

CHAPTER

5



ABOUT THE AUTHOR

Joel Gurin is a leading expert on Open Data—accessible public data that can drive new company development, business strategies, scientific innovation, and ventures for the public good. He is the author of the book *Open Data Now* and senior advisor at the Governance Lab at New York University, where he directs the Open Data 500 project. He previously served as chair of the White House Task Force on Smart Disclosure, as chief of the Consumer and Governmental Affairs Bureau of the U.S. Federal Communications Commission, and as editorial director and executive vice president of *Consumer Reports*.

DRIVING INNOVATION WITH OPEN DATA

BY JOEL GURIN



Key Takeaways

Open Data, like Big Data, is a major driver for innovation. Unlike privately held Big Data, Open Data can be used by anyone as a free public resource and can be used to start new businesses, gain business intelligence, and improve business processes.

While Open Data can come from many sources—including social media, private sector companies, and scientific research—the most extensive, widely used Open Data comes from government agencies and offices. Governments at all levels need to develop policies and processes to release relevant, accessible, and useful Open Data sources to enable innovation, foster a better-informed public, and create economic opportunity.

Hundreds of new companies have launched using Open Data, operating in all sectors of the economy and using a wide range of business models and revenue sources.

The chapters in this report provide ample evidence of the power of data and its business potential. But like any business resource, data is only valuable if the benefit of using it outweighs its cost. Data collection, management, distribution, quality control, and application all come at a price—a potential obstacle for companies of any size, though especially for small- and medium-sized enterprises.

Over the last several years, however, the “I” of data’s return on investment (ROI) has become less of a hurdle, and new data-driven companies are developing rapidly as a result. One major reason is that governments at the federal, state, and local level are making more data available at little or no charge for the private sector and the public to use. Governments collect data of all kinds—including scientific, demographic, and financial data—at taxpayer expense.

Now, public sector agencies and departments are increasingly repaying that public investment by making their data available to all for free or at a low cost. This is Open Data. While there are still costs in putting the data to use, the growing availability of this national resource is becoming a significant driver for hundreds of new businesses. This chapter describes the growing potential of Open Data and the data-driven innovation it supports, the types of data and applications that are most promising, and the policies that will encourage innovation going forward.

Market Opportunity in Data-Driven Innovation

Today’s unprecedented ability to gather, analyze, and use large amounts of data—both Open Data and privately held data—is creating qualitatively new kinds of business opportunities.

Data analysis can increase efficiency and reduce costs through what can be called *process innovation*. Logistics companies are using data to improve efficiency. UPS, for example, is using data analytics to determine the optimal routes for its drivers. A recent analysis by The Governance Lab (GovLab) at New York University, an academically based research organization, showed how data analysis can increase efficiency in healthcare, both for the National Health Service in the UK and potentially for other healthcare systems as well.¹ And as discussed throughout this report, data generated by the Internet of Things can reveal new correlations that lead to insight

and innovation. For example, data analysis of manufacturing processes can reveal opportunities to improve efficiency.

Companies can also use data to reach customers more effectively, improve brand reputation, or design products and services for specific audiences (i.e., *marketing innovation*). For example, Google analyzes search behavior to target advertising, and Amazon uses customer data to increase sales through personalized recommendations.

Lastly, data drives *business innovation and creation*. There are a growing number of companies throughout myriad industries (e.g., finance, healthcare, energy, education, etc.) that simply would not exist without today's data science. They are not using data to run their delivery network more efficiently or sell more books. They are using data to deliver entirely new products and services—to build businesses that are innovative from the ground up.

This chapter looks at this third kind of innovation, not because it will necessarily have the largest short-term economic impact but because of the impact it will have over time. These new data-driven companies will have a multiplier effect: once the first companies show the way, others will follow, leading to cumulative job growth and wealth creation. These companies rely on data as a business resource, and public government data is an especially cost-effective source for them. By making more data freely available, government agencies can make a critical difference in fostering this kind of business innovation.

While Big Data has attracted a lot of interest, Open Data may be more important for new business creation. As the diagram below shows, Big Data and Open Data are related but different concepts. Some Big Data is anything but open. Customer records held by businesses, for example, are meant to be used exclusively by the companies that collect it to improve their business processes and marketing. Open Data, in contrast, is designed for public use. It is a public good that supports and accelerates businesses across the economy, not just specific companies in specific sectors.

When Big Data is also Open Data, as is the case for much open government data, it is especially powerful. A recent McKinsey study estimated

the value of Open Data globally at more than \$3 trillion a year.² While that study covered several kinds of Open Data, government data and large government datasets make up a significant part of that calculation. National governments around the world, with the United States and the UK in the lead, are realizing that the data they collect in areas as diverse as agriculture, finance, and population dynamics can have tremendous business value. They are now working to make those datasets more widely available, more usable, and more relevant to business needs.

