McKinsey & Company

Public & Social Sector Practice

Government data management for the digital age

Public institutions can make more of their data resources. Five actions can help them modernize their data infrastructures and unlock significant value across state, economy, and society.

by Axel Domeyer, Solveigh Hieronimus, Julia Klier, and Thomas Weber

September 2021

Digital society's lifeblood is data—and governments have lots of data, representing a significant latent source of value for both the public and private sectors.[1] If used effectively, and keeping in mind ever-increasing requirements with regard to data protection and data privacy, data can simplify delivery of public services, reduce fraud and human error, and catalyze massive operational efficiencies.

Despite these potential benefits, governments around the world remain largely unable to capture the opportunity. The key reason is that data are typically dispersed across a fragmented landscape of registers (datasets used by government entities for a specific purpose), which are often managed in organizational silos. Data are routinely stored in formats that are hard to process or in places where digital access is impossible. The consequence is that data are not available where needed, progress on digital government is inhibited, and citizens have little transparency on what data the government stores about them or how it is used.

Only a handful of countries have taken significant steps toward addressing these challenges. As other governments consider their options, the experiences of these countries may provide them with valuable guidance and also reveal five actions that can help governments unlock the value that is on their doorsteps.

As societies take steps to enhance data management, questions on topics such as data ownership, privacy concerns, and appropriate measures against security breaches will need to be answered by each government. The purpose of this article is to outline the positive benefits of modern data management and provide a perspective on how to get there.

Interoperable and connected government data offer significant benefits

The COVID-19 crisis has highlighted the importance of data, which have been at the heart of managing the impact of the pandemic in many countries. Germany, for instance, uses data to track ICU beds and manage spikes in hospitalizations.[2] Meanwhile, some other countries, including the United Kingdom, Ireland, Portugal, and South Korea, have developed dashboards to help decision makers track the ups and downs of the pandemic.[3]

Of course, the value of public-sector data extends beyond the pandemic to numerous aspects of society and the economy. The challenge for many governments is accessibility, which is often restricted by a reliance on multiple registers dedicated to narrow purposes. Better linking of information can yield significant benefits.

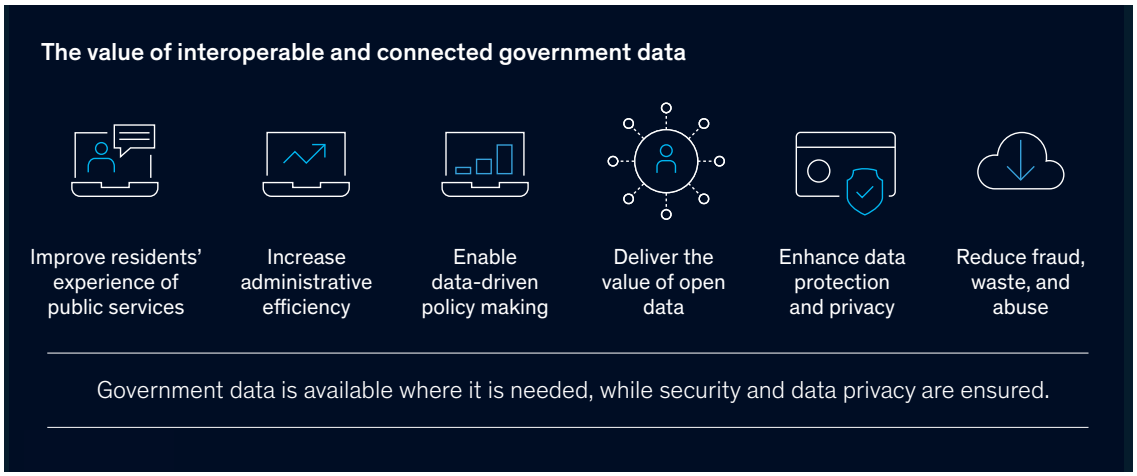
To unlock their data potential, governments can develop an *interoperable and connected data landscape*, in which data collected by any government entity are available where needed, where security and privacy are safeguarded, and where adequate measures (legal, technical, and organizational) prevent misuse of data. If governments can achieve that, there are benefits in six key areas (Exhibit 1).

