

Public Disclosure Authorized

Public Disclosure Authorized

Public Disclosure Authorized



# BIG DATA

*in ACTION for GOVERNMENT*

*Big Data Innovation in Public Services,  
Policy and Engagement*



# Contents

Introduction.....	1
Strengthening Government Service Delivery.....	3
Smarter Policymaking.....	7
Deepening Citizen Engagement.....	11
Putting Big Data into Action for Government.....	14
Additional Resources.....	16

# Introduction



**Big data is a term widely used to describe the exponential growth of data, particularly the data flowing from ubiquitous mobile phones, satellites, ground sensors, vehicles and social media. It also explains the rise of the computing technologies and algorithms that harness big data for valuable insights. In the public sector, big data typically refers to the use of non-traditional data sources and data innovations to make government solutions more responsive and effective.**

Governments have an opportunity to harness big data solutions to improve productivity,

performance and innovation in service delivery and policymaking processes. The UK's Digital Government Unit estimates the country's government can save around US\$20-41 billion annually through big data analytics – a figure based on efficiency gains, reduced fraud and error, and improved tax collection alone. McKinsey estimates big data can reduce the European Union's administrative costs by 15 to 20 percent, creating the equivalent of at least US\$223-446 billion in new value<sup>1</sup>. In developing countries, governments have an opportunity to adopt big data solutions and leapfrog traditional administrative approaches.

Big data is a pressing issue, particularly at a time when many are concerned about the role of information in political change. Governments in the United Kingdom, the United States, Singapore and Korea are among many<sup>2</sup> that have adopted high-level big data strategies. These frameworks focus on the

1. See <http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/public-sector-digitization-the-trillion-dollar-challenge> || <http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation>

2. Other countries with big data strategies include Australia, France, Japan, New Zealand

various roles of government as **producer**, **consumer** and **facilitator** of big data to enable better service and policy outcomes:

- **Producer:** More than a million datasets from governments around the world are available online, from tax returns and unemployment claims to hospital funding and energy use.
- **Consumer:** Governments can use big data (both their own and from other sources) to promote responsive government. The

US Census Bureau buys commercial data for address verification, for example, and transportation agencies use information from GPS navigation systems to inform traffic management.

- **Facilitator:** Governments should invest in big data research and stewardship, as well as establish robust legal and technical frameworks to ensure big data is accessible and responsibly used for public good.

## PROMISING APPLICATIONS FOR BIG DATA IN GOVERNMENT

The potential for big data to transform government is vast. This brief focuses on big data solutions with applications in service delivery, policymaking and citizen engagement – areas where big data can play a transformational role. It also discusses key initiatives where the government needs to facilitate action for effective data use.

### SERVICE DELIVERY

Big data analytics can be used by governments to improve existing services and to draw on novel datasets to drive entirely new public services

### POLICYMAKING

Policymakers are using satellite imagery, cell phone data and more to produce alternative economic indicators for new – and real-time – policy insights

### CITIZEN ENGAGEMENT

By applying machine learning to online and social media, governments can be more responsive to citizen sentiment, ushering in a new dimension of civic engagement

*Application Areas for BIG DATA in Government*



# Strengthening Government Service Delivery

These days, consumers are accustomed to commercial products and services growing more personalized and always on-demand. People interact with government services every day in health, employment, education and business, just to name a few. Society is eager to use big data to make public service delivery equally as smart, responsive and personalized.

One area of tremendous opportunity is in the use big data algorithms to detect fraud, non-compliance and bottlenecks in government operations. Governments are huge producers of data, most of which is unstructured and text-intensive. Text analytics and machine learning algorithms are indispensable to analyze administrative data for insights. They can automate systematic, multilevel checks on insurance, procurement and tax records to flag entries requiring further examination, or to identify performance bottlenecks that need attention.

## PROCUREMENT AND FINANCIAL FLOWS:

Big data techniques are transforming processes to detect fraud, collusion and money laundering through the analysis of procurement and financial information. Algorithms are used to trawl procurement records, invoices, bank information, certificates of origin and other data to identify fraud, collusion and shell companies.

