

Data in Government
Creating Data-Driven Systems for State Effectiveness & Accountability
By Pat Austria

Introduction

The proliferation of technology and data has enabled new pathways between citizen and government. It has changed both state capacity and citizen expectations. However, the promises of the data revolution have yet to translate to tangible performance improvements or economic value for many governments. A report by McKinsey found that open data could generate more than \$3 trillion in global economic value and that public sector data alone is worth between 0.4% and 1.5% of an economy's GDP.

While interest and excitement abound, governments have been slow to fully integrate data in their processes. In 2014, UN Secretary-General Ban Ki-Moon commissioned a group of independent advisors to explore entry points for integrating data and sustainable development. In their report, they called upon "governments and the UN to act to enable data to play its full role in the realisation of sustainable development by closing key gaps in access and use of data: between developed and developing countries, between information-rich and information poor people, and between the private and public sectors."

Coordination between countries and groups remains incredibly important but it overlooks the critical role of the state ecosystem. Government have yet to fully leverage the power of data and yet many assume that government bodies are ready and capable partners in the data revolution. While investment in open data occurs, delivering tangible impact for governments and citizens demands more transformative changes. Leveraging data for a more effective state requires the creation of a robust and dynamic data ecosystem. It is a critical foundation for sectoral data applications and the primary step in creating more data-driven governments. The following paper examines two key areas in which the state must engage with data: (1) laying foundations for sustainable data ecosystems and (2) delivering value within state functions.

Laying the Foundations: Building Scalable and Sustainable Data Ecosystems

Governments must be ready to challenge their conventional practices to invest in efficient and robust data ecosystems. While innovation and technology are traditionally pursued rapidly and iteratively, governments must adopt a whole of system approach to data to ensure that data is responsibly and sustainably utilized.

1) Create a cohesive data strategy for a whole-of-government approach

Data collection occurs in numerous facets of modern life and databases are often created in isolation. Absent a data strategy, many states find themselves with incongruent databases that reinforce the traditional silos of bureaucracy. The strategy should include strategic ambitions, but also processes and protocols such as standard file formats, platforms and reporting schedules.

Source: Global Open Data Index¹

Box 1. The top findings from the 2015/2016 Global Open Data Index report were deeply rooted in the incongruity of government datasets.

- *Data findability is a major challenge.* We have data portals and registries, but government agencies under one national government still publish data in different ways and different locations. Moreover, they have different protocols for licenses and formats. This has a hazardous impact - we may not find open data, even if it is out there, and therefore can't use it. Data findability is a prerequisite for open data to fulfill its potential and currently most data is very hard to find.
- *A lot of 'data' IS online, but the ways in which it is presented are limiting their openness.* Governments publish data in many forms, not only as tabular datasets but also visualizations, maps, graphs, and texts. While this is a good effort to make data relatable, it sometimes makes the data difficult or even impossible for reuse. It is crucial for governments to revise how they produce and provide data that is in good quality for reuse in its raw form. For that, it is crucial to be aware what best form of raw data is needed, which varies from category to category.
- *Platforms: Encouraging complimentary or single platforms fosters greater compatibility across ministries.* This helps avoid duplicative or incongruent initiatives, but also allows governments to leverage economies of scale when signing contracts with technology providers.²
- *Processes: Harmonizing interactions between government processes and data collection is critical to promoting usability.* Government data should reflect current government ontologies and be crafted with the input of department heads and topical experts. To the extent possible, data inputs and outputs should be easily integrated into established government processes. This includes using existing platforms as data sources and making data easily integrated into established processes, reporting schedules, and data formats. For example, if an office's standard is Microsoft Excel, all databases should have an option to download an Excel version of the data..
- *Regularity: Data utility is exemplified when it can be compared across time.* This allows for greater insight on historical trends and predictive progress, but it demands regularity and diligence in reporting. Stakeholders are motivated by different incentives to collect and analyze data. It is imperative that reporting expectations are clear and that guidelines exist to ensure uniformity or comparability. One way that many governments are promoting regularity is making "open" the default for government datasets.

