

A CLOSER LOOK AT OPEN DATA: OPPORTUNITIES FOR IMPACT

Local and regional governments can use open data to make substantial improvements to education, transportation, and health care—but first they will have to overcome some big challenges.

MICHAEL CHUI is a partner at the McKinsey Global Institute, and DIANA FARRELL is a director in McKinsey's Washington, DC, office.

¹Data sets range from completely open to completely closed across four dimensions: *accessibility* (the range of users permitted to access the data); *machine readability* (the ease with which the data can be processed automatically); *cost* (the price of obtaining data); and *rights* (limitations on the use, transformation, and distribution of the data). We use the terms “open data” and “liquid data” interchangeably.

Public institutions and private companies are aware of the opportunities and challenges associated with big data analytics. But as these organizations take steps to harness the power of big data, it's important that they are also aware of a complementary global trend; open data—the release of information by governments and private institutions and the sharing of other data within and across industries—extends the power of big data and makes possible entirely new products and services, while also enabling better decision making and improved operations in both the private and public sectors.¹ In fact, our research suggests that open data can help unlock more than \$3 trillion annually in seven domains of the global economy: consumer finance,

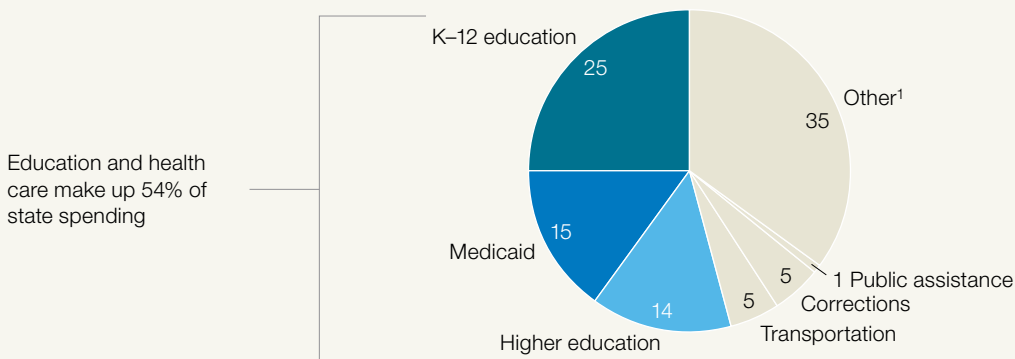
consumer products, education, electricity, health care, oil and gas, and transportation. Many regional and local governments have already opened up their data to increase transparency and accountability, promote civic engagement, and improve basic public services. In the United States, for example, residents of Texas can help their government identify potential cost savings by scrutinizing publicly available budgetary information. Some third parties have used “liquid” information (data that are open, widely available, and in shareable formats) to create smartphone apps that tell commuters when the next bus will arrive or that guide drivers to the nearest available parking spot.

These applications, while promising, are just the beginning. Open data can help local and regional governments transform the way they deliver services while also offering significant cost savings. Moreover, by sharing, where appropriate, the vast amount of data at their disposal, governments can catalyze the use of open data in other sectors, which can in turn help create value in the broader economy. (For more, see “How government can promote open data and help unleash over \$3 trillion in economic value.”) Realizing this potential will not be easy and will require creating safeguards for personal privacy, making investments in technology, changing mind-sets and work processes, and embrac-

Exhibit

Education accounts for nearly 40 percent of state spending in the United States.

State spending, % of budget



¹“Other” includes care for residents with disabilities, economic development, health benefits for public employees, parks and recreation, pensions, state police, and general aid to local communities.

Source: *State Expenditure Report: Examining Fiscal 2010–2012 State Spending*, National Association of State Budget Officers, Dec 2012, nasbo.org

ing new concepts such as MyData—a term that describes the sharing of information with the individual about whom it has been collected. (See sidebar, “About MyData.”) Here, we discuss the open-data opportunity—and the accompanying challenges—in three of the seven domains we researched that have the closest ties to regional- and local-government budgets: education, transportation, and health care (exhibit).

Education: Opportunities inside and outside the classroom

More than one billion students are enrolled in schools around the world, and those students are served by more than 58 million teachers. Annual spending on kindergarten through grade 12 and postsecondary education exceeds \$4 trillion worldwide.² In some countries, public education is funded primarily at a national level; in others, it is largely the responsibility of state and local govern-

ments. In the United States alone, state and local governments spend about \$1 trillion a year on education; as the exhibit shows, in fiscal year 2011, 25 percent of state dollars were spent on schooling for K–12 students. With so many resources dedicated to public education, there are substantial opportunities to improve current systems.