***Solution Spotlight:*** The [International Consortium of Investigative Journalism](#) (ICIJ) provides nearly 40 years of information on companies in over 200 countries and territories. ICIJ hosts the Mossack Fonseca (aka Panama) papers – 11.5 million emails, database entries, images and documents which big data analytics helped to rapidly sort, categorize, index and search. The analysis found several irregularities and triggered multiple investigations.

***Solution Spotlight:*** The Counter Fraud Centre at the UK's Chartered Institute of Public Finance and Accountancy (CIPFA) manages the **London Counter Fraud Hub**, which supplies big data analytics, investigations and recovery services for the 32 London Local Authorities and the City of London Corporation.

**HEALTHCARE:** Valuable information about public health is found in government records, as well as unconventional sources such as social media, discussion forums, mailing lists, health websites and news outlets. Mobile phone data is also being widely used for monitoring and improving health services. These channels can complement traditional sources to help authorities monitor results and feedback in real-time, and improve performance and outcomes in hospital management operations, insurance processing, vaccinations, nutrition and disease.

**Solution Spotlight:** **HealthMap** is a platform that automatically tracks and analyzes multiple data-feeds in 15 languages – including official documents, news reports, social media, international health organizations, government websites and health workers' blogs – to produce an online visualization of disease trends. This approach can facilitate the early detection of disease outbreaks, enabling prompt and targeted responses.

**Solution Spotlight:** In Brazil, the Ceara State government is using **Proactive listening**, a mobile-phone based system to provide the government with real-time data on the coverage and quality of public health services. Real-time information enables the government to identify emerging problems, such as bribes for health services, and intervene in a timely and effective manner.

**Solution Spotlight:** During West Africa's 2014 Ebola crisis, Swedish nonprofit **Flowminder** used cellphone data to create models predicting population mobility and potential routes of disease transmission, while experts from around the world created the **Ebola Open Data Initiative**. This organized access to multiple big data resources and created a tool to analyze them, leading to the development of an open-source global model for Ebola, which allowed health officials to test different scenarios and interventions.

**PUBLIC UTILITIES:** Remote sensing data from satellites and ground-based sensors can provide a wealth of real-time or near-real-time information to monitor the provision and quality of public utilities such as water and energy.

**Solution Spotlight:** **Shanghai's** municipal government has installed sensors throughout the city's 3,300-kilometer waterpipe network. These sensors allow the authorities to identify precise locations of issues such as leaks, dramatically reducing response times, infrastructure damage, maintenance costs and service disruption.



**Solution Spotlight:** The [India.Nightlights](#) platform uses night light data from satellite images to monitor electricity provision over time to all 600,000 villages in India.

**Solution Spotlight:** The SWEET Lab is working with local governments and service providers in developing countries to understand water and sanitation services. [Sensors](#) on water pumps in Kenya and Ethiopia provide usage statistics and monitor functionality, to help keep pumps working.

**EDUCATION:** Personal data from devices, exam data and other sources can be used in innovative ways to monitor student performance, better understand teaching practices, and help parents and students identify the best fit with a school. Big data can complement traditional sources to enable new delivery regimes that tailor lessons to performance, and monitor progress towards government and global goals.

**Solution Spotlight:** In **Kenya**, the national government uses a mapping platform to show areas where educational resources are lacking. Tanzania has established the website **Shule.info** to help assess the quality of schools.

**Solution Spotlight:** In Mexico, the **ENOVA** platform that uses data analytics and data from student interaction and feedback to continuously troubleshoot educational processes and improve learning tools.



**LAND ADMINISTRATION:** High-resolution imagery from satellites and Unmanned Aerial Vehicles (UAVs or drones), combined with cloud computing and handheld computing devices, can deliver capabilities to revolutionize land registration processes and survey methods. These near-real-time methods will inevitably replace traditional surveying approaches. Land registration improves public revenues by increasing tax collection, enables efficient resolution of competing boundary claims and provides the unbanked with property rights that are the basis for access to credit and finance.

**Solution Spotlight:** In **Kosovo**, UAVs are being tested to **secure property rights**. The drone imagery is processed in real-time and used in combination with a handheld tablet to record boundaries in an online system to formally register land. The entire process is completed in hours and days, as opposed to weeks and months with traditional survey methods.

