Data for the Public Good

How Data Can Help Citizens and Government

Alex Howard

CLOUD MEETS
BIG DATA

EMC²

O'REILLY®

Strata
CONFERENCE
Making Data Work

HADOOP WORLD NOW AT STRATA

Join us at
Strata New York
Oct. 23 – 25, 2012

Strata Conference
Making Data Work
October 23 – 25, 2012
New York, NY

Strata Conference covers the latest and best tools and technologies for this new discipline, along the entire data supply chain—from gathering, analyzing, and storing data to communicating data intelligence effectively.

Strata Conference is for developers, data scientists, data analysts, and other data professionals.

The call for speakers is now open. Registration opens in early June.

strataconf.com/stratany2012

Data for the Public Good

Alex Howard

Data for the Public Good

by Alex Howard

Cover Designer: Karen Montgomery
Interior Designer: David Futato

February 2012:   First Edition.

Revision History for the First Edition:
    2012-02-17   First release

See *http://oreilly.com/catalog/errata.csp?isbn=9781449329761* for release details.

Table of Contents

Data for the Public Good

From Healthcare to Finance to Emergency Response, Data Holds Immense Potential to Help Citizens and Government

Can data save the world? Not on its own. As an age of technology-fueled transparency, open innovation and big data dawns around the world, the success of new policy won't depend on any single chief information officer, chief executive or brilliant developer. Data for the public good will be driven by a distributed community of media, nonprofits, academics and civic advocates focused on better outcomes, more informed communities and the new news, in whatever form it is delivered.

Advocates, watchdogs and government officials now have new tools for data journalism and open government. Globally, there's a wave of transparency that will wash over every industry and government, from finance to healthcare to crime.

In that context, open government is about much more than open data — just look at the issues that flow around the #opengov hashtag on Twitter, including the nature identity, privacy, security, procurement, culture, cloud computing, civic engagement, participatory democracy, corruption, civic entrepreneurship or transparency.

If we accept the premise that Gov 2.0 is a potent combination of open government, mobile, open data, social media, collective intelligence and connectivity, the lessons of the past year suggest that a tidal wave of technology-fueled change is still building worldwide.

The Economist's support for open government data remains salient today:

> "Public access to government figures is certain to release economic value and encourage entrepreneurship. That has already happened with weather data and with America's GPS satellite-navigation system that was opened for full com-

mercial use a decade ago. And many firms make a good living out of searching for or repackaging patent filings."

As Clive Thompson reported at Wired last year, public sector data can help fuel jobs, and "shoving more public data into the commons could kick-start billions in economic activity." In the transportation sector, for instance, transit data is open government fuel for economic growth.

There is a tremendous amount of work ahead in building upon the foundations that civil society has constructed over decades. If you want a deep look at what the work of digitizing data really looks like, read Carl Malamud's interview with Slashdot on opening government data.

Data for the public good, however, goes far beyond government's own actions. In many cases, it will happen despite government action — or, often, inaction — as civic developers, data scientists and clinicians pioneer better analysis, visualization and feedback loops.

For every civic startup or regulation, there's a backstory that often involves a broad number of stakeholders. Governments have to commit to open up themselves but will, in many cases, need external expertise or even funding to do so. Citizens, industry and developers have to show up to use the data, demonstrating that there's not only demand, but also skill outside of government to put open data to work in service accountability, citizen utility and economic opportunity. Galvanizing the co-creation of civic services, policies or apps isn't easy, but tapping the potential of the civic surplus has attracted the attention of governments around the world.

There are many challenges for that vision to pass. For one, data quality and access remain poor. Socrata's open data study identified progress, but also pointed to a clear need for improvement: Only 30% of developers surveyed said that government data was available, and of that, 50% of the data was unusable.

Open data will not be a silver bullet to all of society's ills, but an increasing number of states are assembling platforms and stimulating an app economy.

Results-oriented mayors like Rahm Emanuel and Mike Bloomberg are committing to opening Chicago and opening government data in New York City, respectively.

Following are examples of where data for the public good is already having an impact upon the world we live in, along with some ideas about what lies ahead.

Financial Good

Anyone looking for civic entrepreneurship will be hard pressed to find a better recent example than BrightScope. The efforts of Mike and Ryan Alfred are in line with traditional entrepreneurship: identifying an opportunity in a market that no one else has created value around, building a team to capitalize on it, and then investing years of hard work to execute on that vision. In the process, BrightScope has made government data about the financial industry more usable, searchable and open to the public.