Improve instruction. Many of the most valuable applications of open data will evolve in the classroom itself. Schools can improve instruction by using open data to develop personalized learning plans for students, provide frequent feedback to teachers on their performance, and create targeted professional-development programs. MyData programs will be especially important here, as they allow families to track student progress and correct inaccurate information.

For families and educators to identify areas for improvement, data on student performance and information about specific teaching practices can be combined with information about individual learning preferences and local educational mandates or guidelines. Adaptive-learning programs gather data about how an individual student interacts with specific lesson plans and then they contextualize those measurements—with information about that student’s past performance and the performance of his or her cohort—to identify areas that the student has mastered, or conversely, where he or she might need extra help. When used for remedial-math programs at Arizona State University, personalized learning plans improved pass rates by nearly 10 percent.

This data-driven approach to teaching and learning allows for targeted, rapid responses; by using insights gained from frequent data collection and analysis, problems can be corrected and lesson plans adapted before student performance suffers. Sharing data across institutions and school districts—

²In defining the scope of the effort, we include kindergarten through grade 12 and higher education delivered through public and private systems in the United States, the European Union, Asia–Pacific, and across emerging nations.

About MyData

MyData is not a specific program. It is a concept that entails sharing information collected about an individual (or organization) with that person (or company). MyData will be an important part of capturing the open-data opportunity, especially in the fields of education and health care. For example, some hospitals now provide individual patients with access to their own medical-records data. Providing aggregate statistics (a form of open data) alongside MyData allows useful comparisons; some utilities show consumers how their energy use compares with that of neighbors to encourage conservation. In many cases, individuals can modify or correct the data provided to them, which improves its quality.

making the data more liquid—allows for more robust analysis. Educators can monitor student attainment frequently and at a granular level (measuring mastery of a particular concept, for example) and track how individual student performance correlates to classroom activities (lectures versus exercises or group work). These approaches can also be used to increase motivation, influence mind-sets, and adjust learning strategies to improve student achievement.

Match students to schools and programs.

Open data can be used to help parents and students identify the best fit with a school or program. Peng Shi, a graduate student, created Boston's new methodology for public-school assignments by using information released by the city on the quality and location of schools.³ His algorithm presents parents with choices based on school quality and distance. While school choice in K–12 grades may help students find institutions that can better serve their needs, many districts cannot offer every student a slot in a high-performing school; open data can help families identify alternatives and put pressure on systems to provide more high-quality options.

Ensure efficient system administration.

Schools can reduce procurement costs by using open data to consolidate, standardize, and compare the types of products they purchase. By allowing administrators to analyze past purchases and examine how peer institutions are allocating their budgets, open data can also enable better forecasts for when products or services will be required.

One country in Europe, for example, has used open data to reduce the price its schools pay for a wide range of supplies. Its open-bid platform hosts an online auction where vendors bid on contracts to supply basic

goods to schools across the nation. Suppliers see information about current bids, the quantity of goods requested, and data on previous auctions, such as closing bids, prices, and quantities. Just two years after implementation, the average price of the items offered on the platform had dropped by 24 percent. These efforts not only reduce procurement costs but also allow vendors to better forecast demand.

While schools can reap many of the benefits of enhanced procurement using their own data, sharing their information across schools and districts can lead to even better results through benchmarked price negotiations. Improved purchasing of broadband services is one example. By comparing the prices of broadband service offered by local suppliers—and by benchmarking their costs against their peers'—one school district in northern California was able to reduce broadband costs by 30 percent.

Challenges

The benefits of open data in the field of education are tremendous; in addition to the opportunities discussed here, open data can help match graduating students to appropriate employment and can make the true costs of higher education (and private K–12 education) more transparent. But to capitalize on any of these opportunities, stakeholders must share information across the education system so they can identify areas ripe for improvement, and they must do so in a way that protects the identity of individual students and teachers. They will also need standardized measures for analyzing student, teacher, and school performance and for identifying best practices so that educators can learn about learning—using data to see which approaches work best for students and refining these techniques through ongoing monitoring.

³James Vaznis, "MIT has plan for Boston school assignments," *Boston Globe*, October 28, 2012, bostonglobe.com.



A teacher helps students in class.

© ABLEIMAGES/THINKSTOCK

Transportation: Doing more with less

Open data has already played an important role in improving transportation; we believe it will continue to play a role in addressing the inefficiencies that persist. Open data can help local and state governments reduce transportation spending and improve services in three ways.