**Solution Spotlight:** The **South African** government has worked with Digital Globe to use high-frequency and high-resolution satellite imagery feeds along with mobile phone applications to geo-locate each residential dwelling with an address. This is part of a broad government effort to digitalize census and household survey processes for planning and budgeting.

**PUBLIC SECURITY:** Police forces are drawing on big data and predictive analytics to make better policing decisions. Basic information such as crime type and location can help officers make smarter decisions on patrol. Law enforcement agencies are also using big data, through social media analytics companies such as Crimson Hexagon, to be more effective in public outreach campaigns.

**Solution Spotlight:** **PredPol** is a predictive policing application used in 50 US cities that analyzes basic data, such as crime type, dates and location, to help law enforcement agencies make better policing decisions. Brazil, the UK and the Netherlands use similar approaches.

**Solution Spotlight:** In Bogota, **Colombia**, World Bank researchers are using big data analytics and risk terrain modeling to understand relationships between crime and public infrastructure such as bus stations, public hospitals, schools and drugstores.





# Smarter Policymaking

The role of information in policymaking is a big concern in our information age. Big data is a viable source of high-frequency and granular data that can provide profound insights into human mobility and economic behavior, to better inform policy decisions. The entire globe is now imaged by satellites on a daily basis, with quality and affordability continuously improving. New public-private arrangements, like the Open Transport Partnership and the Workforce Data Initiative, are emerging to make data from social media, professional networks, mobile phones and sensors readily available for policymakers.

Traditionally, policy is based on reports that are largely informed by traditional survey and administrative data and statistics – instruments that are slow and labor intensive. In contrast,

big data can potentially deliver policy insights on a more frequent, disaggregated and cost-effective basis. Traditional surveys – like censuses, official statistics and enterprise data – will always be needed, but it is likely that robust big data proxies will continue to emerge and become more integrated into policymaking processes and decisions. The data from satellites, mobile phones and social media alone can change the dynamic between information and policymaking.

**TRANSPORT AND URBAN PLANNING:** Satellite imagery, cell phones, vehicle sensors, video feeds and social media are being used by policymakers and planners for traffic and urban planning. Remote sensing, mobile phone data and machine learning can provide policymakers and planners with much better understanding of urban mobility, land usage and urban change. Policymakers now and in the future can use these insights to provide access to jobs and make cities more sustainable, inclusive, productive, resilient and livable.

**Solution Spotlight:** **Big Pixel** is among several satellite-based applications that are enabling policymakers to have

systematic understanding of where urbanization is occurring, to what degree and how quickly, showing what impact policy interventions like zone changes and transport corridors have on urbanization.

**Solution Spotlight:** **OpenTraffic** is a global platform that uses anonymized positional data from vehicles and smartphones to visualize real-time and historic traffic conditions. The **Uber Movement** platform aggregates Uber vehicle data to inform urban infrastructure decisions.

**Solution Spotlight:** In Seoul, South Korea, the **OWL bus** uses big data analytics to better serve the needs of nighttime travelers. The project used data from three billion call and text data points, along with five billion points from corporate and private taxi data, to design nighttime bus routes optimized to match the origin and destination of passengers' journeys.

**OFFICIAL STATISTICS:** In many countries, census data is incomplete and outdated. Accurate and reliable information on basic factors such as population and poverty are surprisingly lacking for much of the world. A consensus is emerging in the statistics community that big data and machine learning can be used in certain instances to create viable statistical proxies for monitoring and policy decisions.

**Solution Spotlight:** **Facebook Connectivity Lab** is working in collaboration with the Center for International Earth Science Information Network to produce high-resolution **population Maps**. **Flowminder** is using satellite, mobile phone and household survey data to create **dynamic population maps**.

**Solution Spotlight:** The **Billion Prices Project (BPP)** started at the **MIT Sloan School of Management** aggregates price information from a multitude of online retailers around the world and gives real-time **inflation** predictions.

**Solution Spotlight:** Stanford's sustainability and artificial intelligence lab has developed a **satellite imagery and deep learning** approach to predict poverty. The World Bank is also experimenting with satellite-based methods that complement traditional methods for small-area poverty mapping.