Beyond open government data, three other kinds of Open Data are driving innovation in important ways:

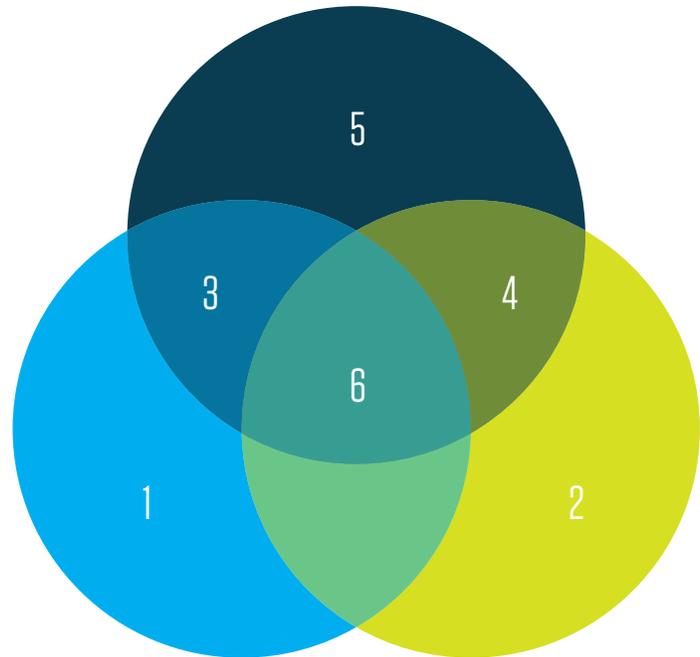
Scientific Data

The results of scientific research have often been closely guarded. Academic researchers hold on to their data until they can publish it, while private sector research is generally not shared until the company that supported it can patent the results. Now, however, scientists in academia and business alike are beginning to test a new model, one where they share data early on to accelerate the pace of everyone's research. This open-science approach was developed most notably in the Human Genome Project, funded by the U.S. government, which ran from 1990 to 2003. The scientists involved agreed to share data openly, and that approach accelerated their progress. Now, pharmaceutical companies are beginning to experiment with a similar model of data-sharing at an early research stage.

Social Media Data

Social media is a rich source of Open Data. Between review sites, blogs, and an average of 200 billion tweets sent each year,³ social media users are creating a huge resource of public data reflecting their opinions about consumer products, services, and brands. The evolving science of sentiment analysis uses text analytics and other approaches to synthesize those public data points into information that can be used for marketing, product development, and brand management. Companies like Gnip and Datasift have built their business on aggregating social media data and making it easy for other companies to study and analyze.

- 1. **NON-PUBLIC DATA**
for marketing, business analysis, national security
- 2. **CITIZEN ENGAGEMENT PROGRAMS** not based on data (e.g., petition websites)
- ● 3. **LARGE DATASETS**
from scientific research, social media, or other non-govt. sources
- ● 4. **PUBLIC DATA** from state, local, federal govt. (e.g., budget data)
- 5. **BUSINESS REPORTING**
(e.g., ESG data); other business data (e.g., consumer complaints)
- ● 6. **LARGE PUBLIC GOVERNMENT DATASETS** (e.g., weather, GPS, Census, SEC, healthcare)



Personal Data

No one is suggesting that personal data on health, finances, or other individual data should be publicly available. There is increasing interest, however, in making each person's individual data more available and open to him or her. New applications are helping people download their health records, tax forms, energy usage history, and more. The model is the Blue Button program that was originally developed to help veterans download their medical histories from the Veterans Administration. The private sector has now adapted it to provide medical records to about 150 million Americans.⁴ A similar program for personal energy usage data, the Green Button program, was developed through government collaboration with utilities.