— *Improve resident experience.* When accessing a public service, citizens and companies often need to provide data and documents that they have already shared. If data collected across government were more accessible, public services could follow a "once only" principle, meaning data must be submitted just a single time, saving time and reducing manual inputs. Another benefit would be the ability to deliver services proactively, with new data automatically triggering a response where required. Estonia has this functionality up and running—for example, the registration of a newborn child automatically leads to the provision of childcare benefits, with data from the tax registry determining how much money should be transferred to which bank account.[4]

[1] For example, the European Data Portal, which gathers public-sector information of the EU27+, currently holds about 1.1 million datasets.
[2] Rodolfo Catena and Matthias Holweg, "We need to relocate ICU patients out of COVID-19 hotspots," *Harvard Business Review*, June 23, 2020, hbr.org.
[3] Observatory of Public Sector Innovation (OPSI), "OPSI COVID-19 Innovative Response Tracker," Organisation for Economic Co-operation and Development, September 2020, oecd-opsi.org.
[4] Jana Silaškova and Masao Takahashi, "Estonia built one of the world's most advanced digital societies. During COVID-19, that became a lifeline," World Economic Forum, July 1, 2020, weforum.org.

Exhibit 1

If governments can achieve an interoperable and connected landscape, the benefits include more efficiency, usability, and value-creating opportunities.



— *Increase administrative efficiency.* Government employees are often required to obtain data manually, both from other government agencies and citizens. Interoperable and connected data would allow governments to streamline this "back end," reducing friction and cutting clearing times. The potential benefits would be substantial in both public-service delivery and periodic activities such as the census. In Germany, fully interoperable and connected government data were estimated to produce a 60 percent reduction in case-processing time for key public services.[5] For the census, technologically advanced countries, such as the Netherlands, pull data entirely from existing databases. This approach incurs up to 99 percent less costs than a traditional survey-based method.[6]

— *Enable data-driven policy making.* High-quality and available data have a positive impact on policy making. In Denmark, for example, the government uses geodata to simulate flooding scenarios, enabling both better crisis management and improved long-term-investment decisions.[7] To build the complex models required, data from several key registers (for example, cadastres, buildings, and addresses) are combined with 3-D topographic data. Research has shown that these kinds of applications produce excellent cost–benefit outcomes.[8]

[5] *Mehr Leistung für Bürger und Unternehmen: Verwaltung digitalisieren. Register modernisieren* [More performance for citizens and companies: Digitize administration. Modernize registers], Nationaler Normenkontrollrat, October 2017, normenkontrollrat.bund.de.
[6] Eric Schulte Nordholt, "Usability of administrative data for register-based consensus," *Statistical Journal of the IAOS*, January 2018, Volume 34, Number 4, pp. 487–498.
[7] "Havvand på land" [Seawater on land], Klimatilpasning, October 7, 2019, klimatilpasning.dk.
[8] Andrew Coote et al., *Assessing the economic value of 3D geo-information*, EuroSDR, November 2017, Official Publication Number 68, eurosdr.net.

— *Deliver the value of open data.* Governments can play a key role as data providers—both in the form of raw data and official statistics—helping to unlock a $3 trillion open-data opportunity for the private sector and civil society.[9] However, many datasets published on open-data portals are taken from information silos. Interoperable and connected registers, with appropriate safeguards, allow for the publication of more comprehensive and insightful datasets.

— *Enhance data protection and privacy.* To some, interoperable and connected government data raises the specter of a "Big Brother" government. However, the current management of government data also poses challenges in terms of data privacy. Citizens are often unable to see their personal data or know where these data are stored and when and why they are accessed. A modern data landscape, by contrast, enables privacy by design. Structured and secure data exchanges reduce the number of people in contact with data and the risk of leaks. Citizens can benefit from more transparency and active management of consent. In addition, governments can show what data are saved, and where, and provide a log of digital interactions. That enables users to opt in or out of use cases. Estonia's data tracker, for example, allows citizens to review data queries relating to their personal information, including the reason for access.[10]