**FOOD SECURITY:** Big data tools are enabling sophisticated analysis of weather and satellite data to analyze crop strategies and formulate optimal agricultural policies.

**Solution Spotlight:** Lobell Labs has developed a **scalable crop yield mapper** that uses satellite-based measurements to predict crop yields in both commercial and smallholder farms. The solution has been tested in the US, as well as in smallholder systems in Africa and India.

**Solution Spotlight:** Information on food prices can be crowdsourced through social media. In Indonesia, Global Pulse has tested social media and crowdsourcing to track **food commodity** prices in near-real-time in areas where other data sources are limited.

**EMPLOYMENT SERVICES:** Government labor agencies are experimenting with big data to inform the most appropriate policies to help individuals back to work, such as tailoring training services for different segments of job seekers.

**Solution Spotlight:** LinkedIn is working with the Australian government to mine data from the LinkedIn **economic graph** to identify trends such as the increasing demand for technology workers to also possess soft skills. These insights will help policymakers to create programs giving Australian workers the right skills and competencies.

**Solution Spotlight:** The World Bank is working with Flowminder in Haiti to analyze mobile phone data to provide Haitians with better information on and access to jobs and opportunities.



**ENVIRONMENTAL POLICY:** Analysis of environmental and climate data from multiple sources is enabling authorities to understand environmental impacts and interventions, from regional, national and global perspectives.

**Solution Spotlight: Global Forest Watch** combines satellite imagery, crowdsourced witness accounts and public datasets to track deforestation around the world, believed to be a leading manmade cause of climate change. The project helps ethical businesses to ensure that their supply chain is not complicit in deforestation.

**Solution Spotlight: Microsoft's Madingley** is a next-generation model of ecosystems and biodiversity across the globe. Madingley provides a working simulation of the global carbon cycle, and aims to model everything from deforestation to animal migration, pollution and overfishing, in a real-time "virtual biosphere."





systems of government will embrace the idea of shared greater accountability to varying degrees. Making government services and policies more responsive to citizen feedback will require strong relationships between policymakers, service providers, civic organizations and citizens.

**Solution Spotlight:** After Haiti's 2010 earthquake, a crowdsourced application developed by **Humanitarian Open Street Map** soon became the default tool for search and rescue teams. More than 600 volunteers traced roads and encampments from aerial images into a computer program. Crowdsourced markers identified resources such as refugee camps and cholera clinics.

**Solution Spotlight:** In **Brazil**, researchers analyzed tweets made during protests surrounding the 2014 Soccer World Cup, to understand citizens' feelings about policies and government. The analysis showed citizens expressing negative sentiment about the national government's low investment in basic services compared to lavish spending on the World Cup – suggesting domestic policy priorities for Brazil's government.

**Solution Spotlight:** The **Ushahidi platform** was created to monitor Kenya's general elections. It has since been used around the globe by citizens to monitor and support solutions for elections, civic problems and crises.







# Putting BIG DATA into Action for Government

The age of big data creates new challenges and opportunities.

Governments need to develop strategies, tools and forms of engagement to better understand dynamic forces and respond accordingly. The solutions featured in this brief show how big data can tackle fraud and corruption, generate administrative savings, and improve service delivery and policymaking processes, making them smarter, more accountable and more responsive to citizen feedback.

However, to realize value from big data, governments must strengthen technical and legal frameworks to access and use data responsibly. Importantly, they need to develop integrated capabilities to put big data insights into action, and to be responsive to citizen feedback on services and policies. Big data innovation is a rich process of learning that will require

perseverance as solutions are tested, adapted and mainstreamed – but the potential value from effectively using big data should make the effort worthwhile.

Big data has the potential to help shape government of the future. However, leveraging the opportunities brought about by big data also requires a cultural shift within the public sector. Data needs to be seen as an asset rather than a byproduct of administrative activity, in order to be valued, curated and shared where appropriate. This could be part of a broader movement to transform the public sector away from the bureaucratic model inherited from the 19<sup>th</sup> century and towards a more agile, 21<sup>st</sup>-century version. Such a transformation could be key to restoring citizen trust in government, which has significantly eroded in recent years.