Market Development — Opportunities for Using Public Data

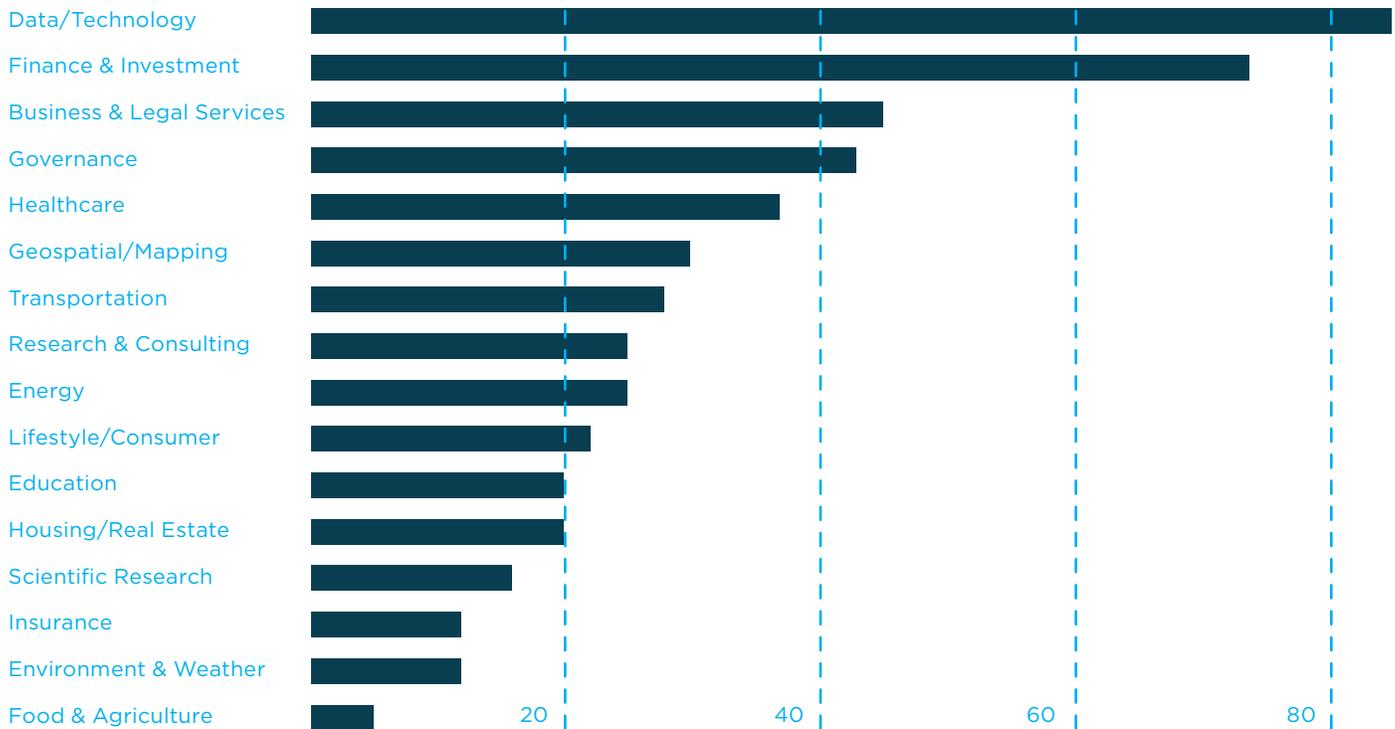
The United States and other national governments have committed themselves to making government data “open by default;” that is, to make it open to the public unless there are security, privacy, or other compelling reasons not to do so. But datasets won't open themselves, and it is not possible to make a country's entire supply of public data available overnight. Since it will take considerable time, money, and work to turn

national government datasets into usable Open Data, it is important to try to evaluate the ROI for this effort. Over the last few years, policy analysts have made several high-level attempts to estimate the economic value of these different kinds of data, Open Data in particular.

The aforementioned GovLab is studying the same issue in a more granular way. The GovLab now runs the Open Data 500 study, a project to find and study roughly 500 U.S.-based companies that use open government data as a key business resource.⁵ While the study has not yet collected systematic financial data on these companies, it has provided a basic map of the territory, showing the categories of companies that use open government data, which federal agencies they draw on as data suppliers, basic information about their business models, and what kinds of open government data have the greatest potential for use.

The Open Data 500 includes companies across business sectors. Several companies are built on two classic examples of open government data: weather data, first released in the 1970s, which has fueled companies like the Weather Channel; and GPS data, made available more recently, which is used by companies ranging from OnStar to Uber. But a look at companies started in the last

TYPES OF COMPANIES



10 years shows diverse uses of data from a wide range of government agencies. The table above offers about 100 examples, organized by business category. These are not meant to be a “best of” list, but rather, examples that show the types of applications in different sectors that are beginning to attract public attention and investor interest.

One striking development is the growing number of companies whose business is to make it easier for other businesses to use Open Data. Categorized as Data/Technology companies in the Open Data 500, they make up the largest single category in that study. These companies provide platforms and services that make open government data easier to find, understand, and use. One of the best examples is Enigma.io, a Manhattan-based company that gained visibility when it won the New York TechCrunch Disrupt competition in May 2013.⁶ Enigma provides a solution for the technical limitations of government datasets by putting their data onto a common, usable platform.

These companies serve a critical function in the Open Data ecosystem. Much government data is incomplete or inaccurate, managed through obsolescent legacy systems, or difficult to find. While many government agencies are working to