Improve infrastructure planning and management. Open data can help governments improve transportation-demand forecasting, prioritize transport infrastructure improvements, and synchronize the ways different modes of transportation interoperate. It is an especially powerful tool when combined with other transportation statistics. For example, in 2012, the city of Moscow used open data to complement its own estimates of how long it would take citizens to commute via different modes of transport; specifically, city officials used Internet resources that measure traffic, as well as mobile-phone-location data, transportation-operation statistics, and

public projections of city and regional development. By combining its own data with open data, the city was able to determine whether extending its subway system into the suburbs was necessary or if other services could do a better job of meeting demand. The transit authority ultimately decided against making a costly investment in the subway extension and instead planned a less costly investment in a suburban-rail extension. This limited Moscow's up-front costs and ensured that services could be flexible enough to meet the needs of a shifting population. The new suburban-rail extension is part of a larger transportation strategy that, in addition to avoiding more than \$1 billion in infrastructure costs, has the potential to reduce average commute times by 16 minutes per trip, saving one week of travel time for each rider every year.

Optimize fleet investment. Open data can help optimize the size and mix of public fleets—all the vehicles owned and managed by

the government—and make operations more efficient. In one US city, for example, open data helped the local government understand how its vehicles were being used, which uncovered an opportunity to cut the fleet by as much as 30 percent. And in California, after the state released budget data on vehicle spending, citizen advocates spotted examples of unnecessarily high costs, and state agencies were able to reduce the fleet by 15 percent.

Inform customer decision making. Open data can help customers make better decisions about when and which type of public transportation to use; when aggregated, these choices can be the source of significant economic value. Public-transit systems have expanded the use of sensors that generate location data, for example, which can then be used to transmit train and bus locations in real time. This information is used by transit agencies to manage their operations, but these data streams also have been made available to entrepreneurial developers such as CUBIC, whose NextBus app lets riders in cities such as Los Angeles, Toronto, and Washington, DC, see current wait times and delays.

Open data on bus and train location and road congestion can shrink the “reliability buffer”—the extra time a traveler builds into a trip to account for possible delays, which can be as much as 70 percent of total trip time.⁴ Open data on vehicle location and on-time performance can also help attract new users because of the increased predictability of services. In Duluth, Minnesota, for example, Google Transit installed transponders that record city-bus locations and make the data public so riders can decide if it offers a suitable alternative to driving. After the transponders were installed, ridership increased by 12 percent, reducing traffic and also decreasing the cost of transportation for many passengers.

Challenges

Using open data to improve transportation delivery can be tricky. Such projects raise privacy concerns, and there are gaps in data standards and capabilities in some transportation organizations. There are political and financial barriers as well; infrastructure projects are expensive, long-run investments that often involve policy makers from multiple jurisdictions. There are also concerns about providing transportation services in rural areas where the financial return on investment in infrastructure, which open data can help estimate, is of secondary importance.

Health care: Better care at lower costs

Using open data to improve health care will be driven largely by stakeholders outside local governments—for example, providers, payors, national governments, and patients who choose to share their information—but these advances will in fact generate savings at the local level, especially in the context of Medicaid spending in the United States. We have identified several levers for capturing value in health care through the use of open and proprietary data. We focus on levers that can be applied in the United States, but many of the levers we identify would work in other nations and have been applied in health-care systems around the world.⁵

Right living. An enormous and largely untapped source of value in health care lies in the use of data by patients to manage their own health to avoid illness and to get better results from treatment if and when they do become ill. The major focus is providing patients with the information they need to make healthy lifestyle choices and to manage their treatments effectively.

Health-care data that are made more liquid can reduce the occurrence of lifestyle-linked conditions and illnesses (such as hypertension and diabetes) by better identifying at-risk

⁴*Travel Time Reliability: Making It There On Time, All The Time*, US Department of Transportation Federal Highway Administration, 2006, dot.gov.

⁵This section relies heavily on research documented in *The ‘big data’ revolution in healthcare: Accelerating value and innovation*, January 2013, mckinsey.com.

individuals. This is done by combining patient information (such as exercise habits) with demographic information and then analyzing outcomes across different patient populations. At-risk patients can then be targeted for health education or for assistance in preventing illness—for example, by recommending screenings or issuing personal reminders.

Open data can also help with drug adherence—failure to use medication as prescribed for chronic conditions costs the US health-care system \$100 billion to \$289 billion a year.⁶ Ginger.io has developed a smartphone app that can detect when a patient is not using drugs correctly or has changed behavior in a way that indicates a potential medical emergency. The app collects motion data to establish a base pattern and then looks for deviations that might signal trouble; a sharp drop in activity, for example, might indicate that a patient has stopped taking antidepressants or has fallen down. Irregular sleep patterns could signal that an anxiety attack is imminent.