— *Reduce fraud, waste, and abuse.* A substantial share of government payments result from error or fraud—funds go to the wrong recipient, an incorrect amount is transferred, or a government payment is used improperly. In the United States, agencies across the government made an estimated $175 billion in improper payments in 2019.[11] Interoperable and connected government data can help mitigate loss risk by reducing errors from manual inputs and inconsistent data across registers and by enabling governments to leverage analytics tools that identify fraud. Estonia, which is leading in this area as well, combines information from agriculture registers with satellite images to analyze whether land subsidized by government grants is cultivated.[12]

Government data today: Scattered, siloed, inaccessible

The COVID-19 pandemic has highlighted the challenges of using government data. For example, aggregating case numbers from laboratories and hospitals has often involved communication via email, phone, and fax.[13] Accessing and merging data collected by different government institutions has proven to be extremely challenging. In most cases, it involves individual requests, manual processes, or customized APIs.

While some countries have started to make progress, government data management is typically hamstrung by three challenges:

— *Data are scattered.* To use data stored in the register, agencies need to be able to identify that data. That is rarely possible. Indeed, governments are often unable to discern which data are in which register, or even which registers exist. This means there is little transparency on whether a specific data point is available somewhere in the government, whether it is available in multiple registers, or where the most current data can be found.

[9] Michael Chui, Diana Farrell, and Kate Jackson, "How government can promote open data," April 2014, McKinsey.com.
[10] Federico Plantera, "Data tracker—tool that builds trust in institutions," e-Estonia, September 2019, e-estonia.com.
[11] "Improper payments: Issue summary," US Government Accountability Office, gao.gov.
[12] "AI–'kratt' strategy," e-Estonia, April 2020, e-estonia.com.
[13] Andreas Stiller, "Coronavirus-Fallzahlen und der Amtsschimmel" [Coronavirus case numbers and the red tape], Heise Medien, March 16, 2020, heise.de; Sarah Kliff and Margot Sanger-Katz, "Bottleneck for U.S. coronavirus response: The fax machine," *New York Times*, July 13, 2020, nytimes.com.

— *Data cannot be accessed digitally.* Many registers are still paper based, which creates a significant cost and administrative burden and renders systematic usage impossible. Even where digitized registers are used, access is often not standardized. For other agencies to access the data, additional infrastructure, such as secure connections or APIs, is required.

— *Data are not interoperable.* There are a number of obstacles that prevent the combination and joint processing of data stored across different agencies (for more, see sidebar "Five obstacles to data interoperability").[14]

A handful of governments have begun to deliver

Many governments around the world have made interoperable and connected data a top priority. In 2020, the United Kingdom published its national data strategy, which includes the mission to transform government's use of data to drive efficiency and improve public services based on an appropriately safeguarded, joined-up and interoperable data infrastructure.[15] In Germany, the federal government has launched a national program to modernize the public-sector data landscape.[16] Still, only a short list of countries (including Estonia, Denmark, and the Netherlands) have made significant progress on these kinds of

[14] Directorate-General for Informatics, *New European interoperability framework: Promoting seamless services and data flows for European public administrations*, European Commission, November 2017, op.europa.eu.
[15] *UK National Data Strategy*, GOV.UK, December 9, 2020, gov.uk.
[16] "Digitalisierung der Verwaltung schreitet voran" [Digitization of administration is advancing], Bundesministerium des Innern, für Bau und Heimat, March 12, 2019, bmi.bund.de; "Eckpunkte des Konjunkturprogramms: Corona-Folgen bekämpfen, Wohlstand sichern, Zukunftsfähigkeit stärken" [Key points of the economic stimulus program: Combat the consequences of corona, secure prosperity, strengthen future viability], Bundesministerium der Finanzen, June 3, 2020, bundesfinanzministerium.de.

Five obstacles to data interoperability

The combination and joint processing of data stored across different agencies typically faces five major obstacles. All of these can be addressed by making data "interoperable."

— *No uniform legal framework for using data.* Regulation often only covers specific applications, which means that building use cases for government data is slow and uncertain. In addition, it is often unclear which data agencies can legally share with each other and under what specific conditions and safeguards. Legal interoperability would comprise a uniform legal framework to control when data can and cannot be accessed, exchanged, or combined.