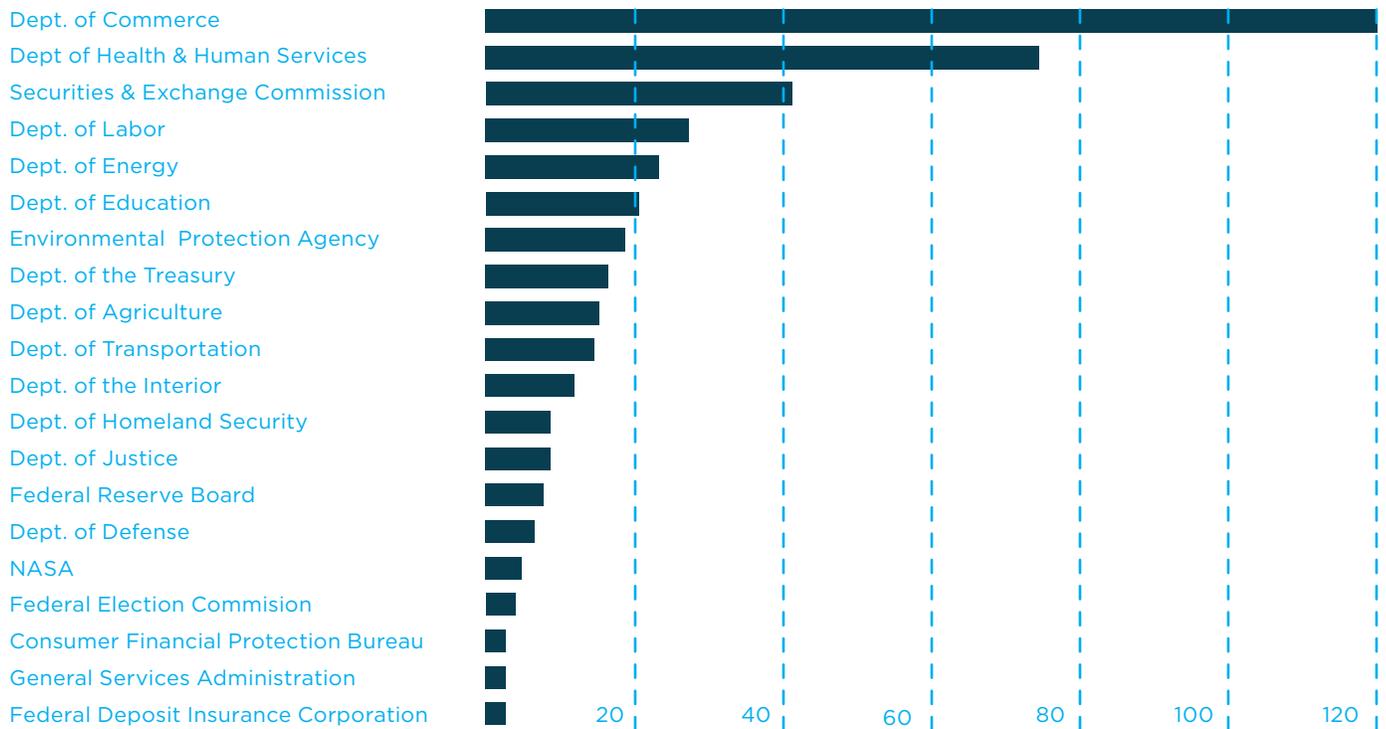
improve their data resources, it is a massive task and one that requires help from the private sector. Given the complexities of government datasets, the current state of much government data, and the lack of funding to improve it rapidly, companies that serve as data intermediaries will continue to have a viable business for years to come. They will also have a multiplier effect: their success will help make many other data-driven companies successful as well.

There are data-driven opportunities for businesses across all industries, with different kinds of Open Data serving as fuel for their innovative fires. Some of these most active sectors and the most important datasets include:

Business and Legal Services

A number of companies are managing, analyzing, and providing Open Data for business intelligence and business operations. Innography, for example, takes data from the U.S. Patent and Trademark Office and combines it with other data to provide analytic tools that businesses can use to learn about potential competitors and partners. In another area, Panjiva uses customs data to facilitate international trade, connecting buyers and suppliers across 190 countries.

NUMBER OF COMPANIES USING EACH AGENCIES' DATA



Education

Data-driven companies are finding value in two kinds of education data. The first is data on student performance, which can be opened to students, parents, and teachers to help tailor education to specific student needs. It is not yet clear how companies will be able to use this sensitive data to connect students with educational resources and programs without running afoul of privacy concerns. If they can, however, they will provide an important public benefit with significant economic value.

The second kind of education data is about the academic institutions themselves, and in particular, the value they offer. College-bound students have long-relied on college rankings from the likes of *U.S. News & World Report* and the *Princeton Review* to find a good college. They have used that information to work with their parents and their high school counselors to figure out whether and how they can afford the college of their choice. What many don't realize, however, is that signing up to go to any given college is like buying a new car—the sticker price is less meaningful than the price you can negotiate. Most colleges are now required to disclose their “true cost;” that is, the expected cost for a particular kind of student

after the college's typical financial aid package is taken into account. Several education websites are now using this kind of data to help students find colleges and perhaps find a school that is more affordable than they thought.

Energy

Growing interest in clean energy and sustainability is creating a new breed of data-driven companies. Several now use a combination of Open Data on energy efficiency with data-gathering sensors to help make residential and commercial buildings more energy efficient. Others are using Open Data to help advance clean energy technologies. Clean Power Finance, for example, uses Open Data to help solar-power professionals find access to financing. A new company, Solar Census, is aiming to make solar power more cost-effective by using geospatial and other data to figure out exactly how to place and position solar panels for maximum efficiency.

Finance and Investment

This is perhaps the most developed category of Open Data businesses, as finance and investment companies have long-used open government data as an essential resource. Data from the Securities and Exchange Commission (SEC) has

powered investment firms for decades, and it is now possible to combine SEC data with other data sources for faster, more accurate, and more usable analysis. For example, Analytix Insight, which runs the website Capital Cube, provides analyses of more than 40,000 publicly traded global companies, updated daily, and presented in formats that make it easy for investors to use.