Due to the efforts of these two entrepreneurs and their California-based startup, anyone who wants to learn more about financial advisers before tapping one to manage their assets can do so online.

Prior to BrightScope, the adviser data was locked up at the Securities and Exchange Commission (SEC) and the Financial Industry Regulatory Authority (FINRA).

"Ryan and I knew this data was there because we were advisers," said Bright-Scope co-founder Mike Alfred in a 2011 interview. "We knew data had been filed, but it wasn't clear what was being done with it. We'd never seen it liberated from the government databases."

While they knew the public data existed and had their idea years ago, Alfred said it didn't happen because they "weren't in the mindset of being data entrepreneurs" yet. "By going after 401(k) first, we could build the capacity to process large amounts of data," Alfred said. "We could take that data and present it on the web in a way that would be usable to the consumer."

Notably, the government data that BrightScope has gathered on financial advisers goes further than a given profile page. Over time, as search engines like Google and Bing index the information, the data has become searchable in places consumers are actually looking for it. That's aligned with one of the laws for open data that Tim O'Reilly has been sharing for years: Don't make people find data. Make data find the people.

As agencies adapt to new business relationships, consumers are starting to see increased access to government data. Now, more data that the nation's regulatory agencies collected on behalf of the public can be searched and understood by the public. Open data can improve lives, not least through adding more transparency into a financial sector that desperately needs more of it. This kind of data transparency will give the best financial advisers the advantage they deserve and make it much harder for your Aunt Betty to choose someone with a history of financial malpractice.

The next phase of financial data for good will use big data analysis and algorithmic consumer advice tools, or "choice engines," to make better decisions. The vast majority of consumers are unlikely to ever look directly at raw datasets themselves. Instead, they'll use mobile applications, search engines and social recommendations to make smarter choices.

There are already early examples of such services emerging. Billshrink, for example, lets consumers get personalized recommendations for a cheaper cell phone plan based on calling histories. Mint makes specific recommendations on how a citizen can save money based upon data analysis of the accounts added. Moreover, much of the innovation in this area is enabled by the ability of entrepreneurs and developers to go directly to data aggregation intermediaries like Yodlee or CashEdge to license the data.

Transit Data as Economic Fuel

Transit data continues to be one of the richest and most dynamic areas for co-creation of services. Around the United States and beyond, there has been a blossoming of innovation in the city transit sector, driven by the passion of citizens and fueled by the release of real-time transit data by city governments.

Francisca Rojas, research director at the Harvard Kennedy School's Transparency Policy Project, has investigated the dynamics behind the disclosure of data by transit agencies in the United States, which she calls one of the most successful implementations of open government. "In just a few years, a rich community has developed around this data, with visionary champions for disclosure inside transit agencies collaborating with eager software developers to deliver multiple ways for riders to access real-time information about transit," wrote Rojas.

The Massachusetts Bay Transit Authority (MBTA) learned from Portland, Oregon's, TriMet that open data is better. "This was the best thing the MBTA had done in its history," said Laurel Ruma, O'Reilly's director of talent and a long-time resident in greater Boston, in her 2010 Ignite talk on real-time transit data. The MBTA's move to make real-time data available and support it has spawned a new ecosystem of mobile applications, many of which are featured at MBTA.com.

There are now 44 different consumer-facing applications for the TriMet system. Chicago, Washington and New York City also have a growing ecosystem of applications.

As more sensors go online in smarter cities, tracking the movements of traffic patterns will enable public administrators to optimize routes, schedules and capacity, driving efficiency and a better allocation of resources.

Transparency and Civic Goods

As John Wonderlich, policy director at the Sunlight Foundation, observed last year, access to legislative data brings citizens closer to their representatives. "When developers and programmers have better access to the data of Congress, they can better build the databases and tools that let the rest of us connect with the legislature."

That's the promise of the Sunlight Foundation's work, in general: Technology-fueled transparency will help fight corruption, fraud and reveal the influence behind policies. That work is guided by data, generated, scraped and aggregated from government and regulatory bodies. The Sunlight Foundation has been focused on opening up Congress through technology since the organization was founded. Some of its efforts culminated recently with the publication of a live XML feed for the House floor and a transparency portal for House legislative documents.

There are other horizons for transparency through open government data, which broadly refers to public sector records that have been made available to citizens. For a canonical resource on what makes such releases truly "open," consult the "8 Principles of Open Government Data."

