Academy of ICT Essentials for Government Leaders

Module on

Realizing Data-Driven Governance

UNITED NATIONS
APCICT - ESCAP

Academy of ICT Essentials for Government Leaders

Realizing Data-Driven Governance

APCICT ASIAN AND PACIFIC TRAINING CENTRE FOR INFORMATION AND COMMUNICATION TECHNOLOGY FOR DEVELOPMENT

Academy of ICT Essentials for Government Leaders

Realizing Data-Driven Governance

Correspondence concerning this report should be addressed to the email: apcict@un.org

Contact :

Asian and Pacific Training Centre for Information and Communication Technology for Development (APCICT/ESCAP)

5th Floor G-Tower, 175 Art Center Daero, Yeonsu-gu, Incheon, Republic of Korea

Tel +82 32 458 6650

Fax +82 32 458 6691/2

Email apcict@un.org

http://www.unapcict.org

Printed in Republic of Korea

ISBN: 979-11-88931-02-6

ABOUT THE MODULE SERIES

In today's "Information Age", easy access to information is changing the way we live, work and play. The "digital economy", also known as the "knowledge economy", "networked economy" or "new economy", is characterized by a shift from the production of goods to the creation of ideas. This underscores the growing, if not already central, role being played by information and communication technologies (ICTs) in the economy in particular, and in society as a whole.

As a consequence, governments worldwide have increasingly focused on ICTs for development (ICTD). For these governments, ICTD is not only about developing the ICT industry or sector of the economy, but also encompasses the use of ICTs to stimulate economic growth, as well as social and political development.

However, among the difficulties that governments face in formulating ICT policy is unfamiliarity with a rapidly changing technology landscape and the competencies needed to harness ICTs for national development. Since one cannot regulate what one does not understand, many policymakers have shied away from ICT policymaking. But leaving ICT policy to technologists is also wrong because often, technologists are unaware of the social and policy implications of the technologies they are developing and using.

The Academy of ICT Essentials for Government Leaders module series has been developed by the Asian and Pacific Training Centre for Information and Communication Technology for Development (APCICT) for:

1. Policymakers at the national and local government level who are responsible for ICT policymaking;
2. Government officials responsible for the development and implementation of ICT-based applications; and
3. Managers in the public sector seeking to employ ICT tools for project management.

The module series aims to develop familiarity with the substantive issues related to ICTD from both a policy and technology perspective. The intention is not to develop a technical ICT manual, but rather to provide a good understanding of what the current digital technology is capable of achieving and where technology is headed, and what this implies for policymaking. The topics covered by the modules have been identified through a training needs analysis and a survey of other training materials worldwide.

The modules are designed in such a way that they can be used for self-study by individuals or as a resource in a training course or programme. The modules are stand-alone as well as linked together, and effort has been made in each module to link to themes and discussions in the other modules in the series. The long-term objective is to make the modules a coherent course that can be certified.

Each module begins with a statement of module objectives and target learning outcomes against which readers can assess their own progress. The module content is divided into sections that include case studies and exercises to help deepen understanding of key concepts. The exercises may be done by individual readers or by groups of training participants. Figures and tables are provided to illustrate specific aspects of the discussion. References and online resources are listed for readers to look up in order to gain additional perspectives.

The use of ICTD is so diverse that sometimes case studies and examples within and across modules may appear contradictory. This is to be expected. This is the excitement, challenge and the promise of this discipline, as countries leverage the potential of ICTs as tools for development.

Supporting the Academy of ICT Essentials for Government Leaders module series in print format is an online distance learning platform—the APCICT Virtual Academy (http://e-learning.unapcict.org) with virtual classrooms featuring the trainers' presentations in video format and PowerPoint presentations of the modules.

APCICT has developed a repository platform for ICTD (http://www.unapcict.org/resources), a dedicated online site for ICTD practitioners and policymakers to enhance their learning and training experience. The repository platform gives access to knowledge resources on different aspects of ICTD and provides a space for sharing knowledge and experiences on advancing ICTD.

REALIZING DATA-DRIVEN GOVERNANCE

Governments continue to struggle with using data to improve governance. Even in the midst of a data revolution, some of them face the traditional challenges of lack of data, low quality data and outdated data. Others face the more contemporary challenge of making sense of mountains of data that are available to them. How can we enhance the use and analysis of data to support governance and government decision-making?

This module, designed for government officials in developing countries, aims to help them understand data-driven governance and assist them in using data-driven governance in their jurisdictions. The module is divided into seven sections.

Section 1 discusses the data revolution, its causes and social effects.

Section 2 examines data-driven governance.

Section 3 describes traditional and contemporary data sources.

Section 4 highlights data governance and data management.

Section 5 delves into data analytics, bias and intuition.

Section 6 explores the enabling environment for data-driven governance.

Section 7 considers the data culture in the public sector.

MODULE OBJECTIVES

THE MODULE AIMS TO:

1. Discuss the social consequences of the data revolution as the context for data-driven governance;
2. Examine data-driven governance through a discussion of data-driven decisions, evidence-based policy, results-based management, and the challenges in monitoring and implementing the United Nations Sustainable Development Goals;
3. Provide a description of traditional and contemporary data sources;
4. Study the role of data governance and data management in ensuring data accessibility, reliability, quality and timeliness for users;
5. Discuss data analytics—the process of examining data in order to draw conclusions, the dangers of algorithmic bias, and the role of intuition in decision-making;
6. Explore policy and other issues related to developing and implementing data-driven governance in developing countries; and
7. Consider the data culture in the public sector.

LEARNING OUTCOMES

After working on this module, users should be able to:

1. Discuss the social consequence of the data revolution;
2. Understand data-driven governance;
3. Appreciate and explain the facilitating role of data governance, data management and data analytics in data-driven governance;
4. Identify and act upon key enabling policies, programmes and activities that promote data-driven governance; and
5. Appreciate the importance of data culture in the public sector.

ACKNOWLEDGEMENTS

The module was prepared by Emmanuel C. Lallana under the overall guidance of Kiyoung Ko, Director of Asian and Pacific Training Centre for Information and Communication Technology for Development Asian and Pacific Training Centre for Information and Communication Technology for Development (APCICT)..

Substantive comments were provided by APCICT's partners who participated in multiple rounds of review. The module benefited from comments during the Expert Group Meeting on Data-Driven Governance held on 6-7 December 2017 in Manila and a Consultative Meeting held on 29-31 August 2018 in Bangkok. Valuable inputs were also received from Tiziana Bonapace, Atsuko Okuda, Siope Vakataki 'Ofa and Matthew Perkins of the ICT and Disaster Risk Reduction Division, ESCAP, and Gemma Van Halderen and Rikke Munk Hansen of the Statistics Division, ESCAP.

Module development was coordinated by Robert de Jesus and Nuankae Wongthawatchai. Editing support was provided by Christine Apikul. Byeongjo Kong, Kevin Drouin, Michael Santiago and Yunjin Lee helped proofread the manuscript. Joo-Eun Chung and Ho-Din Ligay provided administrative support and other necessary assistance for the issuance of this module.

TABLE OF CONTENTS

LIST OF BOXES

LIST OF TABLES

ACRONYMS

APCICT	Asian and Pacific Training Centre for Information and Communication Technology for Development
APEC	Asia-Pacific Economic Cooperation
API	Application Programming Interface
CD	Compact Disc
CEO	Chief Executive Officer
COO	Chief Operating Officer
DAMA-DMBOK2	Data Management Association International's Guide to the Data Management Body of Knowledge, Second Edition
ESCAP	Economic and Social Commission for Asia and the Pacific
EU	European Union
GDPR	General Data Protection Regulation
GPHIN	Global Public Health Intelligence Network
GPS	Global Positioning System
HCFAC	Health Care Fraud and Abuse Control
HHS-OIG	Office of Inspector General for the Department of Health and Human Services (United States of America)
ICT	Information and Communication Technology
ICTD	Information and Communication Technologies for Development
IoT	Internet of Things
MfDR	Monitoring for Development Results
OECD	Organisation for Economic Co-operation and Development
SDG	Sustainable Development Goal
UNDG	United Nations Development Group
UNDP	United Nations Development Programme
UNESCO	United Nations Educational, Scientific and Cultural Organization

1. THE DATA REVOLUTION AND DATA JUSTICE

This section aims to:

- Describe the information and communication technology (ICT) foundation of the data revolution;
- Explain digitization, digitalization, digital transformation and datafication; and
- Discuss data justice.

Data — "the facts and statistics collected together for reference or analysis"[1] — is said to be the new oil. According to the Economist:

> Data is to this century what oil was to the last one: a driver of growth and change. Flows of data have created new infrastructure, new businesses, new monopolies, new politics and – crucially – new economics [...] Many a battle will be fought over who should own, and benefit from, data data.[2]

In developing countries, greater access and use of data can help:

Improve governance by enhancing transparency and accountability, introducing new efficiencies into service delivery, and increasing information sharing within government departments;

Empower citizens by improving their capacity to make decisions and widening their choices, and by acting as a catalyst for social mobilization;

Create economic opportunity by enabling business creation, job creation and new forms of innovation, and more generally, by spurring economic growth; and

Solve complex public problems by improving situational awareness, bringing a wider range of expertise and knowledge to bear on public problems, and allowing policymakers, civil society and citizens to better target interventions and track impact.[3]

1 Oxford Living Dictionaries, "Data". Available at https://en.oxforddictionaries.com/definition/data (accessed on 8 January 2019).

2 The Economist, "Data is giving rise to a new economy", 6 May 2017. Available at https://www.economist.com/briefing/2017/05/06/data-is-giving-rise-to-a-new-economy.

3 Stefaan Verhulst and Andrew Young, "The Evidence that Open Government Data Improves Developing Economies", *ICTworks*, 18 June 2018. Available at https://www.ictworks.org/the-evidence-that-open-government-data-improves-developing-economies/#.W43755Mzau4.

The World Bank suggests that "like oil, unprocessed data has relatively little value and needs to be mined, refined, stored and sold on to create value".[4] Thus, the contemporary challenge for governments is to "extract value from data to improve service delivery in the same way that private companies have learned to do it for profit".[5]

But data is not like oil in crucial ways.[6] Oil is a finite resource but data is "effectively infinitely durable and reusable".[7] Extracting oil is expensive and difficult, while producing data is increasingly cheaper and easier. Oil is costly to move, while data "can be replicated indefinitely and moved around the world at the speed of light, at very low cost".[8]

1.1 Data and Technology

There is no doubt that we are amidst a data revolution: "An explosion in the volume of data, the speed with which data is produced, the number of producers of data, the dissemination of data, and the range of things on which there is data."[9]

In 2002, researchers from the University of California, Berkeley estimated that the world produced "5 exabytes of new information".[10] This is equivalent to 37,000 new libraries with more than 162 million books, manuscripts, maps, photographs, etc. (roughly the size of the world's largest library "— the United States Library of Congress). Five exabytes is also double the amount of information produced in 1999.