Other new companies provide a wide range of financial information and services to businesses and consumers. Brightscope uses information filed with the Department of Labor to evaluate the fees charged by different pension plans and helps employers and employees make more informed choices. Companies like Credit Sesame and NerdWallet compare different options and recommend credit cards and other financial services to consumers based on their credit ratings. Bill Guard uses data from the Consumer Financial Protection Bureau and information submitted by consumers to help protect people from fraudulent charges.

Some financial information companies are now processing financial data in the interest of helping small- and medium-sized enterprises (SMEs) get the capital they need—another example of the multiplier effect. These companies have realized that SMEs suffer because lenders cannot afford to do due diligence for small companies and thus don't have the confidence to give them the funds they need. On Deck now uses a number of public data sources to do that risk assessment and help small businesses get access to much-needed business loans. In a similar way, the British company Duedil serves to facilitate funding for SMEs in the UK and Ireland.

Food and Agriculture

In this area, perhaps more than any other sector besides healthcare, Open Data has the potential to revolutionize an industry that is essential to society and human wellbeing. The Climate Corporation, an iconic example of a successful Open Data company, has pioneered what is now being called “precision agriculture”—using Open Data to help farmers increase their efficiency and the profitability of their farms. The Climate Corporation, which was sold to Monsanto in the fall of 2013 for about \$1 billion, built value by combining different Open Data sources (ranging from satellite data to information on rainfall and soil quality) and subjecting it to sophisticated

analysis.⁷ The result is a set of services that can help farmers decide which crops to plant and when and help them prepare for the impact of climate change. Other companies, like FarmLogs, are beginning to offer some similar services.

Governance

Local government data is often no easier to use than federal data. Different cities use different and often unwieldy systems to track their government operations. Companies like OpenGov and Govini are providing platforms that municipal governments can use to organize their data and share it with their citizens. Organized and presented in clear charts and graphs, city data can become a tool for town meetings, city planning, and dialogue with city leaders. These tools also make it possible to compare operations in similar cities. For example, local data can allow a comparison between police overtime hours in Palo Alto and those in San Mateo, potentially revealing the reason for any disparity.

Housing and Real Estate

Real estate websites (which emerged about a decade ago) do much more than aggregate listings from brokers. Sites like Redfin, Trulia, and Zillow now offer data on schools, walkability scores, crime rates, and many other quality of life indicators, using data from national and local sources. In a country where historical averages show about one-fifth of the population moving every year, we can expect these sites to compete increasingly on the depth of information they offer and their ease of use.

Lifestyle and Consumer

In May 2013, the White House released the report of the Task Force on Smart Disclosure, a group chaired by this author to study how open government data can be used for consumer decision making.⁸ Federal agencies like the Consumer Financial Protection Bureau, the Department of Health and Human Services, and others now have data on a wide range of consumer services, including credit cards, mortgages, healthcare services, and more. Websites like FindTheBest have begun to use this kind of data to provide consumer guidance on a range of products and services.

As the idea of Smart Disclosure takes hold, we can expect to see more websites tailored to particular consumer needs and concerns. GoodGuide, for

example, uses data from more than 1,500 datasets to create a service for consumers who want to choose the products they buy with an eye towards their environmental impact, health concerns, or other factors. GoodGuide's analysis is not only being used by consumers but also by companies that want to use the data to "go green."

Transportation

The availability of new, usable transportation data is transforming this sector as well. The applications include companies that provide detailed directions and traffic advisories (HopStop, Roadify Transit), traffic analytics to help transportation planners (Inrix Traffic), and safety data to improve the trucking industry (Keychain Logistics).

The Biggest Opportunity for Data-Driven Disruption — Healthcare

While all these categories of data-driven companies have significant growth potential, it is in healthcare that new uses of data may bring the greatest opportunity for disruptive innovation. We can expect more efficient systems for tracking patients and their care, leading to lower costs and fewer medical errors. We can look forward to more data-driven diagnostics, treatment plans, and predictive analytics to more scientifically determine the best treatments. And we will see a new era of personalized medicine, where data about an individual—ranging from genetic makeup to exercise habits—is used to algorithmically determine a strategy for care.

Healthcare has become a proving ground that shows how the four different kinds of data—Big Data, Open Data, personal data, and scientific data—can be used together to great effect. By analyzing *Big Data* (the voluminous information on public health, treatment outcomes, and individual patient records), healthcare analysts are now able to find patterns in public health, healthcare costs, regional differences in care, and more. *Open Data* on healthcare is becoming more available through the Centers for Medicare and Medicaid Services (CMS) and recent data releases from the U.S. Food and Drug Administration. With *personal data*, the third piece of the puzzle, people are getting more data to help them understand and manage their own health issues, both through Blue Button and similar programs and through personal health monitoring devices. And we're seeing a rapid increase in open *scientific data*, particularly data about the human genome, which can be used to improve medical care.