Creating a whole-of-system strategy seems intuitive but it can be in tension with conventional practices around innovation. Innovation is typically marked by fast and iterative delivery, and managers in the private sector are often told to allow innovation to occur without restraint. In the public sector, this would mean promoting data use in any way individual departments see fit. By contrast, a whole-of-system data strategy calls upon the public sector to think systemically and promote greater uniformity across different groups. While more difficult in the short-term, this can help governments leverage the value of data across departments by fostering compatibility and regularity early on.

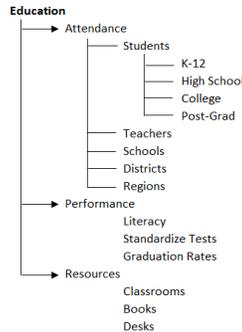
¹ Global Open Data Index, "The GODI 2016/17 Report," <https://index.okfn.org/insights/>

²Tomer, Adie and Ranjitha Shivaram, "Modernizing government's approach to transportation and land use data," The Brookings Institution, https://www.brookings.edu/wp-content/uploads/2017/07/modernizingapproachtodata_report1.pdf

2) Develop a common taxonomy and benchmarks

Creating robust data ecosystems require a whole-of-government approach that is built upon a common taxonomy. Taxonomies should be multi-tiered to balance specificity and applicability, with detail provided through topical tagging such as themes or focus areas. These classifications should be specific to the government and developed with the current institutional structure in mind. Creating this taxonomy can be a highly politicized process but is critical to ensuring organization and accessibility.

Figure 1. Sample Taxonomy and Tagging



Database Name	Tier 1	Tier 2	Tier 3	Tier 4	Lead Ministry	Source	School Year	Uploaded On
Kerala Standardized Test 2009 Scores	Education	Performance	Standardized	TK-12	Ministry of Education	Government	2009-2010	2/20/2018
Student Attendance Region IV 2010	Education	Attendance	Students	K-12	Ministry of Education	Government	2010-2011	5/4/2014
Desk Purchases Region IV 2009	Education	Resources	Classrooms	High School	Ministry of Education	Government	2009-2010	3/15/2017
Female Literacy Rates	Education	Performance	Literacy	High School	Ministry of Education	World Bank	2010-2013	11/23/2017

Taxonomies can also be a powerful tool for making strategic ambitions actionable. Operationalizing the lofty ambitions of a strategic vision are difficult for any administration. Specific dashboards can be created to track progress on policy objectives by providing a platform for measurement and accountability at lower management levels.

Figure 2. Example Outcome Indicator Structure

Outcome	Output	Input	Lead Ministry
President's Strategic Goal: Achieve gender equality	Improve access and quality of women's health	85% of births attended by healthcare professional	Ministry of Health
		Reduce incidents of sexual violence by 40%	Ministry of Justice
	Improve valuable and gainful opportunities for women	Bring girls' school participation rate to 76%	Ministry of Education
		Increase percentage of women managers by 30%	Ministry of Commerce

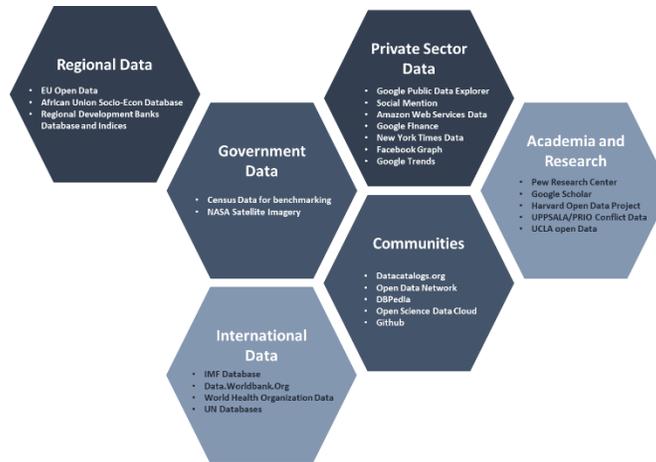
In addition to indicator frameworks, benchmarks are a critical component of effective data systems. Benchmarks provide context for data points, whether it be a historical starting point or a comparative indicator such as regional or global averages. Having benchmarks alongside data points allows information to be more descriptive and valuable to users.