— *No connected view of data.* Data about a single citizen, company, or building stored in different registers are often not connected—even though they concern the same object. Substantial interoperability means that there would be a link between entries, typically through unique identifier numbers associated with a citizen, company, building, or entity.

— *No "data provider" mindset.* Government agencies typically do not see themselves as data providers that generate value for citizens and corporations by managing and storing their data. Organizational interoperability means that agencies could view themselves as service providers that enable secure and reliable access to anyone who has the right.

— *No consistent logic across data.* Different agencies may, for example, store data on companies but follow different semantic conceptions—such as production sites, legal entities, or headquarters—rendering combined use impossible. Semantic interoperability means that there would be a shared logic that ensures the precise meaning of exchanged information can be interpreted unambiguously by different systems.

— *No uniform technical format.* Government data cannot be combined and connected when they are stored in different formats. Technical interoperability means that data shared between different databases could be accepted and processed without the need to change data format or other characteristics.

endeavors (for more, see sidebar "International examples of successful integrated data-management implementations").

How to succeed: Five actions to deliver interoperable and connected government data

Despite the challenges, there are manifest benefits to be gained from interoperable and connected government data. To support the transition, governments can deploy five actions, informed by the experiences of pioneering countries.

1. Set a clear vision based on tangible use cases

If governments were to redesign their data landscapes entirely, the task would be too large. They should therefore begin by sketching out a vision that focuses on clear and tangible use cases. The Austrian government, for example, decided in 2000 that the following year's census would be the last nondigital version and began preparations for its first digital census in 2011.[17] Denmark focused on improving quality and access for a small set of frequently used "grunddata"—basic data—regarding people, organizations, and places.[18] In this context, initial use cases should be representative of a larger set of relevant applications (for example, comparable government services) to ensure transferability later on and to avoid developing initiatives with a too narrow focus.

[17] Thomas Körner et al., *Registernutzung in Zensus und Bevölkerungsstatistik in Österreich und der Schweiz* [Use of registers in census and population statistics in Austria and Switzerland], Statistisches Bundesamt, October 2017, normenkontrollrat.bund.de.
[18] *The digital path to future welfare: eGovernment strategy 2011–2015* (English version), The Danish Agency for Digitisation (Digitaliseringsstyrelsen), December 2011, digst.dk.

International examples of successful integrated data-management implementations

A small group of countries have made significant progress in making data interoperable and connected across the public sector:

— *Estonia* was one of the first countries to implement the "once only" principle. The Public Information Act, approved in 2000, prohibited the establishment of separate databases for the collection of the same data.[1] The Estonian government has unique identifiers for its citizens and companies and connects its registers via the X-Road data exchange system, which processes nearly 1 billion queries per year.[2]

— *Denmark* implemented the "grunddata" program to make basic data relating to people, companies, and buildings (as well as geodata and climate data) available free of charge to authorities, businesses, and citizens.[3] The Danish administration defined an overarching data model, describing which data are stored in what format and how different pieces of information are connected. It designed a central "data distributor" to enable access to data from 11 registers through a standardized interface. This component was also linked to public-service portals, enabling the once-only principle for all basic data.[4]

— *The Netherlands* integrated 12 base registers into one system, the "Stelsel van Basisregistratie," in 2003. The system contains general information, such as personal addresses, business names, and income. It made citizens' lives easier by combining data from different databases and providing prepopulated forms for tax declarations.

[1] "Public Information Act," Riigi Teataja, November 14, 2013, riigiteataja.ee.
[2] "Interoperability services," e-Estonia, January 11, 2021, e-estonia.com.
[3] "Grunddata" [Basic data], Digitaliseringsstyrelsen, September 2020, digst.dk.
[4] "Finansieringen af grunddataprogrammet er på plads" [The financing of the basic data program is in place], Digitaliseringsstyrelsen, December 12, 2016, digst.dk.