New companies are launching to put all this data to work. Venture capitalists reportedly invested more than \$2 billion in digital health startups in the first half of 2014.⁹ There are three categories of companies that are growing most rapidly.

Healthcare Selection

A number of websites now use a combination of Open Data and consumer feedback to provide information on the quality and cost of different healthcare options. ZocDoc and Vitals help people find doctors and clinics and book appointments. Aidin uses data from CMS and other sources to help hospital discharge planners work with their patients to find better post-hospital care. Drawing on Open Data from the U.S. National Provider Identifier Registry, iTriage lets you use a website or smartphone to log in symptoms, get quick advice on the kind of care needed, and get a list of nearby facilities that can help. And TrialX connects patients with clinical trials of new treatments. As CMS releases more data on both the quality and cost of care, data-driven healthcare companies will have an opportunity to help individuals and drive down national healthcare costs.

Personal Health Management

The movement to electronic medical records will open new ways for individuals and their doctors to combine public and personal information to improve their healthcare. Amida Technology Solutions is building on the Blue Button model to accelerate the use of personal health records. At the same time, other companies are tapping the power of personal data in different ways. Propeller Health uses inhaler sensors, mobile apps, and data analytics to help doctors identify asthma patients who need additional help to control their chronic disease. Iodine combines large healthcare datasets with individualized health information to provide patient guidance. And several companies have developed wristbands and other wearable monitors that track personal biometric data as an aid to wellness programs and medical treatment.

Data Management and Analytics

As more health data is opened up, more companies are finding ways to analyze it. Evidera uses data from CMS, databases of clinical trials, and other sources to develop models predicting how different treatment interventions will affect different kinds of patients. In a similar way, Predilytics uses machine learning to help health plans and providers deliver care more effectively

and reduce costly admissions (and readmissions) to the hospital.

Business and Revenue Models for Data-Driven Companies

Open Data poses a business paradox. How can one hope to build a business worth millions or even billions of dollars by using data that is free to the public? Open Data startups have succeeded by bringing new ideas, analytic capabilities, user-focused design, and other added value to the basic value inherent in Open Data. As with many startups, the revenue model for many of these companies is still a work in progress. They have focused on functionality first, monetization second. (In a few years, we will know whether this has been a wise strategy.) Nevertheless, several business models are starting to emerge.

In a 2012 study,¹⁰ Deloitte surveyed a large sample of Open Data companies and identified five business archetypes:

Suppliers publish Open Data that can be easily used.

Aggregators collect Open Data, analyze it, and charge for their insights or make money from the data in other ways.

Developers “design, build, and sell Web-based, tablet, or smart-phone applications” using Open Data as a free resource.

Enrichers are “typically large, established businesses” that use Open Data to “enhance their existing products and services,” for example, by using demographic data to better understand their customers.

Enablers charge companies to make it easier for them to use Open Data.

The Open Data 500 study has found a number of companies that combine several Deloitte archetypes, particularly among companies that the Open Data 500 categorizes as “Data/Technology.” For example, Enigma.io, as described above, has *aggregated* about 100,000 government datasets, *supplied* that data to the public in a more useful form, and served as an *enabler* by consulting with companies that have special uses for certain kinds of datasets (e.g., risk analysis).

While Deloitte’s categories describe the different ways in which companies use Open Data to deliver business value, the Open Data 500 study has focused on a different part of the business model—the ways companies generate revenue from their work. The Open Data 500 has found a variety of revenue sources that are available to companies across the archetypes identified by Deloitte.