3) Do not be limited to traditional data sources

Developments in analytics, mobile, and geospatial technology have created new sources of data. Fostering innovation in government necessitates establishing partnerships across data streams and sectors. Using other data sources to complement or replace government collection can help avoid duplication and free up resources for other initiatives. Figure 3 highlights several open data sources that can be used within a government's data ecosystem. Many of these resources also

use Application Programming Interfaces (APIs) that can automate updating of the data to reduce the burden on government resources.

Figure 3. General Sources of Open Data



Through consumer electronics and web platforms, the private sector owns a sizable portion of valuable data. Using these avenues, private companies have the greatest access to citizen perception and behavior. While governments almost never require private sector as an intermediary in the fulfillment of the social contract, but the changing dynamic is a necessary consideration in creating effective and responsible government data ecosystems.

4) Balance aggregation and granularity to promote utility and protect privacy

Technologies can be downloaded over the internet and citizen insights can be collected in real-time. These are far divergences from traditional capital acquisition and survey methodologies. As such, traditional regulatory models should be reformed to reflect modern capacities and liabilities. Simply applying traditional processes can slow innovation and undermine citizen rights.

Source: McKinsey Blockchain for Government ³

Box 2. Blockchain for Citizen Data Privacy

Estonia, a country at the forefront of incorporating technology in government processes, uses Keyless Signature Infrastructure (KSI) to safeguard public-sector data. “KSI creates hash values, which uniquely represent large amounts of data as much smaller numeric values. The hash values can be used to identify records but cannot be used to reconstruct the information in the file itself. The hash values are stored in a blockchain and distributed across a private network of government computers.”

Responsible data governance is particularly important when managing data at various levels. The promulgation of granular data holds many opportunities, but it also creates dangerous avenues for misuse. Finding the right balance between aggregation and granularity is critical for privacy and greatly aids analysis. Localized data can allow for focused and actionable

³McKinsey, “Using blockchain to improve data management in the public sector,” <https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/using-blockchain-to-improve-data-management-in-the-public-sector>

insights, but it also can be used to target or discriminate against individuals or groups. Multi-tier views allow for different users to explore specific cuts of the data. For example, while directors may be interested in larger demographic trends, more granular data can help local managers craft focused interventions for improvement.

5) Promote utilization through an enhanced user experience and trained data champions

Fostering a data-driven culture requires having data champions who can integrate data within an office's strategies, processes, and decision-making. These champions can be focused data-scientists or simply high-performing individuals that are trained specifically in data analysis and platforms. Regardless of training, integration in the ministry is critical. A survey from the Pew Research Center indicated that most Americans believe that the value of open data was in providing journalists and researchers better information to hold governments accountable. In that same survey, very few respondents had confidence that data would lead to improvements in government performance.⁴ While a well-informed media is foundational to an effective and accountable state, it does not discount the importance of governments using the data within their own systems and processes.

To capitalize on the potential power of data in government, data must inform decisions and policy. To do so, a data system must not only have openness but accessibility, findability, and utility. Creating a tagging taxonomy, incorporating a search function, and outlining important data components like publication year can help users access and analyze the data more easily. Data stories can foster momentum for maintaining interest and investment in data, highlighting scalable stories of success, and identifying areas of inefficiency.

Limitations of Data

Leveraging the power of data requires a fundamental understanding of its limits. Data should be enabling not divisive – it must be used to inform decision-making and analyzed within other contextual factors. Some fundamental limits include:

- *Indicators are largely informed by a Western perspective of progress.* Traditional development indicators should be analyzed, critiqued, and augmented to reflect the benchmarks most useful to local policy-makers.
- *Data is limited in its ability to capture cultural nuance.* Governments should continue to use qualitative information and local expertise to provide a more comprehensive view. The importance of social nuances should not be foregone in the pursuit of data analytics and trends.
- *Data and analytics can have intrinsic biases toward certain groups.* Data collection through technological devices are particularly vulnerable to elite capture given that digital literacy is concentrated in areas with higher income or educational attainment. In addition, rudimentary analytical models that search for largest swaths of people will favor urbanized populations over marginalized ones. For example, analytics may promote service expansion to only urbanized populations given that the largest amount of people live in these areas. Governments should be purposeful in building analytical models that not only maximize cost-benefit, but equitably serve even the most marginalized sections of the population.

⁴ Pew Research Center, "Americans' Views on Open Government Data," <http://www.pewinternet.org/2015/04/21/open-government-data/>