For instance, while gerrymandering has been part of American civic life since the birth of the republic, one of the best policy innovations of 2011 may offer hope for improving the redistricting process. DistrictBuilder, an open-source tool created by the Public Mapping Project, allows anyone to easily create legal districts.

"During the last year, thousands of members of the public have participated in online redistricting and have created hundreds of valid public plans," said Micah Altman, senior research scientist at Harvard University Institute for Quantitative Social Science, via an email last year.

"In substantial part, this is due to the project's effort and software. This year represents a huge increase in participation compared to previous rounds of redistricting — for example, the number of plans produced and shared by members of the public this year is roughly 100 times the number of plans submitted by the public in the last round of redistricting 10 years ago," Altman said. "Furthermore, the extensive news coverage has helped make a whole new set of people aware of the issue and has re framed it as a problem that citizens can actively participate in to solve, rather than simply complain about."

Principles for Data in the Public Good

As a result of digital technology, our collective public memory can now be shared and expanded upon daily. In a recent lecture on public data for public good at Code for America, Michal Migurski of Stamen Design made the point that part of the global financial crisis came through a crisis in public knowledge, citing "The Destruction of Economic Facts," by Hernando de Soto.

To arrive at virtuous feedback loops that amplify the signals that citizens, regulators, executives and elected leaders inundated with information need to make better decisions, data providers and infomediaries will need to embrace key principles, as Migurski's lecture outlined.

First, "data drives demand," wrote Tim O'Reilly, who attended the lecture and distilled Migurski's insights. "When Stamen launched crimespotting.org, it made people aware that the data existed. It was there, but until they put visualization front and center, it might as well not have been."

Second, "public demand drives better data," wrote O'Reilly. "Crimespotting led Oakland to improve their data publishing practices. The stability of the data and publishing on the web made it possible to have this data addressable with public links. There's an 'official version,' and that version is public, rather than hidden."

Third, "version control adds dimension to data," wrote O'Reilly. "Part of what matters so much when open source, the web, and open data meet government is that practices that developers take for granted become part of the way the public gets access to data. Rather than static snapshots, there's a sense that you can expect to move through time with the data."

The Case for Open Data

Accountability and transparency are important civic goods, but adopting open data requires grounded arguments for a city chief financial officer to support these initiatives. When it comes to making a business case for open data, John Tolva, the chief technology officer for Chicago, identified four areas that support the investment in open government:

1. Trust — "Open data can build or rebuild trust in the people we serve," Tolva said. "That pays dividends over time."

2. Accountability of the work force — "We've built a performance dashboard with KPIs [key performance indicators] that track where the city directly touches a resident."

3. Business building — "Weather apps, transit apps ... that's the easy stuff," he said. "Companies built on reading vital signs of the human body could be reading the vital signs of the city."
4. Urban analytics — "Brett [Goldstein] established probability curves for violent crime. Now we're trying to do that elsewhere, uncovering cost savings, intervention points, and efficiencies."

New York City is also using data internally. The city is doing things like applying predictive analytics to building code violations and housing data to try to understand where potential fire risks might exist. "The thing that's really exciting to me, better than internal data, of course, is open data," said New York City chief digital officer Rachel Sterne during her talk at Strata New York 2011. "This, I think, is where we really start to reach the potential of New York City becoming a platform like some of the bigger commercial platforms and open data platforms. How can New York City, with the enormous amount of data and resources we have, think of itself the same way Facebook has an API ecosystem or Twitter does? This can enable us to produce a more user-centric experience of government. It democratizes the exchange of information and services. If someone wants to do a better job than we are in communicating something, it's all out there. It empowers citizens to collaboratively create solutions. It's not just the consumption but the co-production of government services and democracy."

The Promise of Data Journalism

The ascendance of data journalism in media and government will continue to gather force in the years ahead.

Journalists and citizens are confronted by unprecedented amounts of data and an expanded number of news sources, including a social web populated by our friends, family and colleagues. Newsrooms, the traditional hosts for information gathering and dissemination, are now part of a flattened environment for news. Developments often break first on social networks, and that information is then curated by a combination of professionals and amateurs. News is then analyzed and synthesized into contextualized journalism.

Data is being scraped by journalists, generated from citizen reporting, or gleaned from massive information dumps — such as with the Guardian's formidable data journalism, as detailed in a recent ebook. ScraperWiki, a favorite tool of civic coders at Code for America and elsewhere, enables anyone to collect, store and publish public data. As we grapple with the consumption challenges presented by this deluge of data, new publishing platforms are also

empowering us to gather, refine, analyze and share data ourselves, turning it into information.