In 2013, 4.4 zettabytes of data—"as many bits as stars in the physical universe"—were created worldwide.[11] In the same year, it was estimated that by 2020 the amount of data would grow to 44 zettabytes. It was also projected that in 2017, developing economies would overtake developed economies in creating data.

In 2017, it was projected that "in 2025 the world will create and replicate 163 zettabytes of data, representing a ten-fold rise from the amount of data created in 2016".[12]

---

4 World Bank, *Information and Communications for Development 2018: Data-Driven Development* (Washington D.C., 2019), p. 1. Available at https://www.worldbank.org/en/topic/digitaldevelopment/publication/data-driven-development.print.

5 Ibid.

6 Amol Rajan, "Data is not the new oil", *BBC*, 9 October 2017. Available at https://www.bbc.com/news/entertainment-arts-41559076.

7 Bernard Marr "Here's Why Data Is Not The New Oil", *Forbes*, 5 March 2018. Available at https://www.forbes.com/sites/bernardmarr/2018/03/05/heres-why-data-is-not-the-new-oil/#59bc073f3aa9.

8 Ibid.

9 United Nations Secretary-General's Independent Expert Advisory Group on the Data Revolution for Sustainable Development, "A World that Counts: Mobilizing the Data Revolution for Sustainable Development", November 2014, p. 6. Available at http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf.

10 Regents of the University of California, "How Much Information? 2003: Executive Summary", 27 October 2003. Available at http://groups.ischool.berkeley.edu/archive/how-much-info-2003/execsum.htm.

11 EMC Digital Universe, "The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things", Info Brief, April 2014. Available at https://www.emc.com/collateral/analyst-reports/idc-digital-universe-2014.pdf.

12 David Reinsel, John Gantz and John Rydning, "Data Age 2025: The Evolution of Data to Life-Critical", IDC White Paper, April 2017.

Box 1. A Zettabyte in Numbers

1 kilobyte	1,000
1 megabyte	1,000,000
1 gigabyte	1,000,000,000
1 terabyte	1,000,000,000,000
1 petabyte	1,000,000,000.000,000
1 exabyte	1,000,000,000,000,000,000
1 zettabyte	1,000,000,000,000,000,000,000

Extracted from: Carascojames, "Era Zeta Bytes: How many bytes in Zeta bytes, how will it affect us?" steemKR, https://steemkr.com/technology/@carascojames/era-zeta-bytes-how-many-bytes-in-zeta-bytes-how-will-it-affect-us (accessed on 8 January 2019).

Observers partly attribute the data revolution to technology—increasing computing power, faster broadband connections, inexpensive sensors and pervasive mobile phones.[13] Developments in storage (such as cloud computing) and database systems have also enabled the collection, storage and processing of very large amounts of data.

The widespread use of technology has allowed us to evolve from information consumers to information producers.[14] We used to watch television, listen to the radio and read newspapers. In 2016, we tweeted 456,000 times, posted 46,740 Instagram photos, Googled 3.6 million searches, sent 103,447,520 spam emails and published 600 new page edits on Wikipedia **every minute**.[15]

As information producers, we create **digital footprints**—the sum of all data that we produce as a result of our online activities.[16] This includes data that we create when we use social media (such as Facebook and Twitter) and when we Google.

At the same time, we produce **data exhaust**—the by-product of our online activities.[17] For example, when we make a call using our mobile phones, a call log—that consists of our phone number and that of the party we are calling, call time and duration, and information about our device's interactions with cellular towers—is created.[18] We also produce data exhaust when we use our web browser to search or buy online. Our data exhaust provides significant information about our online behaviour, and when processed can lead to valuable insights about our offline habits and preferences.

---

13 United Nations Global Pulse, "Big Data for Development: Challenges and Opportunities", May 2012, p. 9. Available at http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseMay2012.pdf.

14 Kirsty, "Where Did the 'Data Explosion' Come From?" *Bime Blog*. Available at https://blog.bimeanalytics.com/english/where-did-the-data-explosion-come-from (accessed on 8 January 2019).

15 Tom Hale, "How Much Data Does the World Generate Every Minute?" *IFL Science*, 26 July 2017. Available at http://www.iflscience.com/technology/how-much-data-does-the-world-generate-every-minute/.

16 Margaret Rouse, "Data Exhaust", *Whatls.com*, April 2015. Available at https://whatis.techtarget.com/definition/data-exhaust.

17 Ibid.

18 Jessica Leber, "Mobile Call Logs Can Reveal a Lot to the NSA", *MIT Technology Review*, 18 June 2013. Available at https://www.technologyreview.com/s/516181/mobile-call-logs-can-reveal-a-lot-to-the-nsa/.

While we have become information producers, we are no longer the sole producers of data. Our devices that are connected to the Internet are also creating data.[19] This development is called the Internet of Things (IoT): "A system of interrelated computing devices, mechanical and digital machines, objects, animals or people that are provided with unique identifiers, and the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction."[20] In short, "Internet-enabled communications between everyday objects".[21]

In 2017, over 20 billion devices were estimated to be connected to the Internet.[22] This number is expected to grow to almost 31 billion in 2020 and then to about 75 billion devices in 2025. As a result, by 2025, the "IoT will generate over 2 zettabytes of data, mostly generated by consumer electronics devices".[23]

Humans will be hyper connected in the face of the IoT. It is predicted that by 2025, "an average connected person anywhere in the world will interact with connected devices nearly 4,800 times per day or **one interaction every 18 seconds**".[24]

IoT will not only provide important insights about individual behaviour, it will also help solve society's challenges.

---

19 Steve Ranger, "What is the IoT? Everything you need to know about the Internet of Things right now", *ZDNet*, 21 August 2018. Available at http://www.zdnet.com/article/what-is-the-internet-of-things-everything-you-need-to-know-about-the-iot-right-now/.

20 Margaret Rouse, "Internet of Things (IoT)", *TechTarget IoT Agenda*, June 2018. Available at https://internetofthingsagenda.techtarget.com/definition/Internet-of-Things-IoT.

21 Hyea Won Lee, "Agriculture 2.0: how the Internet of Things can revolutionize the farming sector", *World Bank Information and Communications for Development Blog*, 17 August 2017. Available at http://blogs.worldbank.org/ic4d/agriculture-20-how-internet-things-can-revolutionize-farming-sector.

22 Statista, "Internet of Things (IoT) connected devices installed base worldwide from 2015 to 2025 (in billions)". Available at https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/ (accessed on 8 January 2019).

23 Ian Scales, "IoT 2025: 27 billion devices spewing 2 zettabytes of data and generating $3 trillion", *TelecomTV*, 5 August 2016. Available at http://www.telecomtv.com/articles/iot/iot-2025-27-billion-devices-spewing-2-zettabytes-of-data-and-generating-3-trillion-13872/.

24 Andrew Cave, "What Will We Do When the World's Data Hits 163 Zettabytes in 2025?" *Forbes*, 13 April 2017. Available at https://www.forbes.com/sites/andrewcave/2017/04/13/what-will-we-do-when-the-worlds-data-hits-163-zettabytes-in-2025/#550efad8349a.

Box 2. IoT and Water Security

IoT technology can provide greater comprehension of the complex challenges surrounding water security, enabling governments to better define priorities for water supply, consumer demand and governance. Additionally, IoT applications can help agencies better coordinate response among stakeholders by capturing the specific impacts of each policy. This can be achieved using predictive models or through real-time measurement that enables A/B testing (to compare two versions, A and B).

Increasing water supply is often the first option considered as water inventories drop, and traditionally, companies have invested heavily in finding new sources of water. As new sources dry up, however, utilities may instead focus on improving the yield for delivery—since the infrastructure is often old, and water-supply systems lose 16 per cent on average during delivery. One of the challenges the IoT can solve is determining exactly where to repair to improve yield, and whether the volume saved for that area will offset the capital cost of repair. Sensors can provide a more precise understanding of water flows and help prioritize improvements, even at the level of individual homeowners not typically engaged with the state of water infrastructure. Stopping or slowing in-home leaks, which can waste up to 10,000 gallons a year, can further boost the yield on sanitized water. Products such as LeakSmart, for example, combine a simple sensor and actuator to detect when a pipe has burst and shut off the water.

Extracted (with modifications) from: Max Meyers, Claire Niech and William D. Eggers, "Anticipate, sense, and respond: Connected government and the Internet of Things", Deloitte Insights, 28 August 2015. Available at https://www2.deloitte.com/insights/us/en/focus/internet-of-things/iot-in-government.html.

1.2 From Digitization to Datafication

The data revolution can also be understood from the lens of "digitization", "digitalization", "digital transformation" and "datafication".

Digitization is the process of converting analog into digital.[25] Our telephones used to be analog, they are now digital. Where before our data were stored in paper and in cabinet drawers, we now have them in binary code (ones and zeros) and in databases.[26]

25 Margaret Rouse , "Digitization", *WhatIs.com*, April 2007. Available at http://whatis.techtarget.com/definition/digitization.

26 Gil Press, "A Very Short History of Digitization", *Forbes*, 27 December 2015. Available at https://www.forbes.com/sites/gilpress/2015/12/27/a-very-short-history-of-digitization/#bf522f349ac2.

Box 3. Ways We Have Digitized Our World

Snail mail to email: Mail has transitioned to a digital format. Most companies offer some form of online billing so customers never have to see a paper bill, and we can access catalogues online rather than get them in our mailboxes. From business correspondence to letters among friends, most of our daily mail is now found on our computers, tablets or smartphones.

CDs to MP3s: Compact discs (CDs) have been replaced by MP3s on a large scale. Most people download their music to play on their phones or MP3 players, rather than stack up a massive collection of CDs. Turning music into media files rather than actual records is the norm.

Road maps to GPS: When was the last time you bought a map? It's now easy to get our hands on a global positioning system (GPS) device like Garmin or TomTom, or simply use navigation features on smartphones. Or we can hop online to get directions from Google. All our maps can now be accessed through computers and smartphones with rapidly updated information, rather than paper maps that are quickly outdated.

Extracted (with modifications) from: Jaymi Heimbuch, "7 major ways we're digitizing our world, and 3 reasons we still want hardcopies", Treehugger, 11 October 2010. Available at https://www.treehugger.com/clean-technology/7-major-ways-were-digitizing-our-world-and-3-reasons-we-still-want-hardcopies.html.

Digitalization means "turning interactions, communications, business functions and business models into (more) digital ones".[27]

In business, it refers to improving business operations using digital technologies.