Several key areas require government action if big data is to be used effectively for better government:

1. **Provide clear regulation and guidelines for data use:** Governments need to provide legal and policy guidelines on data

ownership, quality and sharing, privacy, civil liberties and equality. They must make decisions on privacy in conjunction with public opinion, informing citizens about the trade-offs between the privacy and security risks of sharing data, versus the benefits.

2. **Foster Public-Private Partnerships:** The majority of high-value data assets reside with the private sector. Public-private partnerships can coordinate governments, businesses and nonprofits to make big data accessible for the public good. The Global Partnership for Sustainable Development Data is a multi-stakeholder network of over 150 organizations to promote the sharing and responsible use of private-sector data.

3. **Promote transparency in algorithms:** Governments must encourage good practices in the design and documentation of algorithms, including transparency and accountability mechanisms to correct and contest decisions based on algorithms.

4. **Invest in big data capacities:** Many datasets span departments, meaning governments need to manage and integrate datasets from departmental silos. Big-data 'control centers' are needed for implementing strategies and best practices, and providing essential resources in advanced computing technologies. Governments also need change-management and data science skills to adapt process, systems and practices to put big data to action.



# Additional Resources

*Useful starting points on the use of big data to improve government include:*

**Big Data Challenge, Pioneering Data Driven Approaches to Development:** World Bank report that profiles the application and lessons of several World Bank big data applications in the development sector.

**United Nations Global Pulse:** An initiative to promote awareness of opportunities to “harness big data for development and humanitarian action.” It is a network of innovation labs that coordinate research and “strengthen the big data innovation ecosystem.”

**UN International Conference on Big Data for Official Statistics 2016:** Proceedings of the 2016 conference in Dublin. “Going beyond the existing examples of isolated use of mobile phone data, social media data or satellite imagery data, innovations are needed in the daily production of official statistics, which require real partnerships with the private sector, new skills and infrastructure, and clear links between these available Big Data sources and the SDG indicators.”

**UN Big Data Global Working Group: Big Data Project Inventory:** A catalog of Big Data projects that are relevant for official statistics needed for decision-making on public policies, as well as for management and monitoring of public sector programs/projects. It is a joint project of the World Bank and United Nations Statistics Division.

**Big Data Applications in the Government Sector: Communications of the ACM** [Association for Computing Machinery], March 2014. Good overview of business and government use of big data, including issues of volume, velocity, variety, and security. The public sector is learning from fast-growing commercial use of big data and analytics.

**Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights:** Executive Office of the President, May 2016. “This report examines several case studies from the spheres of credit and lending, hiring and employment, higher education, and criminal justice to provide snapshots of opportunities and dangers, as well as ways that government policies can work to harness the power of big data and avoid discriminatory outcomes.”

**The Federal Big Data Research and Development Strategic Plan:** Written by the Big Data Senior Steering Group, Subcommittee on Networking and Information Technology Research and Development, for the National Science and Technology Council, Executive Office of the President, United States.

**Pathways for Progress, Data Driven Development:** World Economic Forum report on vision of how Big Data might be used to address the challenges of sustainable development.

**Bringing probability judgments into policy debates via forecasting tournaments:** A tournament sponsored by the US intelligence community revealed ways in which forecasters can better use probability estimates to make predictions—even for seemingly “unique” events—and showed that tournaments are a useful tool for generating knowledge.

**Beyond prediction: Using big data for policy problems:** Machine-learning prediction methods have been extremely productive in applications ranging from medicine to allocating fire and health inspectors in cities. However, there are a number of gaps between making a prediction and making a decision, and underlying assumptions need to be understood in order to optimize data-driven decision-making.

**World Development Report, 2016: Digital Dividends:** This report looks at digital opportunities of the new economy and how countries can prepare by ensuring that institutions are accountable.

**World Development Report, 2017, Governance and the Law:** This report looks at fundamental questions which are at the heart of development, policymaking and policy implementation.

# Contact

This solutions brief is an input into the 2017 World Government Summit workshop and high-level panel on Big Data for Government. It was produced in collaboration with the World Bank Governance Practice and *Innovations in Big Data Analytics* program, which resides within the Global Operations and Knowledge Management unit of the World Bank.

For additional information, please contact Trevor Monroe ([tmonroe@worldbank.org](mailto:tmonroe@worldbank.org)).