Right care. Research shows that suboptimal care is often the result of poor communication between patients and doctors and among doctors themselves. This leads to inappropriate or redundant care that can result in complications and raise costs.

Electronic medical records can help avoid these problems by creating a single consolidated record for each patient and giving access to all the patient's caregivers. In the United States, participating providers can offer patients access to their own medical records—their medical MyData—through an initiative called Blue Button, which was pioneered by the Department of Veterans Affairs. In France, patients can carry their entire medical histories on a smart card—a card with a memory chip that any doctor or care facility can use. The system can also flag potentially harmful drug inter-

actions for a specific patient before a physician selects a drug treatment or performs a radiologic test.

Open data can also be used to mount a vigorous response to disease outbreak—another aspect of right care. Public-health agencies collect data from emergency rooms and other sources to detect outbreaks of certain diseases so that members of the public can take steps to protect themselves. Google has shown that it can map a flu outbreak in real time by monitoring the number and locations of users searching for “flu” and related topics. A start-up called Propeller Health has created a GPS-enabled tracker that monitors inhalers; it can be used to detect when environmental conditions such as high pollen counts might trigger severe asthma problems.

Right provider. “Right provider” means identifying the doctors and care settings that provide high-quality care for a specific task and determining which resources are needed to deliver it (for example, determining which procedures are best done by a physician assistant or nurse rather than a doctor). Optimizing provider decisions depends on many organizations having access to performance data. Performance data can be used to align incentives with outcomes (for example, paying doctors for successful treatment rather than for procedures).

This approach can give patients better options and raise the quality of care. For example, in Arkansas, all Medicaid providers who treat upper-respiratory infections, pregnancy, attention-deficit-hyperactivity disorder, and select diseases receive a report on the cost and quality of treatment. The aggregated data on all a provider's claims help doctors understand how their practices compare with other providers in the state, giving them a way to identify opportunities to improve. New York

⁶“Meera Viswanathan et al., “Interventions to improve adherence to self-administered medications for chronic diseases in the United States: A systematic review,” *Annals of Internal Medicine*, 2012, Volume 157, Issue 11, pp. 785–95.

State in the United States and the National Health Service in the United Kingdom have opened up data about hospital performance. These data can be used by patients, doctors, payors, and the public to identify the best care available in their communities.

Right value. Open data can help increase value in health-care delivery by controlling costs and improving quality. This includes using data to eliminate fraud, waste, or abuse in the Medicare and Medicaid programs, which can exceed \$70 billion annually.⁷ Regulators can combine claims data with public information about patients (such as date of birth, address, employment, or registry of motor-vehicle information) to confirm patient identity; this can help reduce instances of billing for services that were not provided. Using open and proprietary data to reduce costs, maintain quality, and fight fraud could enable value of \$50 billion to \$100 billion a year.

Challenges

Capturing the full value of open data in the US health-care system would require several changes. The most fundamental one would be a shift in medicine and caregiving to data-driven approaches, in which physician decisions about treatment are informed by results from thousands of patients. Payment systems would also need to be adapted because conventional means of controlling costs, such as negotiating prices of per-procedure fees, are not geared toward taking advantage of the insights that open data provide.

There are also technical and organizational barriers to realizing the full value of open data in health care, including the inability of many health-care data systems to provide standardized information. Finally, there are concerns about privacy and confidentiality—the consequences of mishandled medical data

can be extremely serious. Government, providers, and payors must ensure that effective systems are in place to keep shared medical records confidential.



Making information more transparent is an important first step toward improving public services, increasing transparency, and saving taxpayer money, but it is not sufficient to capture the value of open data. Governments, like private institutions, will need to prioritize opening the data sets that will create the greatest benefits; they will also need to invest in technology and cultivate a vibrant network of third-party developers that can transform raw data into valuable tools. Additionally, they will need to acquire and develop the talent, processes, and cultures to complement their technological investments. Key skills include the ability to perform analyses, create useful reports and tools based on open data, and incorporate data into managerial decision-making processes. In addition to meeting the technological requirements of capturing the value of open data, governments must also create thoughtful policies that protect intellectual property and ensure privacy of both consumers and institutions. ■

⁷*Medicare and Medicaid Fraud, Waste, and Abuse: Effective Implementation of Recent Laws and Agency Actions Could Help Reduce Improper Payments*, US Government Accountability Office, 2011, gao.gov.