Box 4. Taking Meeting Minutes

Imagine a business that holds morning C-Suite meetings. Since the meetings are important, everyone agrees that the assistant of the Chief Executive Officer (CEO) needs to take the minutes. However, the attendees are split on how they should take and share the minutes.

The CEO suggests that the assistant handwrites the minutes and then photocopies them. The CEO claims this will be quick and easy.

The Chief Operating Officer (COO) suggests that the assistant types the minutes and then emails the attendees a PDF copy. It is harder to lose an email than it is a photocopy, says the COO.

The Chief Technology Officer (CTO) suggests that they create a morning C-Suite meeting page and template on a software-as-a-service application that services this need. This way everyone knows where previous minutes are saved, can easily access and share these minutes, can add items to meeting pages, and link to meetings.

Extracted (with modifications) from: David Burkett, "Digitisation And Digitalisation: What Means What?" WorkingMouse, 19 December 2017. Available at https://workingmouse.com.au/innovation/digitisation-digitalisation-digital-transformation.

27 I-Scoop, "Digitization, digitalization and digital transformation: the differences". Available at https://www.i-scoop.eu/digitization-digitalization-digital-transformation-disruption/ (accessed on 8 January 2019).

For governments, digitalization can help increase efficiency and transparency, and improve social services.

A 2018 survey of American federal officials showed that "82 per cent of public sector agencies are planning digitalization projects".[28] Inter-agency collaboration (34 per cent), benefits tracking (31 per cent) and benefits applications (29 per cent) are among the key functions that will be digitalized.

Box 5. The Benefits of Digitalization for Governments

Efficiency and cost saving:

Digitalization typically results in better efficiency. Various institutions have different drivers for digitalization, such as accessibility, cost-cutting, tracking, or even simply for the sake of increasing digital adoption itself. All these will—to a certain extent—lead to efficiency. In the most basic sense, a government agency that wants to go paperless (by transferring all its printed materials into the cloud) will be able to save and share space with other government agencies, as well as reduce its carbon footprint and traffic (by reducing the need to deliver and dispose of materials). In addition, government officers can save time when looking for specific documents due to digital indexing, which can further enhance productivity. Estonia, for example, claims to have saved 800 years of working time per year as a result of its digital campaign.

Improved social services:

Digitalization can potentially improve the citizenry's quality of life. An example would be renewing one's driver's license, which could involve travelling to the national transportation agency, filling out forms and waiting in long queues. The entire process can be a matter of minutes only rather than hours if the government transportation agency embrace a digital approach. Many processes can be accomplished online.

Promote transparency:

Given how most governments are promoting honesty and transparency programmes, digitalizing transactions can help provide better visibility and clarity. Corruption can occur during cash transactions between citizens and government agencies. Making transactions digital will not only help state auditors monitor cash flow, but will also encourage citizens and government agencies to uphold ethical practices. With digital platforms, every transaction can be effectively tracked and monitored, while at the same time, reducing bureaucracy and corruption.

Extracted (with modifications) from: Christian Lauron and Irsyad Stamboel, "Digitalization of Government: Suits The C-Suite", BusinessWorld, 18 February 2018. Available at http://www.bworldonline.com/digitalization-of-government/.

Some use digitalization to describe a bigger phenomenon—"the integration of digital technology into everyday life".[29]

---

28 Mathew Chase, "The state of data management in the public sector in 2018", *Experian*, 8 February 2018. Available at https://www.edq.com/blog/the-state-of-data-management-in-the-public-sector-in-2018/.

29 Heikki Otsolampi, "Digitalization – the first true revolution in business history". Available at https://www.avaus.fi/en/blog/digitalization-the-first-true-revolution-in-business-history/ (accessed on 8 January 2019).

Box 6. The Consequences of Digitalization

The process of digitalization provides network users with new and innovative ways to increase and share their social, cultural and economic capital. The connection opportunities offer individuals a rich framework of solutions, based on collective intelligence. But digitalization of our daily lives is also becoming increasingly invasive. Scholars and researchers are exploring the dark side of the network to identify the negative externalities of this digitalization process. At the dawn of this new digital age, a vision linked to a digital risk society is being established in which are highlighted new forms of coercive power, risks of social anomie, collective obsessions, dystopian drifts and forms of social alienation due to the excessive dependence on digital tools.

Adapted from: Michele Bonazzi, "For a Critical Theory of the Digitalization of Everyday Life", 14 July 2016. Available at https://isaconf.confex.com/isaconf/forum2016/webprogram/Paper77619.html.

Digital transformation is the use of digital technology in all aspects of an enterprise to fundamentally change how it creates and delivers value. It is also "a cultural change that requires organizations to continually challenge the status quo, experiment and get comfortable with failure".[30]

Box 7. Digital Transformation and Technology

Governments need to go beyond digitizing existing processes and services. They need to harness the power of digital technologies and data to fundamentally re-imagine and transform the business models of government.

While much progress has been made by many governments around the world, the full potential of digital government remains largely untapped. Many transactional and payment services are still not available end-to-end online. Digital services that do exist often are not optimized for mobile devices. The functionality and user experiences of online services designed and run by governments usually leave a lot to be desired compared to the best practices of commercial organizations.

Why? Legacy systems get in the way, legislation and regulations are hard to change, and security and privacy concerns are complex issues. But despite these challenges, governments recognize the potential of social, mobile, data and cloud technologies to drive transformation in the public sector.

Successful digital transformation requires strong leadership at the highest levels; investments in science, technology, engineering and mathematics skills; and cultural and behavioural change. Likewise, governments need to maximize their digital investments by strategically leveraging new technologies, as well as data and advanced analytics, to optimize policies, programmes, payments and systems.

Extracted (with modifications) from: Boston Consulting Group, "Digital Transformation and Technology: How Governments are Upping their Game in Digital". Available at https://www.bcg.com/industries/public-sector/digital-transformation-technology.aspx (accessed on 8 January 2019).

30 Enterprisers Project, "What is Digital Transformation?" Available at https://enterprisersproject.com/what-is-digital-transformation (accessed on 8 January 2019).

Jason Bloomberg provides a good way to distinguish these trends: "We digitize information, we digitalize processes and roles that make up the operations of a business, and we digitally transform the business and its strategy."[31] In the public sector, whole-of-government e-government initiatives, such as Gov 2.0, is equivalent to digital transformation.

Datafication turns previously invisible process/activity into data that can be monitored, tracked, analysed and optimized.[32] To "datafy" is to "render into data many aspects of the world that have never been quantified before".[33]

Take emotions—anger, fear, joy, amazement, etc. Until the widespread use of social media, emotions were very hard to track and analyse. Now we create a digital footprint of our emotions every time we like, laugh, love or get mad at a Facebook post. As noted by Mayer-Schonberger and Cukier: "Social networking platforms don't simply offer us a way to find and stay in touch with friends and colleagues, they take intangible elements of our everyday life and transform them into data that can be used to do new things."[34]

Box 8. The Datafication of Everyday Life

Timo Elliott wrote on his blog site:

My exercise is now datafied. I went for a run this morning, and my Fitbit One device recorded exactly how long I ran for, how many strides I took, and how many calories I burned in the process. For the first time, it's very easy for me to track and monitor my exercise progress.

And that's just one small example. A lot of my daily activity is now automatically tracked. My network of friends is now datafied with Facebook. My network of professional connections is datafied with LinkedIn. My location is datfied with Foursquare. My latest random thoughts are datafied on Twitter. My music preferences are datafied with Spotify.

Even reading books is now datafied. While I'm reading on my Kindle device, it's actually watching me. Amazon tracks my reading data and uses it to provide useful services. For example, it knows what page I'm on, so I can easily switch between different devices. It uses my reading speed to estimate how long it's going to take me to finish a book. And Amazon has incorporated some aspects of the wisdom of the crowd idea—for example, I can choose to see which passages other people have highlighted as the most interesting.

---

31 Jason Bloomberg "Digitization, Digitalization, and Digital Transformation: Confuse Them At Your Peril", *Forbes*, 29 April 2018. Available at https://www.forbes.com/sites/jasonbloomberg/2018/04/29/digitization-digitalization-and-digital-transformation-confuse-them-at-your-peril/#5c2560f82f2c.

32 Margarita Shilova, "The Concept Of Datafication; Definition & Examples", *Apiumhub*, 15 June 2017. Available at https://apiumhub.com/tech-blog-barcelona/datafication-examples/.

33 Kenneth Cukier and Viktor Mayer-Schoenberger, "The Rise of Big Data: How It's Changing the Way We Think About the World", *Foreign Affairs*, vol. 92, no. 3 (May/June 2013), p. 29. Available at https://www.foreignaffairs.com/system/files/pdf/articles/2013/92305.pdf.

34 Viktor Mayer-Schonberger and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work and Think* (London, John Murray, 2013) p. 91.

That data is also being collected and analysed by Amazon to optimize book sales. For example, when I recently finished a book in a series by Ken Follet, I received an email the very next morning, giving me a special offer on the next book in the series (interestingly, at a price that was higher than the current "normal" price…)

Extracted (with modifications) from: Timo Elliott, "The Datification Of Our Daily Lives", Digital Business & Business Analytics, 9 July 2013. Available at https://timoelliott.com/blog/2013/07/the-datification-of-our-daily-lives.html.

Datafication means that all elements can now be analysed for patterns and correlations. In the case of a digitized book, not only words or phrases but sentences and paragraphs can be analysed.

Datafication is already changing business.[35] Human resources departments use data from mobile phones, apps or social media usage to identify potential employees' specific characteristics such as risk-taking profile and personality. Datafication will make personality tests or analytical thinking tests obsolete. It will also enable more personalized products and services.

Additionally, datafication is streamlining and improving business processes (the set of tasks and activities to deliver a service or product to a client). Table 1 gives some examples.

Table 1. The Impact of Datafication on Industries

Industry	Impact of Datafication
Micro and "short" supply chains	Eases the creation of short supply chains, creating micro supply chain business processes encapsulated via low-cost technologies such as mobile devices.
Agriculture and food	Increases the traceability and removal of intermediaries in the industrial supply chain.
Manufacturing	Feedback from products "in use" provides improved product development practices. Shared data across multiple manufacturers allows increased efficiencies across supply chains.
Commercial real estate management	Redefines how different parts of a city are classified for industrial use. Increases the level of detail for real estate customers in understanding where to locate their businesses for best impact and foot flow.

Source: Ericsson, The Impact Of Datafication On Strategic Landscapes (Stockholm, 2014). Available at https://www.ericsson.com/assets/local/news/2014/4/the-impact-of-datafication-on-strategic-landscapes.pdf.

35 Rahul Zingre, "The Increasing Datafication Of Our Lives", *LinkedIn*, 22 February 2018. Available at https://www.linkedin.com/pulse/increasing-datafication-our-lives-rahul-zingre.