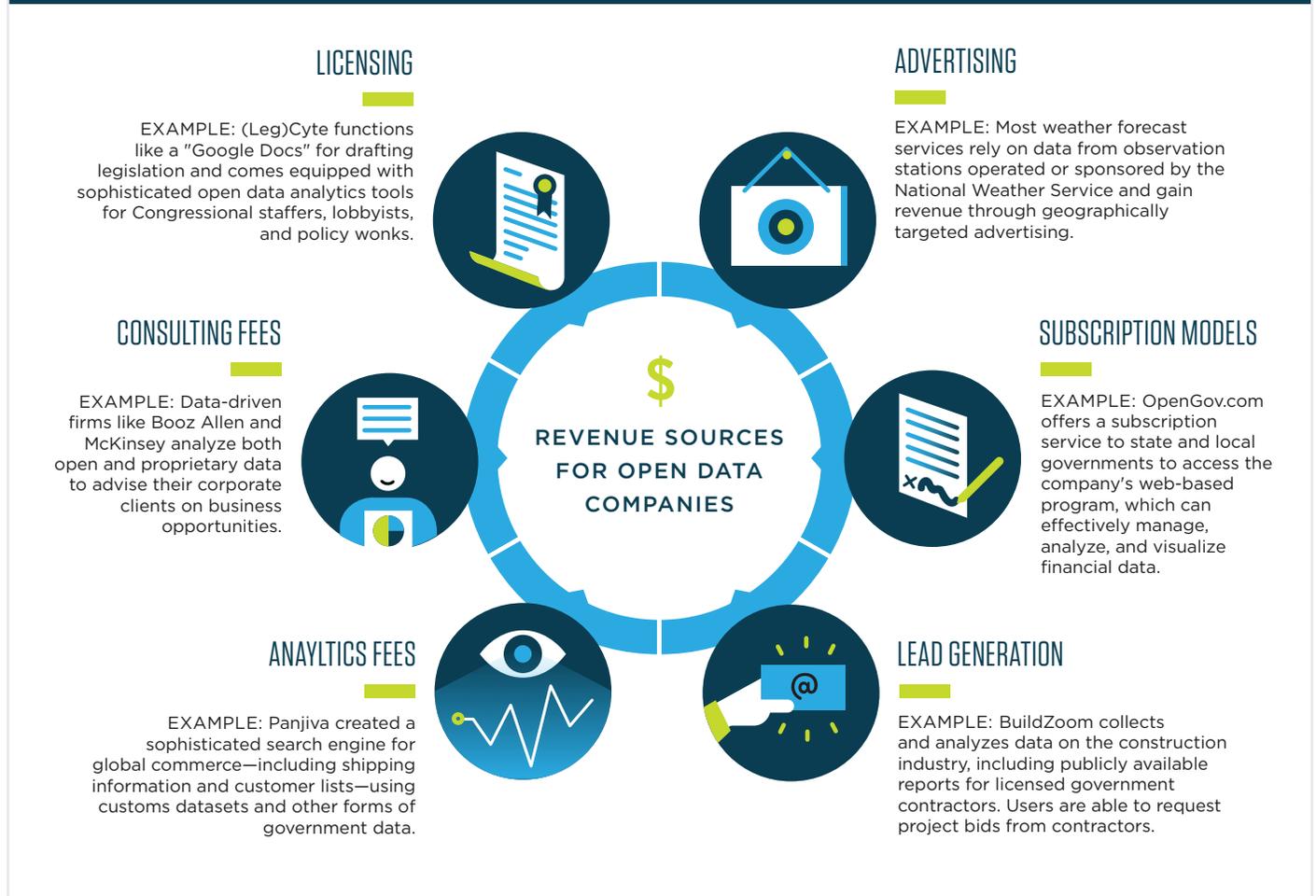
Advertising, a common revenue source for websites, may be a good source for some Open Data companies. Websites are increasingly relying on “native advertising;” that is, sponsored content that is written at an advertiser’s direction but presented in a way that looks like the site’s regular content. This kind of advertising, however, may be at odds with Open Data companies that base their business on the promise of providing objective and unbiased information.

Subscription models, in contrast, can be a natural fit for these data-driven companies. Many add value to Open Data as they combine datasets, analyze data, visualize it, or present it in ways that are tailored to the user’s needs. Willingness to pay depends on the relevance, complexity, uniqueness, and value of the information. While a consumer would be unlikely to subscribe to a simple website that helps him or her choose a credit card, a farmer could easily find it worthwhile to subscribe to a service that uses data to help improve the farm’s profitability.

Lead generation is another natural revenue source for data-driven companies that evaluate business services or help consumers find products and services. Real estate sites are a classic example. They collect a broker’s fee when someone uses the website to find and buy a home. The challenge with this revenue model, however, is that it can give companies an incentive to game the system and refer consumers to service providers that pay the highest referral fees. Over time, this model could generate consumer distrust and become less effective. Companies that use this revenue source should consider establishing a voluntary code of conduct that would include transparency about their business models—enough to let users know how the company earns its revenue and give them the assurance that their information is unbiased.

Fees for data management and analytics provide revenue to companies that help clients learn more from the data available to them. They may work

REVENUE SOURCES FOR OPEN DATA COMPANIES



with businesses, governments, or both. Several companies now help government agencies manage and analyze their own data—or even sell agencies' data back to them in an improved form, as Panjiva does with customs data from the federal government.

Consulting fees are yet another revenue source. Some data-driven firms, like Booz Allen and McKinsey, analyze both open and proprietary data to advise their corporate clients on business opportunities, while investment firms use increasingly diverse data sources to predict market trends.

Finally, licensing fees are a source of revenue for the kinds of companies that Deloitte calls enablers. They can license software, tools, platforms, database services, cloud-based services, and more to enable new data-driven companies to build their business.

Potential Barriers and Ways to Overcome Them

Any discussion of data-driven business has to deal with the issue of data privacy. Chapter 7 addresses privacy concerns surrounding the use of proprietary Big Data, such as personal data gathered online, through data brokers, through customer records, or in other ways. Companies driven by Open Data do not generally use this kind of personal information, but they may need to access datasets that aggregate personal data and present it in a way that masks personal information. Healthcare companies, for example, may want to use anonymous patient records in a way that enables them to detect patterns in treatment outcomes, such as correlations between prescription drugs, lifestyle, and therapeutic results.

There is an ongoing debate about whether it is truly possible to anonymize data like this or whether any system of anonymization can

ultimately be defeated. This debate is likely to play out over the next few years and could have a major impact on data-driven innovation. If successful technologies for anonymization are developed, they will open up new opportunities for data analysis and publication. On the other hand, if experts and the public come to believe that individuals' identities can always be deciphered from the data, then a number of paths to innovation will be cut off.