2. Understand and navigate the relevant data landscape

Once the vision is clear, governments can enable transparency over the relevant data landscape for the prioritized use cases. This requires mapping relevant registers, including semantic and technical characteristics. This mapping enables governments to "know what they know," that is, what data is available and where. Governments will see where critical data is lacking and where data is stored in multiple registers. In Estonia, this task was undertaken by the Information System Authority, which maintains RIHA—a catalog for government information systems. The catalog provides information on more than 2,600 information systems, portraying an overview of the purpose, storage, and management of the data.[19]

3. Offer relevant infrastructure components centrally

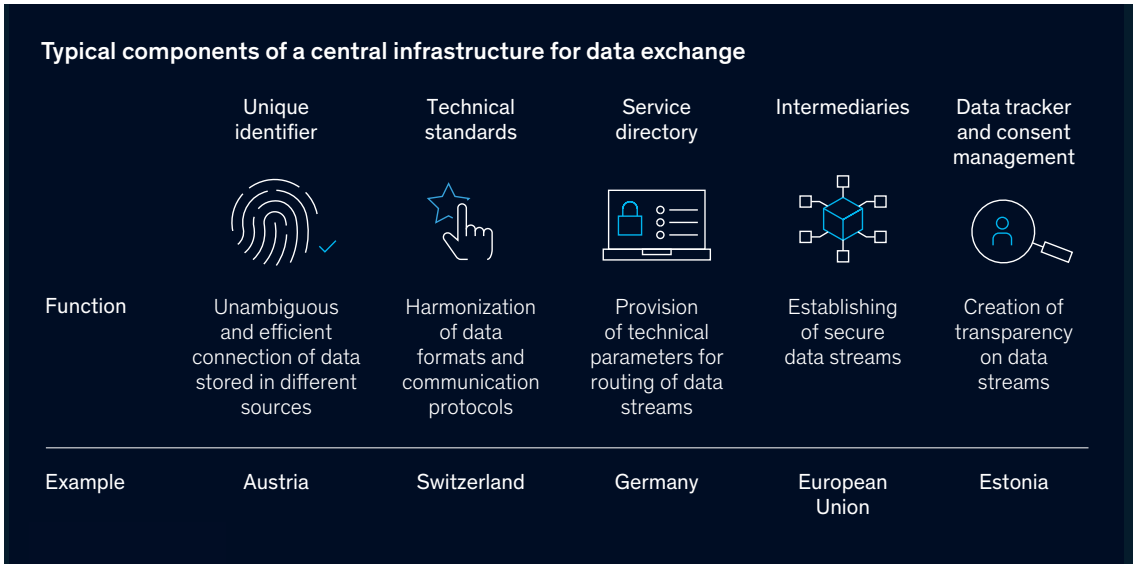
Governments can build a set of standardized components that enable data sharing for a wide range of use cases. In this way, government data stored in silos can be made interoperable and connected at scale. The leading example is the X-Road system, developed by the Estonian government in 2001.[20]

A central infrastructure for secure data exchange between public-sector entities and third parties will typically include at least five specific components (Exhibit 2):

— *Unique identifiers.* Identity management is key to interoperability. Numeric or alphanumeric identifiers (for instance, individuals,

Exhibit 2

A central infrastructure for secure data exchange between public-sector entities and third parties will include at least five specific components.



[19] "Administration system for the state information system RIHA," Republic of Estonia Information System Authority, November 12, 2020, ria.ee.; "Avaleht" [Home page], Riigi infosüsteemi haldussüsteem, September 2020, ria.ee.
[20] "X-Road® history," Nordic Institute for Interoperability Solutions, September 2020, x-road.global.

companies, or buildings) are necessary for the unambiguous and efficient connection of data stored in different sources to a single entity. The Austrian government established sector-specific identifiers for increased data protection.[21] In France, the government data-management team built on existing infrastructure and leveraged social security numbers as overarching identifiers.[22]

— *Technical standards.* Fast and automated data exchange is only possible through harmonized data formats and standards. In Switzerland, public databases storing similar information use consistent data formats across all repositories, and there is an established process for data exchange through a secure common protocol (called sedex or secure data exchange).[23]