Datafication is also driving the use of algorithms for decision-making.[36] Already some are advocating that "many decisions, judgements and forecasts now made by humans should be turned over to algorithms".[37]

While there will be those who welcome this development, others point out that datafication and algorithmic decision-making have both negative and unintended consequences.[38]

1.3 Data Justice

The data revolution has profound social repercussions.

For some, it has opened up new economic opportunities and greater freedom. For others, it has meant more intensive and unwanted surveillance.

We now have a new term—dataveillance—surveillance through data footprints. It was coined in the mid-1980s to "draw attention to the substantial shift … from (expensive) physical and electronic surveillance of individuals to (cheap) surveillance of people's behaviour through the increasingly intensive data trails that their behaviour was generating".[39] Dataveillance also "specifically indicates the ability of reorienting, or nudging, individuals' future behaviour by means of four classes of actions: (1) recorded observation; (2) identification and tracking; (3) analytical intervention; and (4) behavioural manipulation".[40]

Data can support existing power asymmetries or enable greater distributive justice through making the poor visible.

G. Thomas Kingsley of the Urban Institute argues that "we need to use data in ways that will make change happen, not just track it".[41] Getting the right data to decision makers will determine "whether the world's urban future will be a story of inclusion and prosperity or a tragedy (over a billion people living in abject poverty in urban slums with scant water supply, sanitation, or other services—and highly at risk of environmental disaster)".[42]

---

36 Marijn Janssen, Yannis Charalabidis and Helmut Krcmar, "Open Data, Information Processing and Datafication of Government", *Proceedings of the 50th Hawaii International Conference on System Sciences* (2017), p. 2670. Available at https://scholarspace.manoa.hawaii.edu/bitstream/10125/41478/1/paper0329.pdf.

37 Andrew McAffee and Erik Brynjolfsson, *Machine, Platform, Crowd: Harnessing Our Digital Future* (New York and London, W. W. Norton & Company, 2017), p. 64.

38 Olivera Marjanovic and Dubravka Cecez-Kecmanovcic, "Understanding Datafication Effects of Open Government Information Systems – A Contemporary Systems Thinking Approach", *Proceedings of the 50th Hawaii International Conference on System Sciences* (2017). Available at https://pdfs.semanticscholar.org/2d72/1ffcba1a30d3e259a06cbd2905f0dbb41419.pdf.

39 Roger Clarke, "Dataveillance and Information Privacy". Available at http://www.rogerclarke.com/DV/#SurvD (accessed on 8 January 2019).

40 Sara Degli-Esposti, "When big data meets dataveillance: The hidden side of analytics", *Surveillance & Society*, vol. 12, no. 2 (May 2014), p. 210. Available at https://www.researchgate.net/publication/262493771_When_big_data_meets_dataveillance_The_hidden_side_of_analytics.

41 G. Thomas Kingsley, "Global development demands a data revolution that will make change happen, not just track it", *Urban Institute*, 4 May 2017. Available at https://www.urban.org/urban-wire/global-development-demands-data-revolution-will-make-change-happen-not-just-track-it.

42 Ibid.

Linnet Taylor defines data justice as: "Fairness in the way people are made visible, represented and treated as a result of their production of digital data."[43] Its central concern is "how to balance and integrate the need to be seen and represented appropriately with the needs for autonomy and integrity".[44]

Data justice has three pillars: (1) visibility; (2) engagement with technology; and (3) non-discrimination.[45]

The data justice pillar of visibility deals both with privacy and representation of those in the margins. How much of the population are invisible? And when they are made visible, are they seen as subjects/clients to be controlled or as citizens with sovereign power?

Globally, more than 1.1 billion people worldwide are "invisible", and more than one-third of them are children.[46] In the developing world, lack of data about children can lead to two things. First, it can perpetuate patterns of poverty and inequality due to poor information used in decision-making.[47] Second, they will be "more difficult to trace and so easy victims of trafficking, commercial sexual exploitation, illegal adoption, infanticide and other forms of abuse".[48]

Those in the margins are also invisible. National-level population censuses/surveys omit questions relating to sexual orientation and gender identity or expression.[49] The consequence of this is "a dearth of understanding about the experiences of the lesbian, gay, bi-sexual, transgender and queer community, and how gender identity and sexuality influence other demographics such as race, ability or geopolitics".[50]

The second data justice pillar is engagement with technology. This pillar deals with the freedom not to use specific digital technologies, and more importantly, how not to become part of commercial databases (as a by-product of development interventions). It also includes the freedom to control the terms of one's engagement with data markets.

Of the 3.2 billion people with Internet, 2 billion are from developing countries. Having overcome the access challenge, they now face a different one—surveillance through data collection.

---

43 Linnet Taylor, "What is data justice? The case for connecting digital rights and freedoms globally", 16 February 2017, p. 1. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2918779.

44 Ibid., p.18.

45 Ibid., p. 15.

46 Agence France-Presse, "More than 1 billion 'invisible people' worldwide have no proof of identity", 22 October 2017. Available at https://www.pri.org/stories/2017-10-22/more-1-billion-invisible-people-worldwide-have-no-proof-identity.

47 Caroline Ford, "Making Invisible Populations Count – The development agenda of the 21st century", *Consortium for Street Children*, 12 October 2018. Available at https://www.streetchildren.org/news-and-updates/making-invisible-populations-count-the-development-agenda-of-the-21st-century/.

48 Lucia Hanmer and Marina Elefante, *The Role of Identification in Ending Child Marriage: Identification for Development (ID4D)* (Washington D.C., World Bank, 2016), p. 7. Available at http://documents.worldbank.org/curated/en/130281472492551732/pdf/107932-WP-P156810-OUO-9-Child-Marriage.pdf.

49 Charlie Whittington, "Invisible In Data: The Lack of LGBT Data Collection", *Georgetown Public Policy Review*, 17 July 2018. Available at http://gppreview.com/2018/07/17/invisible-data-lack-lgbtq-data-collection/.

50 Ibid.

The collection of personal data has become the norm in our Internet-mediated world. Businesses routinely collect data from their customers. Aside from continuing to collect administrative data, governments now monitor social media accounts of their citizens.[51] Data brokers "obtain personal data from public and private parties to combine for resale to businesses".[52]

Like the digital divide, socio-economic position, gender, ethnicity and place of origin determine who gets surveilled. These factors determine which databases an individual is part of, how those systems use one's data and the kinds of influence they can have over individuals.[53] For example:

> A teenager from an immigrant family, living in a low-income area, whose parents are poor and who belongs to a minority ethnic group and religion is exponentially more likely to be targeted for surveillance by both protective (social services) and preventive (law enforcement) authorities, and is also likely to have less opportunity to resist that surveillance or intervention than her friend who lives in a high-income area and belongs to the majority ethnic group.[54]

Furthermore, recent trends suggest decreasing ability to control the terms of one's engagement with data markets. The European Data Privacy Supervisor notes that "ever-increasing amounts of personal information are being collected and processed in increasingly opaque and complex ways".[55] Even worse, "the next generation of personal data is likely to be even less accessible to the individuals to whom it relates".[56] The European Data Privacy Supervisor recognizes these developments have important consequence on human dignity, individual freedom and the functioning of society.[57]

Non-discrimination, the third pillar of data justice, is composed of the power to identify and challenge bias in data use, as well as, the freedom from prejudicial treatment.

The case of the indigenous Polynesian people of New Zealand is illustrative. The New Zealand Data Futures Forum reported that:

> Many Maori do not perceive themselves as having benefited much from the collection and use of data. They perceive a real and immediate risk of greater data availability being used for ethnic profiling to their detriment. Despite widespread demands on them for data in the past, the data seems to be rarely used in ways that might benefit them… Collection, storage and use of data often occur in ways that do not respect Maori *tikanga*.[58]

---

51 Heidi Swart, "Government surveillance of social media is rife. Guess who's selling your data?" *Daily Maverick*, 25 April 2018. Available at https://www.dailymaverick.co.za/article/2018-04-25-government-surveillance-of-social-media-is-rife-guess-whos-selling-your-data/.

52 World Bank, *Information and Communications for Development 2018: Data-Driven Development* (Washington D.C., 2019), p. 54. Available at https://www.worldbank.org/en/topic/digitaldevelopment/publication/data-driven-development.print.

53 Linnet Taylor, "What is data justice? The case for connecting digital rights and freedoms globally", 16 February 2017, p. 3. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2918779.

54 Ibid.

55 European Data Privacy Supervisor, "Towards a new digital ethics: Data, dignity and technology", Opinion 4, 2015, p. 6. Available at https://edps.europa.eu/sites/edp/files/publication/15-09-11_data_ethics_en.pdf.

56 Ibid., p. 9.

57 Ibid.

58 Elizabeth Stuart, Emma Samman, William Avis and Tom Berliner, "The data revolution: Finding the missing millions", Overseas Development Institute Research Report 3, April 2015, p. 43. Available at https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/9604.pdf.

The data justice pillar of non-discrimination implies that:

> Methods have to be devised that can allow for the governance of algorithmic processes and decision-making, and that the responsibility for challenging discrimination on the part of individuals will need to be accompanied by the ability to identify and create penalties for it on the part of government.[59]

A related approach to data justice is structural data justice. This approach focuses on: "The degree to which society contains and supports the data-related institutions, relations and knowledge systems necessary for realization of the values comprised in a good life."[60]

The structural data justice approach argues that it is not enough to legislate open access to data policies. Access to data goes beyond laws, economic costs or physical distance. To truly enable access to data, we need to examine how social structures (structural position, resource, institutional and epistemic control) enable or constrain access. For instance, digitalization of public services in developing countries will primarily benefit those who have Internet access, who are likely to be affluent males living in urban areas.

The *Data Justice for Development Manifesto* is a first step in promoting data justice in a datafied world. The manifesto calls for the following:

1. Demand just and legal uses of development data;
2. Demand data consent of citizens that is truly informed;
3. Build upstream and downstream data-related capabilities among those who lack them in developing countries;
4. Promote rights of data access, data privacy, data ownership and data representation;
5. Promote data system outcomes that address international development goals and priorities, including the goals and priorities of data subjects;
6. Support "small data" uses by individuals and communities in developing countries;
7. Advocate sustainable use of data and data systems;
8. Create a social movement for the "data subalterns" of the global South;
9. Stimulate an alternative discourse around data-intensive development that places issues of justice at its heart;
10. Develop new organizational forms such as data-intensive development cooperatives;
11. Lobby for new data justice-based laws and policies in developing countries (including action on data monopolies); and
12. Open up, challenge and provide alternatives to the data-related technical structures (code, algorithms, standards, etc.) that increasingly control international development[61].

Data justice is necessary for an inclusive data-driven development and governance.

