manage government services from relational databases. Spatial data of boundaries and locations of places exist in geodatabases as well as on open data platforms so that non-cartographers within governments and citizens may find where services exist on a map. While transactional data may have a geographic component like an address, it's no easy feat to join non-spatial data with spatial data. Date and time features in transactional data adds an extra twist as multiple years of data requires some consideration before joining to maps as boundaries and locations

---

[1]Gartner (2020)

may change over time or even completely disappear. While it may make sense to have versions of maps by year, the delivery of open data on web servers may be compromised by needing to load a new map as users toggle through different years of data. In my experience when presenting data with a map, the most recent spatial boundaries and point location are used for all years of data to ensure zippy user experience.

2.1.3 Files

Data is also found in spreadsheets and text files when ad-hoc or short term programs may be created to support some service or business process that may not warrant the involvement of information technology initially. There may not be budget to support the purchase or internal development of software. Department analysts, scientists and administrators may choose to spin up a solution with the business software or statistical programming language they are already experts at to enable the success of these programs and efficiently deliver needed services to internal audiences and the public. Negative consequences from these ventures may only arise when there is a need for reporting data or if files are owned by an individual who leaves their job without saving files to a shared network drive.

2.1.4 Reports

Departments may publish annual reports to share what they accomplished that year that contain tables and charts of the data they use to measure their performance or as a benchmark against peers or some standard. This is the perfect data to include in performance management which often is made up of annual measures. Reports may be constructed using text editors, Business Intelligence tools or programming languages where data, visualizatiosn and narrative are woven together before exporting as a pdf, document, slide deck or website.

2.1.5 Dashboards

Departments with a need for reporting data frequently may purchase Business Intelligence and Visualization tools that are powerful and intuitive. These tools connect to many data sources including spreadsheets, text files, web page, pdfs and even from copy/paste operations into a table. Depending on security policies, dashboards published publicly may enable the download of the data presented in it and occasionally be the sole source of some data.

2.2 How is it used?

2.2.1 Organizational Motivations

While government employees may have access to department level data, gaining access to data from other departments may be difficult, even when content is

not protected or confidential. By opening up high value departmental datasets, governments help break down silos internally and make it easy to share with the public. When an Open Data portal exists, it is also can be used to publish data requested from Public Information Requests especially for popular data and reduce the overhead required by manually filling these requests.

The software used by departments to manage their operations may not have extensive reporting features and require users to develop their own reports through queries to the underlying database or periodic exports of data files at the convenience of system administrators. Open Data platforms can be used as destination for automated extracts and provide users visualizations to quickly examine their data and gain operation insights. Viewing data on a map or over time makes it possible to fine tune staffing decisions. Datasets on Open Data can also be accessed programatically through APIs (Application Programming Interface) which enable users to use their favorite tools like Excel, Power BI, and Tableau with the ability to refresh data as needed.

Informational Technology teams can leverage Open Data to offset downloads from internal server infrastructure to handle the traffic caused by applications built by civic developers relying on transit, public safety and other large datasets.

2.2.2   Civic Concerns

Concerned citizens have used Open Data to dig deeper into news stories like when learning that tips for taxi drivers suddenly increased after a software update.[2] Open Payroll sites have made it easy to check the highest paid state employees and see where sports coaches rank in Connecticut.[3]

- Apps
- Grant applications
- Contests (Kaggle)

2.3   Who Use Open Data?

2.3.1   Top Cities

Of the top 20 cities in population as of the 2019 Census American Community Survey[4], 19 share city data through a portal.[5]

---

[2]Wellington (2020)

[3]Scinto (2019)

[4]Census (2019)

[5]Jacksonville does have a site with police datasets and maps available on GIS site. https://transparency.jaxsheriff.org/, https://maps.coj.net/jaxgis/

City	State	Population	Open Data Site
New York	New York	8,336,817	opendata.cityofnewyork.us
Los Angeles	California	3,979,537	data.lacity.org
Chicago	Illinois	2,693,959	data.cityofchicago.org
Houston	Texas	2,316,797	cohgis-mycity.opendata.arcgis.com
Phoenix	Arizona	1,680,988	phoenixopendata.com
Philadelphia	Pennsylvania	1,584,064	opendataphilly.org
San Antonio	Texas	1,547,250	data.sanantonio.gov
San Diego	California	1,423,852	data.sandiego.gov
Dallas city	Texas	1,343,565	dallasopendata.com
San Jose	California	1,021,786	data.sanjoseca.gov
Austin	Texas	979,263	data.austintexas.gov
Fort Worth	Texas	913,656	data.fortworthtexas.gov
Jacksonville	Florida	911,528	
Columbus	Ohio	902,073	opendata.columbus.gov
Charlotte	North Carolina	885,707	data.charlottenc.gov
San Francisco	California	881,549	datasf.org/opendata
Indianapolis	Indiana	870,340	data.indy.gov
Seattle	Washington	753,655	data.seattle.gov
Denver	Colorado	727,211	denvergov.org/opendata
Washington	District of Columbia	705,749	opendata.dc.gov

2.3.2 Roles

2.3.2.1 Government

- Executive sponsor - from IT or business unit
- Analysts, Scientists
- Program Managers
- Managers
- Equity officers
- Security
- Communications
- Information technology

2.3.2.2 Public

- Civic advocate
- Data journalists
- Residents
- Students

Chapter 3

Use of Open Data

3.1 Examples of Usefulness

- Outcomes

3.2 Measuring Use

- Utilization
- Analytics
- Feedback

1. Who are the real users of this data?
2. How can we tell who they are?
3. Are there personas we haven't imagined?
4. How can we measure actual engagement of these users?
5. Are they using it the way we thought they were? Ex. Hack to "fix" data in between steps.

3.3 Challenges of Open Data

3.3.1 Searching for data

- Paging through results
- Onus of filtering

3.3.2 Peeking at data

Chapter 4

Imagining Open Data

4.1 Sharing

Adding datasets to a platform may benefit power users of data and make it easy consumable by tools and programming languages. However data alone will not advance the knowledge of the community it is intended to serve without also including narrative and insights from the collections of datasets shared to a platform.

- Stories & Narrative
- Live tiles
- Data driven documentation
- Dashboards
- Goals
- Reports
- Info sheets

4.2 Accessing

- Tools
- Open source library
- APIs

4.3 Discovery

- Metadata
- Harvestable
- Standard schemas

4.4   Improving the Ecosystem

- Ownership and stewardship
- Automation
- Data refreshes
- Quality monitoring

Bibliography

Brown, A. (2020). Mapping 5 years of government data.

Census (2019). American community survey (acs).

Gartner (2020). Gartner glossary: Open data.

Scinto, R. (2019). See ct's highest paid state employees.

Wellington, B. (2020). How software in half of nyc cabs generates $5.2 million a year in extra tips.