Data-driven businesses also have to deal with one of the biggest obstacles to growth: poor-quality data. The quality of U.S. government data varies greatly between agencies and sometimes even within the same agency. Government data systems have grown by accretion over decades. They are often housed in obsolete data management systems and, at the same time, may have errors, gaps in the data, or out-of-date information. These are not easy problems for either the government or third parties to solve, but without some solutions, the potential of open government data will be underused.

Government agencies that provide data, and the businesses and non-profits that use it, all have a common interest in making government data as relevant, accessible, actionable, and accurate as possible. None can do it acting alone. What is needed is a way to bring together data providers with data users for a structured, action-oriented dialogue to identify the most important datasets for business and public use and find ways to improve them.

The GovLab has launched a series of Open Data Roundtables to bring together federal agencies with the businesses that use their data. The first such roundtable, held with the Department of Commerce in June 2014, included more than 20 officials and staff from the department and about 20 businesses. This event was the beginning of a process designed to identify specific areas for improving data quality and accessibility. It was followed by an Open Data Roundtable with the U.S. Department of Agriculture. As of this writing, additional GovLab roundtables are being planned with the Departments of Labor, Transportation, and Treasury, as well as with other federal agencies.

A next step could be to develop public-private collaborations to turn government data into machine-readable forms that could make it much more useful and help drive innovation. Some companies, such as Captricity, have developed the technology to convert data from PDF files (a document type common in government data) into more usable formats. By working together, government agencies and these companies could convert large amounts of the most important data into a form that other companies could easily use.

Supporting data-driven innovation will also require government policies that make new sources of data available and encourage companies to use them. The federal Open Data Policy,¹¹ established in May 2013, and the National Open Data Action Plan,¹² released one year later, set out some basic principles for the U.S. government to use in promoting Open Data. The Open Data Policy not only directs federal agencies to release more Open Data, it also requires them to release information about data quality. We can hope and expect that they will do some data cleanup themselves, demand better data from the businesses they regulate, or use creative solutions, like turning to crowdsourcing for help.

The federal government has steadily made Data.gov (the central repository of its Open Data) more accessible and useful. The General Services Administration, which administers Data.gov, plans to keep working to make this key website better still. As part of implementing the Open Data Policy, the administration has also set up Project Open Data on GitHub, the world's largest community for open-source software. These resources will be helpful for anyone working with Open Data, be it inside or outside of government.

As more and better Open Data becomes available, we will learn more about the best ways to use it for business applications, job creation, and economic growth. While it is clear that Open Data can be used in a wide range of industries, we do not yet know exactly which kinds of applications will turn out to be the most promising, the most robust, and the most replicable. We need to learn more about the mechanisms of value creation using Open Data, the kinds of Open Data that will be most important to various sectors, and the ways in which Open

Data fits into different companies' strategic and operating models. Ongoing research on the uses of Open Data by academic institutions, government agencies, and independent organizations will be essential to ensure that our public data resources are used widely and well.

Ultimately, creating a better Open Data ecosystem will take both public and private resources and funding. The payback for government technology

improvements is generally calculated based on short-term savings, like improvements in efficiency and cost reduction. But the release of more and better open government data can have economic benefits that multiply over time. An investment in Open Data now will pay off for years to come.

ENDNOTES

- 1 Stefaan Verhulst, "The Open Data Era in Health and Social Care," *The Governance Lab*, May 2014.
- 2 James Manyika et al., "Open Data: Unlocking Innovation and Performance with Liquid Information," *McKinsey Global Institute*, Oct. 2013.
- 3 Internet Live Stats, "Twitter Usage Statistics" <<http://www.internetlivestats.com/twitter-statistics/>> (19 Aug. 2014).
- 4 Nick Sinai and Adam Dole, "Leading Pharmacies and Retailers Join Blue Button Initiative," *Health IT Buzz*, 7 Feb. 2014.
- 5 The author of this chapter is a senior advisor at The GovLab and project director for Open Data 500.
- 6 Megan Rose Dickey, "This Tech Crunch Disrupt Winner Could Be the Future of Search," *Business Insider*, 2 May 2013.
- 7 "Monsanto Completes Acquisition of The Climate Corporation," *MarketWatch*, 1 Nov. 2013.
- 8 Executive Office of the President, National Science and Technology Council, "Smart Disclosure and Consumer Decision Making: Report of the Task Force on Smart Disclosure," May 2013.
- 9 Eric Whitney, "Power to the Health Data Geeks," *National Public Radio*, 16 June 2014.
- 10 "Open Growth: Stimulating Demand for Open Data in the UK," *Deloitte Analytics*, 2012.
- 11 OMB Memorandum for the Heads of Executive Departments and Agencies M-13-13, "Open Data Policy - Managing Information as an Asset," 9 May 2013.
- 12 The White House, "U.S. Open Data Action Plan," 9 May 2014.