---

59 Ibid., p. 9.

60 Richard Heeks, "A Structural Model and Manifesto for Data Justice for International Development", Development Informatics Working Paper No. 69, University of Manchester, 2017. Available at http://hummedia.manchester.ac.uk/institutes/gdi/publications/workingpapers/di/di_wp69.pdf.

61 Ibid.

2. DATA AND GOVERNANCE

This section aims to:

- Define governance;
- Define data-driven governance and data-driven government;
- Discuss data-driven decision-making, evidence-based policymaking and results-based management as forms of data-driven government; and
- Take up the Sustainable Development Goals as an example of data-driven governance at the global level.

2.1 Governance

Governance is how societies organize to define and achieve their common future.

This definition is consistent with the Peters and Pierre definition of governance as "the definition and pursuit of collective interest", and Prakash and Hart's view of governance as "organizing collective action".[62]

Governance is: "A continuing process through which conflicting or diverse interests may be accommodated and cooperative action may be taken" and includes "formal institutions and regimes empowered to enforce compliance, as well as informal arrangements that people and institutions either have agreed to or perceive to be in their interest".[63] Governance "is not static but represents a continuing pattern of adaption to opportunities and circumstances".[64]

Governance involves both public and private actors. "New governance theory", according to John Ruggie, "rests on the premise that the state by itself cannot do all the heavy lifting required to meet most pressing societal challenges and that it therefore needs to engage other actors to leverage its capacities".[65]

Governance issues can be classified under "demand politics" and "supply politics".[66]

62 B. Guy Peters and Jon Pierre, *Comparative Governance: Rediscovering the Functional Dimension of Governing* (Cambridge, Cambridge University Press, 2016), p. 6; and Aseem Prakash and Jeffrey Hart, "Globalization and Governance: an Introduction" in *Globalization and Governance*, Aseem Prakash and Jeffrey Hart, eds. (London and New York, Routledge, 1999), p. 2.

63 Commission on Global Governance, "Our Global Neighborhood: Chapter One – A New World". Available at https://www.gdrc.org/u-gov/global-neighbourhood/chap1.htm (accessed on 8 January 2019).

64 B. Guy Peters and Jon Pierre, *Comparative Governance: Rediscovering the Functional Dimension of Governing* (Cambridge, Cambridge University Press, 2016), p. 16.

65 John Gerard Ruggie, "Global Governance and 'New Governance Theory': Lessons from Business and Human Rights", *Global Governance*, vol. 20 (2014), pp. 8-9. Available at https://www.hbs.edu/faculty/conferences/2014-business-beyond-the-private-sphere/Documents/Global%20Governance%20and%20%27New%20Governance%20Theory%27.pdf.

66 B. Guy Peters and Jon Pierre, *Comparative Governance: Rediscovering the Functional Dimension of Governing* (Cambridge, Cambridge University Press, 2016), p. 8.

Demand politics is expressions of societal expectations usually directed to government. It includes manifestations of desire for collective action. Supply politics is about the "the capacity of government to address (citizen) expectations and to solve societal problems".[67] It focuses on the ability of civil society (e.g., business sector, non-governmental organizations and other groups) to provide solutions and arrangements to society's common problems.

Governance occurs at different levels—local, national and global.[68]

Local governance is at the community level, within a small area or at the lowest tier of a country's political subdivisions. For the United Nations Development Programme (UNDP), local governance "is the channel closest to the citizens for accessing basic services, for participating in the public decisions that affect their lives, and for exercising their rights and obligations".[69]

National governance occurs in the whole area of a nation-state. It includes "the manner in which power is exercised in the management of a country's economic and social resources for development".[70]

Global governance is: "The way in which actors—individuals, institutions (again both public and private)—attempt to accommodate conflicting interests through processes of collective action decision-making in a range of areas operating beyond (nation-) state borders."[71] It is also "the international process of consensus-forming which generates guidelines and agreements".[72]

Table 2 below provides a snapshot of governance issues at different levels.

Table 2. Governance Issues at Local, National and Global Levels

	Local	National	Global
Demand Politics	Anti-corruption drives	Elections, lobbying, social movements	Seattle protests against the World Trade Organization, anti-child trafficking
Supply Politics	Social services, managing the commons	National laws and regulations, defense and security	Millennium Development Goals, Sustainable Development Goals

67 Ibid.

68 John Pierre and B. Guy Peters, *Governance, Politics, and the State* (London, McMillan Press Ltd., 2000), pp. 75-93.

69 UNDP, "Local Governance and Local Development". Available at http://www.undp.org/content/undp/en/home/democratic-governance-and-peacebuilding/responsive-and-accountable-institutions/local-governance-and-local-development.html (accessed on 8 January 2019).

70 World Bank, *Governance and Development* (Washington D.C., World Bank, 1992), p. 1. Available at http://documents.worldbank.org/curated/en/604951468739447676/pdf/multi-page.pdf.

71 Richard Higgott, "The Theory and Practice of Global and Regional Governance: Accommodating American Exceptionalism and European Pluralism", University of Warwick GARNET Working Paper No. 01/05, pp. 4-5. Available at https://warwick.ac.uk/fac/soc/pais/research/researchcentres/csgr/garnet/workingpapers/0105.pdf.

72 InternationalRelations.org, "Global Governance Definition". Available at http://internationalrelations.org/global-governance/ (accessed on 8 January 2019).

Governance is closely related to government or "formal structures of the public sector and the set of actors exercising state power".[73] Oran Young defines government as "a material entity specialized to the provision of governance".[74]

It is also well accepted that governance is "not limited to government, since other social institutions may provide governance services as well".[75]

Oran Young talks of five different governance systems:

1. Governance by government;
2. Governance by intergovernmental agreement;
3. Private governance through industry self-regulation and codes of conduct;
4. Civil governance or governance by civil society; and
5. Hybrid mechanisms where different governance systems are combined to solve specific problems[76].

Lemos and Agrawal expands on the hybrid mechanism of governance by identifying its four forms:

1. Co-management (between state agencies and communities);
2. Public-private partnerships (between state agencies and market actors);
3. Private-social partnerships (between market actors and communities); and
4. Multi-partner governance (involving all the different types of agents).[77]

Despite the presence of private actors, government remains a key governance actor. A major difference between government and private actors is that government "is an authority with power to compel, to coerce".[78] It also has the authority to collect taxes. As noted by the Commission on Global Governance, "any adequate system of governance must have the capacity to control and deploy the resources necessary to realize its fundamental objectives".[79]

While there are those who treat government as simply one of many governance actors, their weakness is that they "have underestimated the persistence and in some cases extension of forms of political power concerning security, sovereignty and the use and threat of physical force and violence".[80]

---

73 B. Guy Peters and Jon Pierre, *Comparative Governance: Rediscovering the Functional Dimension of Governing* (Cambridge, Cambridge University Press, 2016), p. 5.

74 Oran Young, "Governance for Sustainable Development in a World of Rising Interdependencies", in *Governance for the Environment*, Magali A. Delmas and Organ Young, eds. (Cambridge, Cambridge University Press, 2009), p. 20.

75 Aseem Prakash and Jeffrey Hart, "Globalization and Governance: an Introduction" in *Globalization and Governance*, Aseem Prakash and Jeffrey Hart, eds. (London and New York, Routledge, 1999), p. 2.

76 Oran Young, "Governance for Sustainable Development in a World of Rising Interdependencies", in *Governance for the Environment*, Magali A. Delmas and Organ Young, eds. (Cambridge, Cambridge University Press, 2009), pp. 24-30.

77 Maria Camen Lemos and Arun Agrawal, "Environmental Governance and Political Science", in *Governance for the Environment*, Magali A. Delmas and Organ Young, eds. (Cambridge, Cambridge University Press, 2009), p. 79.

78 Ann Florini, "Global Governance and What It Means", *Brookings*, 16 February 2009. Available at https://www.brookings.edu/on-the-record/global-governance-and-what-it-means/.

79 Commission on Global Governance, "Our Global Neighborhood: Chapter One – A New World". Available at https://www.gdrc.org/u-gov/global-neighbourhood/chap1.htm (accessed on 8 January 2019).

80 Mitchell Dean, *Governing Societies: Political Perspectives on Domestic and International rule* (Berkshire and New York, Open University Press, McGraw Hill Education, 2007) p. 58.

A second definition of governance is also used in this module: "Governance as steering and coordination."[81] This definition is used in the context of data governance—"a collection of practices and processes which help to ensure the formal management of data assets within an organization".[82]

2.2 Data-Driven Governance

The idea of using data to support governance and government decision-making is not new. The history of statistics is closely tied with the emergence of the modernizing 19th century state.[83] What make the present efforts different is that the data revolution has made the ideal more achievable.

Today, we use "data-driven" to describe an activity that is "compelled by data, rather than by intuition or by personal experience".[84] Data-driven also means making "strategic decisions based on data analysis and interpretation".[85]

Data-driven governance is defined as societies' intensive and extensive use of data to define and achieve their common future. It encompasses the pervasive use of data in both demand politics and supply politics.

Data-driven government is "one where, for all critical decisions, actionable information is available when and where needed".[86] It can lead to:

- Sounder governance and control;
- Optimized fraud and error detection, mitigation and prevention;
- Improved services based on insights gained from those being served;
- Improved efficiency through intelligence networks, which can lead to reduced costs; and
- Improved public perception of the agency.[87]

---

81 B. Guy Peters and Jon Pierre, *Comparative Governance: Rediscovering the Functional Dimension of Governing* (Cambridge, Cambridge University Press, 2016), p. 4.

82 Charles Roe, "What is Data Governance?" *Dataversity*, 18 December 2017. Available at http://www.dataversity.net/what-is-data-governance/.

83 Stuart Woolf, "Statistics and the Modern State", *Comparative Studies in Society and History*, vol. 31, no. 3 (July 1989), pp. 588-604. Available at https://www.jstor.org/stable/178772?read-now=1&refreqid=excelsior%3Ae9117494e9c81b206098b366401bfc44&seq=1#page_scan_tab_contents.

84 Wikipedia, "Data-driven". Available at https://en.wikipedia.org/wiki/Data-driven (accessed on 8 January 2019).

85 AT Internet, "Glossary: Data-Driven". Available at https://www.atinternet.com/en/glossary/data-driven/ (accessed on 8 January 2019).

86 Terence Lutes, "Data-driven government: Challenges and a path forward", IBM Analytics White Paper, 2015, p. 3. Available at https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=GQW03008USEN.

87 Ibid.

Box 9. How Data Can Drive Citizen-Centric Government: The Case of Australia

Governments collect, through their interactions with citizens, a huge amount of data.

Data can enable government to improve public services and policy design, bringing wide-ranging social benefits to citizens and improvements in how parts of government are delivered.

Data can unlock ways of helping Australia proactively plan for and respond to some of the biggest challenges facing the nation in areas such as health, social security and indigenous affairs.