There are a growing number of data journalism efforts around the world, from New York Times interactive features to the award-winning investigative work of ProPublica. Here are just a few promising examples:

- Spending Stories, from the Open Knowledge Foundation, is designed to add context to news stories based upon government data by connecting stories to the data used.

- Poderopedia is trying to bring more transparency to Chile, using data visualizations that draw upon a database of editorial and crowdsourced data.

- The State Decoded is working to make the law more user-friendly.

- Public Laboratory is a tool kit and online community for grassroots data gathering and research that builds upon the success of Grassroots Mapping.

- Internews and its local partner Nai Mediawatch launched a new website that shows incidents of violence against journalists in Afghanistan.

Open Aid and Development

The World Bank has been taking unprecedented steps to make its data more open and usable to everyone. The data.worldbank.org website that launched in September 2010 was designed to make the bank's open data easier to use. In the months since, more than 100 applications have been built using the data.

"Up until very recently, there was almost no way to figure out where a development project was," said Aleem Walji, practice manager for innovation and technology at the World Bank Institute, in an interview last year. "That was true for all donors, including us. You could go into a data bank, find a project ID, download a 100-page document, and somewhere it might mention it. To look at it all on a country level was impossible. That's exactly the kind of organization-centric search that's possible now with extracted information on a map, mashed up with indicators. All of sudden, donors and recipients can both look at relationships."

Open data efforts are not limited to development. More data-driven transparency in aid spending is also going online. Last year, the United States Agency for International Development (USAID) launched a public engagement effort to raise awareness about the devastating famine in the Horn of Africa. The FWD campaign includes a combination of open data, mapping and citizen engagement.

"Frankly, it's the first foray the agency is taking into open government, open data, and citizen engagement online," said Haley Van Dyck, director of digital strategy at USAID, in an interview last year.

"We recognize there is a lot more to do on this front, but are happy to start moving the ball forward. This campaign is different than anything USAID has done in the past. It is based on informing, engaging, and connecting with the American people to partner with us on these dire but solvable problems. We want to change not only the way USAID communicates with the American public, but also the way we share information."

USAID built and embedded interactive maps on the FWD site. The agency created the maps with open source mapping tools and published the datasets it used to make these maps on data.gov. All are available to the public and media to download and embed as well.

The combination of publishing maps and the open data that drives them simultaneously online is significantly evolved for any government agency, and it serves as a worthy bar for other efforts in the future to meet. USAID accomplished this by migrating its data to an open, machine-readable format.

"In the past, we released our data in inaccessible formats — mostly PDFs — that are often unable to be used effectively," said Van Dyck. "USAID is one of the premiere data collectors in the international development space. We want to start making that data open, making that data sharable, and using that data to tell stories about the crisis and the work we are doing on the ground in an interactive way."

Crisis Data and Emergency Response

Unprecedented levels of connectivity now exist around the world. According to a 2011 survey from the Pew Internet and Life Project, more than 50% of American adults use social networks, 35% of American adults have smartphones, and 78% of American adults are connected to the Internet. When combined, those factors mean that we now see earthquake tweets spread faster than the seismic waves themselves. Networked publics can now share the effects of disasters in real time, providing officials with unprecedented insight into what's happening. Citizens act as sensors in the midst of the storm, creating an ad hoc system of networked accountability through data.

The growth of an Internet of Things is an important evolution. What we saw during Hurricane Irene in 2011 was the increasing importance of an Internet of people, where citizens act as sensors during an emergency. Emergency management practitioners and first responders have woken up to the potential of using social data for enhanced situational awareness and resource allocation.

An historic emergency social data summit in Washington in 2010 highlighted how relevant this area has become. And last year's hearing in the United States Senate on the role of social media in emergency management was "a turning point in Gov 2.0," said Brian Humphrey of the Los Angeles Fire Department.

The Red Cross has been at the forefront of using social data in a time of need. That's not entirely by choice, given that news of disasters has consistently broken first on Twitter. The challenge is for the men and women entrusted with coordinating response to identify signals in the noise.

First responders and crisis managers are using a growing suite of tools for gathering information and sharing crucial messages internally and with the public. Structured social data and geospatial mapping suggest one direction where these tools are evolving in the field.

A web application from ESRI deployed during historic floods in Australia demonstrated how crowdsourced social intelligence provided by Ushahidi can enable emergency social data to be integrated into crisis response in a meaningful way.