— *Service directory.* To establish automated data exchanges, a directory holding the technical parameters of every data provider is required. This will deliver routing information to data consumers and vice versa. In Germany, the "Administrative Services Directory" serves as an interdisciplinary and cross-administrative infrastructure. It provides secure and reliable automation of services and procedures for communication between, and with, public entities.[24]

— *Intermediaries for secure data exchange.* To prevent inadmissible merging of personal data, data exchanges between government entities are often established via technical intermediaries. The intermediaries can only access the metadata of an exchange—who the sender is, who the recipient is, and the reason for a data exchange—but not its actual content, thereby helping to establish secure data streams. Sometimes, the function of intermediaries is performed by so-called gateways, which offer additional functions such as translating between different data formats and standards. Through secure gateways, registers can be made easily accessible to public-sector entities and third parties, enabling the government to become part of the API economy. For example, the PEPPOL infrastructure for cross-border eProcurement initiated by the EU uses secure "Access Points" as intermediaries to establish secure data connections.[25]

— *Data tracker and consent management.* Citizens expect to have transparency and control over how their governments use their data. This counterbalances the increased technical ease with which public agencies can access their information. Estonia is the pioneer in this area, too. Its data tracker allows citizens to review queries concerning their personal information, including the reason for access.[26] Germany is planning to go one step further and enable citizens to not only track queries but also give and withdraw consent on specific uses of their personal data.

4. Rapidly deliver end-to-end use cases via agile data labs

Despite the complexity and lengthy time horizon of a holistic effort to modernize the data landscape, governments can establish and sustain a focus on rapid, tangible impact. A failure to deliver results from the outset can undermine stakeholder support. In addition, implementing use cases early on helps governments identify gaps in their data landscapes (for example, useful information that is not stored in any register) and missing functionalities in the central data-exchange infrastructure.

[21] "Bereichsspezifische Personenkennzeichen (bPK)" [Area-specific personal identification (bPK)], Bundesministerium für Digitalisierung und Wirtschaftsstandort, September 2020, bmdw.gv.at.
[22] "Numéro d'inscription au répertoire/Numéro de sécurité sociale/NIR" [National registration number], INSEE, May 21, 2019, insee.fr.
[23] "sedex (www.sedex.ch)," Switzerland Federal Statistical Office, September 2020, bfs.admin.ch.
[24] "DVDV—das Dienstverzeichnis der öffentlichen Verwaltung" [DVDV – the public administration service directory], Informationstechnikzentrum Bund, September 2020, itzbund.de.
[25] "PEPPOL eDelivery Network—An overview," OpenPEPPOL, September 2020, peppol.eu.
[26] "Data tracker," September 2019.

To deliver impact quickly, governments may deploy "data labs"—agile implementation units with cross-functional expertise that focus on specific use cases. Solutions are rapidly developed, tested, iterated and, once successful, rolled out at scale. The German government is pursuing this approach in its effort to modernize key registers and capture more value.[27]

5. Establish a central data agency

Organizations such as Estonia's Information System Authority or Singapore's Government Data Office have played a critical role in transforming the data landscape of their respective countries. A central agency can pool scarce data talent and deploy it to implement projects. It can establish joint rules for data governance and data quality management and can own the IT architecture for a common data-exchange infrastructure. In some cases, it can develop and operate critical components. Finally, it can define data-management best practices for public-sector entities and support implementation.

Building an interoperable and connected government data landscape is a significant challenge. It requires substantial resources and can take time. In moving forward, however, governments can lay the groundwork for a fundamentally more effective and efficient public sector and digital society.

[27] "Eckpunkte für die Registermodernisierung" [Cornerstones for the modernization of registers], IT-Planungsrat, April 2020, it-planungsrat.de.

Axel Domeyer is an associate partner in McKinsey's Hamburg office; **Solveigh Hieronimus** is a senior partner in the Munich office, where **Julia Klier** is a partner; and **Thomas Weber** is an associate partner in the Berlin office.

The authors wish to thank Livia Boerner, Luca Flora, Simon Pfeiffer, Simon Prisner, and Frauke Renz for their contributions to this article.

Designed by McKinsey Global Publishing