As well as helping to develop more effective policies, data can act as a key performance indicator for measuring policy outcomes. It can also enable jurisdictions and non-governmental organizations to achieve more without additional funding.

Australians are already seeing the benefits of better use of data in service delivery. The introduction of biometric technology at airports has reduced immigration queues and personally-controlled electronic health records have helped to improve disease management plans.

In summary, data can help the Australian government:

- Develop smarter, evidence-based policy;
- Deliver better, citizen-centric services;
- Increase public sector efficiency and reduce overheads; and
- Evaluate programmes and benchmark third-party providers.

Extracted (with modifications) from: Commonwealth of Australia, Department of the Prime Minister and Cabinet, "Public Sector Data Management", July 2015, p. 12. Available at https://www.pmc.gov.au/sites/default/files/publications/public_sector_data_mgt_project.pdf.

Expanded use of data can improve local, national and global governance. Data can help make citizen services more targeted and effective, and lead to better allocation of scarce resources and to insights into the causes and solutions to costly social problems. Data can also facilitate informed citizen participation in governance.

A 2016 survey of British local governments revealed the following 11 use cases for data:

1. Optimizing management of space and infrastructure;
2. Testing "what works";
3. Intelligent case management;
4. Outcomes-based performance management;
5. Early identification of adverse events and future service pressures;
6. Understanding and responding to citizen needs;
7. Informing public service transformation;
8. Streamlining operational council processes;
9. Opening government;

10. Supporting the local economy, businesses and innovation; and
11. Identifying fraud and error[88].

In the social sector, "data will make the work of social change agents more effective and will build the case for support for the best programmes and enterprises".[89] Data can also improve decision-making.

2.2.1 Data-Driven Decision-Making

Data-driven decision-making "involves collecting data, extracting patterns and facts from that data, and utilizing those facts to make inferences that influence decision-making".[90]

When incorporated into daily operations, data-driven decision-making can transform local and national government from simply reacting to citizen demands and concerns to proactively anticipating issues that matters to citizens most.

How then can governments transition to data-driven decision-making?

A good start is by identifying the organization's level of data use.

The Data Maturity Model, created by Nesta—a British innovation charity—provides a useful and convenient way to gauge that state of data-driven decision-making in governments.[91] In this model, data maturity is determined by a five-point rating scale from "nascent" to "datavore", as follows:

- **Nascent** – Rich in data, poor in intelligence. Data is not a key part of decision-making processes.
- **Basic** – Data is used in reports but usually in a cursory way and with little reference to decisions that have to be made.
- **Intermediate** – Data analysis is usually requested for decision-making, but can be inadequate because analysis is not of high quality, not targeted at the decision to be made or the right data is not available.
- **Advanced** – Some decisions are informed by data on both the frontline and at senior levels, but it is not consistent across the organization.
- **Datavore** – Rich in data intelligence and insight. Data is analysed on specifically for the purposes of key decisions that have to be made, consistently across the organization. Data is available in a timely fashion to support decision-making.[92]

The next step is to develop a plan that will move the organization to higher maturity levels.

---

88 Tom Symons, "Wise Council: Insights From the Cutting Edge of Data-Driven Local Government", *NESTA*, November 2016, p. 7. Available at https://media.nesta.org.uk/documents/wise_council.pdf.

89 Jim Fruchterman, "Using Data for Action and for Impact", *Stanford Social Innovation Review* (Summer 2016). Available at https://ssir.org/articles/entry/using_data_for_action_and_for_impact.

90 Joel Schwartz, "Data-Driven Decision Making: A Primer For Beginners", *Northeastern University Blogs*, 14 December 2017. Available at https://www.northeastern.edu/graduate/blog/data-driven-decision-making/.

91 Tom Symons, "Wise Council: Insights From the Cutting Edge of Data-Driven Local Government", *NESTA*, November 2016, p. 105. Available at https://media.nesta.org.uk/documents/wise_council.pdf.

92 Ibid.

Bernard Marr's ten-step process to improve the use of data in decision-making for business can be adapted to local and national governments.[93] The steps are:

1. **Start with strategy** – Instead of starting with what data you could or should access, start by working out what your organization wants to achieve.
2. **Hone in on the "business" area** – You now need to identify which areas are most important to achieving your overall strategy. If you could only work on improving one or two areas, which would you choose?
3. **Identify your unanswered questions** – By working out exactly what you need to know, you can focus on the data that you really need. Your data requirements, cost and stress levels are significantly reduced when you move from "collect everything just in case" to "collect and measure x and y to answer question z".
4. **Find the data to answer your questions** – It is important to understand that no type of data is inherently better or more valuable than any other type. Focus on identifying the ideal data for your organization—the data that can help you answer your most pressing questions and deliver on your strategic objectives.
5. **Identify what data you already have** – Conduct an inventory of the data your organization has or can access. If the data does not already exist, find ways of collecting it either by putting data collection systems in place or by acquiring or accessing external data.
6. **Work out if the costs and effort are justified** – Once you know the costs of collecting the needed data, you can work out if the tangible benefits outweigh those costs. You should treat data like any other key investment—you need to make a clear case for the investment that outlines the long-term value of data to your strategy.
7. **Collect the data** – Much of this step comes down to setting up the processes and people who will gather and manage your data. You may be buying access to an analysis-ready data set, in which case there is no need to collect data as such. But in reality, many data projects require some amount of data collection.
8. **Analyse the data** – You need to analyse the data in order to extract meaningful and useful business insights.
9. **Present and distribute the insights** – Unless the results are presented to the right people at the right time in a meaningful way, then the size of the data sets or the sophistication of the analytics tools will not really matter. You need to make sure the insights gained from your data are used to inform decision-making and, ultimately, improve performance.
10. **Incorporate the learning into the organization** – Finally, you need to apply the insights from the data to your decision-making, making the decisions that will transform your organization.

Ideally, governments should become "datavores".

93 Bernard Marr, "Data-Driven Decision Making: 10 Simple Steps for Any Business", *Forbes*, 14 June 2016. Available at https://www.forbes.com/sites/bernardmarr/2016/06/14/data-driven-decision-making-10-simple-steps-for-any-business/.

Defining and measuring success is identified as a barrier to more rapid adoption of data-driven decision-making in government.[94] Specifically:

- **Obstacle 1** – It is hard to know what is important. Measures of success can be difficult to define. In government, challenges in defining success and identifying the measures that contribute to it can make it difficult to develop key performance indicators.
- **Obstacle 2** – Departmental objectives may not align with the overall government mission.
- **Obstacle 3** – Self-assessments are ambiguous and subjective. It is unreasonable to expect agency officials to deliver objective self-assessments on their own performance, as the desire to highlight success can introduce ambiguity and subjectivity into the data.
- **Obstacle 4** – Storage formats can impede the use of operational data. Some data highly relevant to the optimization of resource allocation, such as important details on why agency actions succeed or fail, are stored in formats that make them difficult to access. The documents may be scanned copies of handwritten notes or free-text fields in unstructured formats.[95]

These obstacles could be overcome through the following ways:

- **Solution 1** – Identify the most useful measures. The time needed to identify which measures are most informative and most closely connected to mission success is well spent.
- **Solution 2** – Create a "line of sight" from every employee to at least one top-level agency goal. It is very helpful for employees to understand how their work contributes to the overall agency mission. To promote this link, government agencies can disseminate "line of sight" information connecting each employee's role to one or more high-level goals.
- **Solution 3** – Improve the governance and analysis of performance data. Agencies may wish to move responsibility for collecting and reporting on performance measures to an independent entity, potentially under the direction of a chief data officer. Alternatively, increased investment in technological solutions to ensure performance data quality may be considered.
- **Solution 4** – Use cognitive technologies to broaden and deepen performance data. Cognitive technologies are information-processing techniques that can perform specific tasks that, until recently, required human labour. Some of these technologies, such as natural language processing, speech recognition and robotics, are seeing widespread adoption and making rapid progress.[96]

---

94 Mahesh Kelkar and others, "Mission analytics: Data-driven decision making in government", *Deloitte Insights*, 26 September 2016. Available at https://www2.deloitte.com/insights/us/en/industry/public-sector/data-driven-decision-making-in-government.html.

95 Ibid.

96 Ibid.

2.2.2 Evidence-Based Policymaking

Evidence-based policymaking is defined as: "The use of the best available research and information on programme results to guide decisions at all stages of the policy process and in each branch of government."[97]

The four principles of evidence-based policymaking are:

1. Build and compile rigorous evidence about what works, including costs and benefits;
2. Monitor programme delivery and use impact evaluation to measure programme effectiveness;
3. Use rigorous evidence to improve programmes, scale what works, and redirect funds away from consistently ineffective programmes; and
4. Encourage innovation and test new approaches.[98]

Evidence-based policymaking can be used in programme assessment, budget development, implementation oversight, outcome monitoring and targeted evaluation.[99]

Below is a stylized process in developing an evidence-based policy:

1. Determine WHAT constitutes real evidence.
    a. Methodology – Does the analytical approach allow for proper consideration of the problems?
    b. Capacity – Are research skills sufficient to undertake the analysis?
2. Identify WHEN adequate evidence is available to inform decisions.
    a. Time – When to harvest data, gather new data and test the analysis?
    b. Good data – Do high-quality databases support timely analysis?
3. Consider HOW credible evidence is ensured.
    a. Transparency – Is there open debate and discussion to test and educate the public?
    b. Independence – Are there incentives to deliver advice in the public interest (and not corporate/self-interest)?
4. Establish a receptive policy environment.
    a. Is there willingness to test policy options, and are there structures and resources available to do so?[100]

---

97 The Pew Charitable Trusts and MacArthur Foundation, "Evidence-Based Policymaking: A guide for effective government", November 2014, p. 2. Available at http://www.pewtrusts.org/~/media/assets/2014/11/evidencebasedpolicymakingaguideforeffectivegovernment.pdf.

98 Evidence-Based Policymaking Collaborative, "Principles of Evidence-Based Policymaking", September 2016. Available at http://www.evidencecollaborative.org/principles-evidence-based-policymaking.

99 The Pew Charitable Trusts and MacArthur Foundation, "Evidence-Based Policymaking: A guide for effective government", November 2014, p. 5. Available at http://www.pewtrusts.org/~/media/assets/2014/11/evidencebasedpolicymakingaguideforeffectivegovernment.pdf.

100 Gary Banks, "Challenges of evidence-based policy-making", *Australian Public Service Commission*. Available at https://www.apsc.gov.au/challenges-evidence-based-policy-making (accessed on 8 January 2019).

For the proponents of evidence-based policymaking, "greater fidelity to scientific good practice, reduced manipulation or misuse of evidence, and increased application of science will lead to improved social policy outcomes".[101] They also have a preferred methodology in determining best evidence.