The Australian flooding web app includes the ability to toggle layers from OpenStreetMap, satellite imagery, and topography, and then filter by time or report type. By adding structured social data, the web app provides geospatial information system (GIS) operators with valuable situational awareness that goes beyond standard reporting, including the locations of property damage, roads affected, hazards, evacuations and power outages.

Long before the floods or the Red Cross joined Twitter, however, Brian Humphrey of the Los Angeles Fire Department (LAFD) was already online, listening. "The biggest gap directly involves response agencies and the Red Cross," said Humphrey, who currently serves as the LAFD's public affairs officer. "Through social media, we're trying to narrow that gap between response and recovery to offer real-time relief."

After the devastating 2010 earthquake in Haiti, the evolution of volunteers working collaboratively online also offered a glimpse into the potential of citizen-generated data. Crisis Commons has acted as a sort of "geeks without borders." Around the world, developers, GIS engineers, online media professionals and volunteers collaborated on information technology projects to support disaster relief for post-earthquake Haiti, mapping streets on OpenStreetMap and collecting crisis data on Ushahidi.

Healthcare

What happens when patients find out how good their doctors really are? That was the question that Harvard Medical School professor Dr. Atul Gawande asked in the New Yorker, nearly a decade ago.

The narrative he told in that essay makes the history of quality improvement in medicine compelling, connecting it to the creation of a data registry at the Cystic Fibrosis Foundation in the 1950s. As Gawande detailed, that data was privately held. After it became open, life expectancy for cystic fibrosis patients tripled.

In 2012, the new hope is in big data, where techniques for finding meaning in the huge amounts of unstructured data generated by healthcare diagnostics offer immense promise.

The trouble, say medical experts, is that data availability and quality remain significant pain points that are holding back existing programs.

There are, literally, bright spots that suggest what's possible. Dr. Gawande's 2011 essay, which considered whether "hotspotting" using health data could help lower medical costs by giving the neediest patients better care, offered another perspective on the issue. Early outcomes made the approach look compelling. As Dr. Gawande detailed, when a Medicare demonstration program offered medical institutions payments that financed the coordination of care for its most chronically expensive beneficiaries, hospital stays and trips to the emergency rooms dropped more than 15% over the course of three years. A test program adopting a similar approach in Atlantic City saw a 25% drop in costs.

Through sharing data and knowledge, and then creating a system to convert ideas into practice, clinicians in the ImproveCareNow network were able to improve the remission rate for Crohn's disease from 49% to 67% without the introduction of new drugs.

In Britain, researchers found that the outcomes for adult cardiac patients improved after the publication of information on death rates. With the release of meaningful new open government data about performance and outcomes from the British national healthcare system, similar improvements may be on the way.

"I do believe we are at the beginning of a revolutionary moment in health care, when patients *and* clinicians collect and share data, working together to create more effective health care systems," said Susannah Fox, associate director for digital strategy at the Pew Internet and Life Project, in an interview in January. Fox's research has documented the social life of health information, the con-

cept of peer-to-peer healthcare, and the role of the Internet among people living with chronic disease.

In the past few years, entrepreneurs, developers and government agencies have been collaboratively exploring the power of open data to improve health. In the United States, the open data story in healthcare is evolving quickly, from new mobile apps that lead to better health decisions to data spurring changes in care at the U.S. Department of Veterans Affairs.

Since he entered public service, Todd Park, the first chief technology officer of the U.S. Department of Health and Human Services (HHS), has focused on unleashing the power of open data to improve health. If you aren't familiar with this story, read the Atlantic's feature article that explores Park's efforts to revolutionize the healthcare industry through better use of data. Park has focused on releasing data at Health.Data.Gov. In a speech to a Hacks and Hackers meetup in New York City in 2011, Park emphasized that HHS wasn't just releasing new data: "[We're] also making existing data truly accessible or usable," he said, taking "stuff that's in a book or on a website and turning it into machine-readable data or an API."

Park said it's still quite early in the project and that the work isn't just about data — it's about how and where it's used. "Data by itself isn't useful. You don't go and download data and slather data on yourself and get healed," he said. "Data is useful when it's integrated with other stuff that does useful jobs for doctors, patients and consumers."

What Lies Ahead

There are four trends that warrant special attention as we look to the future of data for public good: civic network effects, hybridized data models, personal data ownership and smart disclosure.