On the other hand, critics contend that evidence-based policymaking "obscures the political nature of decisions and, in doing so, 'bias' decisions towards particular outcomes—what can be described as the depoliticization of politics".[102] Evidence-based policymaking critics believe that the policy process is best seen as an arena for competition between contending groups and/or interests.

Justin Parkhurst, acknowledging the important role of evidence and the political nature of policymaking, developed a framework for what he calls the "good governance of evidence". The elements of this framework are:

- **Appropriateness** – Evidence should be selected to address the (multiple) relevant social concerns. Evidence should be considered as to whether it has been created in ways that are useful to achieve policy goals. The applicability of the evidence to the local context should be explicitly considered.
- **Quality** – The pieces of evidence used should be judged on their quality, but quality criteria should reflect the methodological principles pertaining to the form of research utilized (e.g., qualitative interviews versus clinical trials) and the nature of the data generated (e.g., descriptions versus measurements versus estimates).
- **Rigour** – Evidence brought to policy consideration should be rigorously (comprehensively) gathered or synthesized, avoiding selective cherry-picking.
- **Stewardship** – The agent setting the rules and shape of official evidence advisory systems should have a formal mandate.
- **Representation** – The final decision authority for policies informed by evidence should lie with democratically representative and publicly accountable officials.
- **Transparency** – Open information and clear ways for the public to see how the evidence bases informing a decision are identified and utilized should be in place.
- **Deliberation** – The public should be engaged in ways that enable multiple competing values and concerns to be considered in the policy process. This involves giving attention to different points of view, even if not all concerns can be selected in the final policy decisions.
- **Contestability** – Technical evidences and scientific research used in policy decisions should be open to critical questioning and appeals. This involves challenging particular scientific findings, but also enabling challenges over decisions about which evidence to utilize (e.g., to question the appropriateness of the evidence for a specific case).[103]

Can evidence-based policymaking help solve "wicked problems" like climate change, poverty, malnutrition, illiteracy or homelessness?

---

101 Justin Parkhurst, *The Politics of evidence: From evidence-based policy to the good governance of evidence* (Oxford, Routledge, 2017) p. 27.

102 Ibid., pp. 42-43.

103 Ibid., pp. 161-163.

Wicked problems, according to the Australian Public Service Commission, have the following characteristics:

- Difficult to clearly define;
- Have many interdependencies and are often multi-causal;
- Attempts to address wicked problems often lead to unforeseen consequences;
- Often not stable (a moving target);
- Usually have no clear solution;
- Socially complex (not just technically complex);
- Hardly ever sit conveniently within the responsibility of any one organization;
- Involve changing behaviour; and
- Some are characterized by chronic policy failure.[104]

Many believe that governments face policy challenges that are mostly of the wicked variety.

Joshua Newman and Brian Head argue that "a focus on data collection and knowledge translation alone cannot be an effective method for taming wicked problems" because wicked problems have both technical and social features.[105] More data and knowledge may help resolve the technical but not the social feature of wicked problems. Taming wicked problems, according to Brian Head, requires not only better knowledge, but also better consultation and better use of third-party partners.[106]

The Australian Public Service Commission highlights the importance of enhancing the bureaucracy's range of skills and capability to address wicked problems. These skills include working across organizational boundaries, engaging stakeholders and influencing citizens' behaviour, in addition to analytical, conceptual and project management skills.[107]

2.2.3 Results-Based Management

Results-based management is another way in which data can improve governance at the national level.

The United Nations Development Group (UNDG) defines results-based management as: "A management strategy by which all actors, contributing directly or indirectly to achieving a set of results, ensure that their processes, products and services contribute to the achievement of desired results (outputs, outcomes and higher level goals or impact)."[108] In results-based management, actors "use information and evidence on actual results to inform decision-making on the design, resourcing and delivery of programmes and activities, as well as for accountability and reporting".[109]

---

104 Australian Public Service Commission, "Tackling wicked problems: A public policy perspective". Available at https://www.apsc.gov.au/tackling-wicked-problems-public-policy-perspective (accessed on 8 January 2019).

105 Joshua Newman and Brian W. Head, "Wicked tendencies in policy problems: Rethinking the distinction between social and technical problems", *Policy and Society*, vol. 36, no. 3 (2017), pp. 414-429.

106 Brian Head, "Wicked Problems in Public Policy", *Public Policy*, vol. 3, no. 2 (January 2008), p. 114. Available at https://www.researchgate.net/publication/43502862_Wicked_Problems_in_Public_Policy.

107 Australian Public Service Commission, "Tackling wicked problems: A public policy perspective". Available at https://www.apsc.gov.au/tackling-wicked-problems-public-policy-perspective (accessed on 8 January 2019).

108 UNDG, "Results-Based Management Handbook: Harmonizing RBM concepts and approaches for improved development results at country level", October 2011, p. 2. Available at https://undg.org/wp-content/uploads/2016/10/UNDG-RBM-Handbook-2012.pdf.

109 Ibid.

For the United Nations Educational, Scientific and Cultural Organization (UNESCO), the primary goal of the results-based management approach is: "To generate and use performance information for accountability reporting to external stakeholders and for decision-making."[110] Additionally, results-based management enables decision makers and implementers "to take well-informed decisions, be able to learn from their successes or failures, and to share this experience with their colleagues and all other stakeholders".[111]

Despite the difference in focus, both the UNDG and UNESCO (as well as other United Nations agencies) see results-based management as "grounded in objective information and thus places an emphasis on critical inquiry, monitoring and evaluation, and evidence-based decision-making".[112]

The key phases of results-based management include:

- Identifying clear and measurable objectives (results), aided by logical frameworks;
- Selecting indicators that will be used to measure progress towards each objective;
- Setting explicit targets for each indicator, used to judge performance;
- Developing performance monitoring systems to regularly collect data on actual results;
- Reviewing, analysing and reporting actual results vis-à-vis the targets;
- Integrating evaluations to provide complementary performance information not readily available from performance monitoring systems; and
- Using performance information for internal management accountability, learning and decision-making processes, and for external performance reporting to stakeholders and partners.[113]

In the context of this discussion, it is important to remember that a key feature of results-based management is performance measurement—the process of objectively measuring how well an agency is meeting its stated goals or objectives.

Results-based management is one of five programming principles used by the United Nations as "a starting point and guide for the analysis and for all stages of the United Nations Development Assistance Framework".[114] It is also used internally by all United Nations agencies to enhance their respective performance. Donor agencies like USAID (United States of America), DFID (United Kingdom), AustralianAID (Australia), CIDA (Canada), Danida (Denmark), SIDA (Sweden) and the World Bank all promote the use of results-based management.[115] The Asian Development Bank is supporting the institutionalization of results-based management across the Philippine bureaucracy.[116]

---

110 UNESCO, "Results-Based Programming, Management, Monitoring and Reporting (RBM) approach as applied at UNESCO: Guiding Principles", September 2015, p. 8. Available at http://unesdoc.unesco.org/images/0017/001775/177568E.pdf.

111 Ibid.

112 United Nations Joint Inspection Unit, *Results-Based Management in the United Nations System: High-Impact Model For Results-Based Management – Benchmarking framework, stages of development and outcomes* (Geneva, United Nations, 2017). Available at https://www.unjiu.org/sites/www.unjiu.org/files/jiu_note_2017_1_english_0.pdf.

113 OECD, "Results Based Management in the Development Co-Operation Agencies: A Review of Experience (Executive Summary)", February 2000, p. 4. Available at https://www.oecd.org/development/evaluation/dcdndep/31950681.pdf.

114 UNDG, "Guidance Note: Application of the Programming Principles to the UNDAF", January 2010, p. 1. Available at https://undg.org/wp-content/uploads/2016/09/Five-Programming-Principles.pdf.

115 Annette Binnendijk, "Results Based Management in the Development Co-Operation Agencies: A Review of Experience (Executive Summary)", *Development Assistance Committee Working Party on Aid Evaluation*, February 2000. Available at https://www.oecd.org/development/evaluation/dcdndep/31950681.pdf.

116 Asian Development Bank, *Results-Based Management Framework in the Philippines: A Guidebook* (Mandaluyong City, 2013). Available at https://www.adb.org/publications/results-based-management-framework-philippines-guidebook.

Results-based management is closely related to **monitoring for development results** (MfDR)—"a management strategy that focuses on using performance information to improve decision-making".[117] MfDR shifts the focus of management from activities to measurable results, like results-based management.

The MfDR principles are:

- At all phases—from strategic planning through implementation to completion and beyond —focus the dialogue on results for partner countries, development agencies and other stakeholders;
- Align actual programming, monitoring and evaluation activities with the agreed expected results;
- Keep the results reporting system as simple, cost-effective and user-friendly as possible;
- Manage for, not by, results, by arranging resources to achieve outcomes; and
- Use results information for management learning and decision-making, as well as for reporting and accountability.[118]

MfDR's cycle involves five core stages:

1. Setting goals and agreeing on targets and strategies;
2. Allocating the available resources to activities that will contribute to the achievement of the desired results;
3. Monitoring and evaluating whether the resources allocated are making the intended difference;
4. Reporting on performance to the public; and
5. Feeding back information into decision-making.[119]

How is MfDR similar to results-based management?

According to UNDP, MfDR applies the same basic concepts of results-based management—"good planning, monitoring, evaluation, learning and feeding back into planning—but seeks to keep the focus on development assistance demonstrating real and meaningful results". Thus, "MfDR is results-based management in action, but it is oriented more towards the external environment and results that are important to programme countries and less towards an agency's internal performance".[120]

While evidence-based policymaking and results-based management contribute to data-driven governance at the national level, these approaches are equally useful for governance at the local and global levels.

---

117 Managing for Development Results. Available at http://www.mfdr.org/1about.html (accessed on 8 January 2019).

118 OECD and World Bank, "Part 1. MfDR Concepts, Tools and Principles", in *Emerging Good Practice in Managing for Development Results: Sourcebook*, first edition (2006), pp. 10-14. Available at http://www.mfdr.org/Sourcebook/1stEdition/4-MfDRPrinciples.pdf.

119 OECD, "Managing for Development Results: Information Sheet", September 2008. Available at http://www.mfdr.org/About/Final-MfDR-information-sheet.pdf.

120 UNDP, *Handbook on Planning, Monitoring and Evaluating for Development Results* (New York, 2009), p. 6. Available at http://web.undp.org/evaluation/handbook/documents/english/pme-handbook.pdf.

2.3. Challenges in Data-Driven Governance: The Case of the United Nations Sustainable Development Goals

The 2030 Agenda for Sustainable Development, which includes the Sustainable Development Goals (SDGs), is the most recent global governance effort that underscores the importance of data.