Civic Network Effects

Community is a key ingredient in successful open government data initiatives. It's not enough to simply release data and hope that venture capitalists and developers magically become aware of the opportunity to put it to work. Marketing open government data is what repeatedly brought federal Chief Technology Officer Aneesh Chopra and Park out to Silicon Valley, New York City and other business and tech hubs.

Despite the addition of topical communities to Data.gov, conferences and new media efforts, government's attempts to act as an "impatient convener" can only go so far. Civic developer and startup communities are creating a new

distributed ecosystem that will help create that community, from BuzzData to Socrata to new efforts like Max Ogden's DataCouch.

Smart Disclosure

There are enormous economic and civic good opportunities in the "smart disclosure" of personal data, whereby a private company or government institution provides a person with access to his or her own data in open formats. Smart disclosure is defined by Cass Sunstein, Administrator of the White House Office for Information and Regulatory Affairs, as a process that "refers to the timely release of complex information and data in standardized, machine-readable formats in ways that enable consumers to make informed decisions."

For instance, the quarterly financial statements of the top public companies in the world are now available online through the Securities and Exchange Commission.

Why does it matter? The interactions of citizens with companies or government entities generate a huge amount of economically valuable data. If consumers and regulators had access to that data, they could tap it to make better choices about everything from finance to healthcare to real estate, much in the same way that web applications like Hipmunk and Zillow let consumers make more informed decisions.

Personal Data Assets

When a trend makes it to the World Economic Forum (WEF) in Davos, it's generally evidence that the trend is gathering steam. A report titled "Personal Data Ownership: The Emergence of a New Asset Class" suggests that 2012 will be the year when citizens start thinking more about data ownership, whether that data is generated by private companies or the public sector.

"Increasing the control that individuals have over the manner in which their personal data is collected, managed and shared will spur a host of new services and applications," wrote the paper's authors. "As some put it, personal data will be the new 'oil' — a valuable resource of the 21st century. It will emerge as a new asset class touching all aspects of society."

The idea of data as a currency is still in its infancy, as Strata Conference chair Edd Dumbill has emphasized. The Locker Project, which provides people with the ability to move their own data around, is one of many approaches.

The growth of the Quantified Self movement and online communities like PatientsLikeMe and 23andMe validates the strength of the movement. In the

U.S. federal government, the Blue Button initiative, which enables veterans to download personal health data, has now spread to all federal employees and earned adoption at Aetna and Kaiser Permanente.

In early 2012, a Green Button was launched to unleash energy data in the same way. Venture capitalist Fred Wilson called the Green Button an "OAuth for energy data."

Wilson wrote:

> "It is a simple standard that the utilities can implement on one side and web/ mobile developers can implement on the other side. And the result is a ton of information sharing about energy consumption and, in all likelihood, energy savings that result from more informed consumers."

Hybridized Public-Private Data

Free or low-cost online tools are empowering citizens to do more than donate money or blood: Now, they can donate, time, expertise or even act as sensors. In the United States, we saw a leading edge of this phenomenon in the Gulf of Mexico, where Oil Reporter, an open source oil spill reporting app, provided a prototype for data collection via smartphone. In Japan, an analogous effort called Safecast grew and matured in the wake of the nuclear disaster that resulted from a massive earthquake and subsequent tsunami in 2011.

Open source software and citizens acting as sensors have steadily been integrated into journalism over the past few years, most dramatically in the videos and pictures uploaded after the 2009 Iran election and during 2011's Arab Spring.

Citizen science looks like the next frontier. Safecast is combining open data collected by citizen science with academic, NGO and open government data (where available), and then making it widely available. It's similar to other projects, where public data and experimental data are percolating.

Public Data Is a Public Good

Despite the myriad challenges presented by legitimate concerns about privacy, security, intellectual property and liability, the promise of more informed citizens is significant. McKinsey's 2011 report dubbed big data as the next frontier for innovation, with billions of dollars of economic value yet to be created. When that innovation is applied on behalf of the public good, whether it's in city planning, transit, healthcare, government accountability or situational awareness, those effects will be extended.

We're entering the feedback economy, where dynamic feedback loops between customers and corporations, partners and providers, citizens and governments, or regulators and companies can both drive efficiencies and leaner, smarter governments.

The exabyte age will bring with it the twin challenges of information overload and overconsumption, both of which will require organizations of all sizes to use the emerging toolboxes for filtering, analysis and action. To create public good from public goods — the public sector data that governments collect, the private sector data that is being collected and the social data that we generate ourselves — we will need to collectively forge new compacts that honor existing laws and visionary agreements that enable the new data science to put the data to work.