This agreement, adopted on September 2015, aims at ending poverty, protecting the planet and ensuring prosperity.[121] The 2030 Agenda for Sustainable Development has 17 goals, 169 targets and 232 indicators to be achieved by 2030. The goals are:

- Goal 1. No Poverty
- Goal 2. Zero Hunger
- Goal 3. Good Health and Well-Being
- Goal 4. Quality Education
- Goal 5. Gender Equality
- Goal 6. Clean Water and Sanitation
- Goal 7. Affordable and Clean Energy
- Goal 8. Decent Work and Economic Growth
- Goal 9. Industry, Innovation and Infrastructure
- Goal 10. Reduced Inequalities
- Goal 11. Sustainable Cities and Communities
- Goal 12. Responsible Consumption and Production
- Goal 13. Climate Action
- Goal 14. Life Below Water
- Goal 15. Life on Land
- Goal 16. Peace, Justice and Strong Institutions
- Goal 17. Partnerships for the Goals.[122]

It is important to note that the SDGs stress the importance of monitoring development results, and the importance of data and statistics for sustainable development.[123] In particular, SDG target no. 17.18 aims to "increase significantly the availability of high-quality, timely and reliable data disaggregated by income, gender, age, race, ethnicity, migratory status, disability, geographic location and other characteristics relevant in national contexts". SDG target no. 17.19 calls for new measures of sustainable development progress beyond gross domestic product.

---

121 UNDP, "World leaders adopt Sustainable Development Goals", 25 September 2015. Available at http://www.undp.org/content/undp/en/home/presscenter/pressreleases/2015/09/24/undp-welcomes-adoption-of-sustainable-development-goals-by-world-leaders.html.

122 UNDP, "Sustainable Development Goals", n.d. Available at https://www.undp.org/content/dam/undp/library/corporate/brochure/SDGs_Booklet_Web_En.pdf.

123 UNDP, "Guidance Note: Data for Implementation and Monitoring of the 2030 Agenda for Sustainable Development", September 2017, p. 3. Available at http://www.undp.org/content/undp/en/home/librarypage/poverty-reduction/guidance-note--data-for-implementation-and-monitoring-of-the-203.html.

United Nations member countries agreed to follow up and review mechanisms for SDG implementation. They agreed to "be rigorous, evidence-based, informed by country-led data which is high-quality, accessible, timely, reliable and disaggregated by income, sex, age, race, ethnicity, migration status, disability and geographic location and other characteristics relevant in national contexts".[124]

Furthermore, in order to ensure effective implementation, the United Nations General Assembly in July 2017 adopted the SDG indicators—"a solid framework of indicators and statistical data to monitor progress, inform policy and ensure accountability of all stakeholders". The SDGs have 232 indicators to measure progress.

However, data collection and the methodology used to monitor the SDG indicators are key challenges.

The SDG indicators are categorized using a three-tier system based on methodology and data issues:

- Tier 1 – An established methodology exists, and data is widely available;
- Tier 2 – An established methodology exists, but data is not readily available; and
- Tier 3 – An internationally-agreed methodology is yet to be developed.[125]

As of 2017, only one-third of the indicators have available data that can be used for monitoring of the SDGs, close to a quarter have a methodology but no data, and 38 per cent do not have an agreed upon methodology and data.[126]

As noted by UNDP: "With at least 232 indicators and counting, an integrated and indivisible framework, and a fundamental principle of 'leaving no one behind', the 2030 Agenda increases by orders of magnitude the scale and scope of data required to implement and monitor sustainable development, challenging even countries with the best statistical capacity."[127]

The good news is that developing countries "have strengthened their ability to collect data and process them".[128] Based on World Bank's Statistical Capacity Indicator, low- and middle-income countries have improved from an average value of 65.3 in 2004 to an average of 68.8 in 2015.[129]

In Asia, the data challenges faced by governments include:

- The need to strengthen national statistical systems to report on the SDGs and encourage innovations in data collection to simplify the tasks at hand;
- The need for countries to focus on the indicators that are most meaningful for their implementation priorities (all 232 will not be relevant in all cases); and
- Overcoming monitoring that is driven by the availability of data rather than its relevance for national priorities.[130]

---

124 Ibid., p. 6.

125 Ibid., p. 9.

126 OECD, *Development Co-operation Report 2017: Data for Development* (Paris, 2017), p. 24. Available at https://read.oecd-ilibrary.org/development/development-co-operation-report-2017_dcr-2017-en#page25.

127 UNDP, "Guidance Note: Data for Implementation and Monitoring of the 2030 Agenda for Sustainable Development", September 2017, p. 4. Available at http://www.undp.org/content/undp/en/home/librarypage/poverty-reduction/guidance-note--data-for-implementation-and-monitoring-of-the-203.html.

128 Neil Webster and Helle Munk Ravnborg, "Monitoring the implementation of the Sustainable Development Goals – The role of the data revolution", *European Parliament Directorate-General for External Policies*, July 2016. Available at http://www.europarl.europa.eu/RegData/etudes/STUD/2016/578020/EXPO_STU(2016)578020_EN.pdf.

129 Ibid., p. 11.

130 Asian Development Bank, "From Goals to Action: Implementing the Sustainable Development Goals – Seminar Background Note", 2017, p. 6. Available at https://www.adb.org/sites/default/files/publication/301696/goals-action-sdgs.pdf.

Data can help governments in SDG implementation. With the data revolution, "policymaking can be improved by exploiting the massive streams of accurate, timely and granular data".[131]

It is also the case that:

> Data and often good data are available in the vast majority of developing countries, but they still need to be improved in many ways. The need is to ensure that the right data, the relevant data, the data that are of good quality, are collected on a timely basis.[132]

Despite the push for the greater use of data in governance, a 2017 survey of public officials and development practitioners from 126 low- and middle-income countries revealed that "leaders use data or analysis more to conduct retrospective assessments of past performance than inform future policies and programmes".[133]

The slow adoption of data-driven governance can be attributed to technical and political reasons.

The technical issues include: lack of data standards; poor data quality; absence of interoperability framework; and poor information governance.[134]

However, "bureaucratic politics" is among the biggest obstacle to its widespread and rapid adoption. In a world where information is power, there are still officials who are unwilling to modernize data collection and production and/or afraid of sharing/releasing data for fear of losing it.

Fortunately, strong seeds for data-driven governance have already been planted. Enlightened officials, strong demand from civil society groups and pressure from international agencies are pushing for improved data collection, the processing and sharing of public data, and their use in government.[135]

---

131 OECD, *Development Co-operation Report 2017: Data for Development* (Paris, 2017), p. 27. Available at https://read.oecd-ilibrary.org/development/development-co-operation-report-2017_dcr-2017-en#page27.

132 Neil Webster and Helle Munk Ravnborg, "Monitoring the implementation of the Sustainable Development Goals – The role of the data revolution", *European Parliament Directorate-General for External Policies*, July 2016, p. 33. Available at http://www.europarl.europa.eu/RegData/etudes/STUD/2016/578020/EXPO_STU(2016)578020_EN.pdf.

133 Takaaki Masaki and others, "Decoding Data Use: How do leaders source data and use it to accelerate development? – Executive Summary", *AidData*, n.d., p.1. Available at http://docs.aiddata.org/ad4/pdfs/Decoding_data_use--Executive_summary.pdf.

134 Kathleen Hickey, "What's really needed for data-driven government", *GCN*, 22 December 2016. Available at https://gcn.com/articles/2016/12/22/data-governance-challenges-solutions.aspx.

135 Ben Rossi, "Data-driven government: Oxymoron or reality?" *Information Age*, 14 January 2016. Available at http://www.information-age.com/data-driven-government-oxymoron-or-reality-123460782/.

3. A WORLD AWASH IN DATA

This section aims to:

- Define official statistics;
- Discuss big data;
- Explain real-time data;
- Take up small data; and
- Examine citizen-generated data.

Despite the data deluge that the world is experiencing, high-quality data for decision-making remains a challenge in developing countries. According to *A World That Counts*: "Too many countries still have poor data, data arrives too late and too many issues are still barely covered by existing data."[136]

In this section, we will explore the various data sources for data-driven governance.

3.1 Official Statistics

While there is increasing interest in big data, traditional sources of data like official statistics remain important.

Official statistics are "numerical data sets, produced by official governmental agencies mainly for administrative purposes".[137] It is "a subcategory of 'scientific data' that more precisely help to understand how the societies function and evolve".[138]

The purpose of official statistics, according to the United Nations Economic Commission for Europe, is: "To produce and disseminate authoritative results designed to reliably reflect economically and socially relevant phenomena of a complex and dynamic reality in a given country."[139]

136 United Nations Secretary-General's Independent Expert Advisory Group on the Data Revolution for Sustainable Development, "A World that Counts: Mobilizing the Data Revolution for Sustainable Development", November 2014, p. 11. Available at http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf.

137 Sage Research Methods, "Official Statistics". Available at http://methods.sagepub.com/book/key-concepts-in-social-research/n34.xml (accessed 8 January 2019).

138 Walter J. Radermacher, "The Future Role of Official Statistics", Power from Statistics Outlook Report, p. 2. Available at https://www.researchgate.net/publication/320616460_The_Future_Role_of_Official_Statistics.

139 United Nations Economic Commission for Europe, "How Should a Modern National System of Official Statistics Look?" January 2008. Available at https://www.unece.org/fileadmin/DAM/stats/documents/applyprinciples.e.pdf.

Governments use official statistics in planning, decision-making, and monitoring or assessments of policies.[140] The private sector uses them to support business decisions. Citizens use official statistics to judge government's performance and make government accountable.

Unlike private sector data, government data is:

- **Comprehensive** – Statistics agencies aim to cover as much of the population as possible;
- **Consistent** – Statistics agencies have a long-term focus and provide consistent definitions of key measures over time, which is necessary to interpret the most recent estimates; and
- **Credible** – Statistics agencies must meet transparency requirements imposed by law and international bodies.[141]

The last point needs to be emphasized. Official statistics are based on common principles, standards, methodologies and technologies established in accordance with a professional code of ethics.[142]

Box 10. The Fundamental Principles of Official Statistics

The following ten principles have been adopted by the United Nations Economic and Social Council in its 46th plenary meeting on 24 July 2013:

1. Official statistics provide an indispensable element in the information system of a democratic society, serving the government, the economy and the public with data about the economic, demographic, social and environmental situation. To this end, official statistics that meet the test of practical utility are to be compiled and made available on an impartial basis by official statistical agencies to honour citizens' entitlement to public information.
2. To retain trust in official statistics, the statistical agencies need to decide according to strictly professional considerations, including scientific principles and professional ethics, on the methods and procedures for the collection, processing, storage and presentation of statistical data.
3. To facilitate a correct interpretation of the data, the statistical agencies are to present information according to scientific standards on the sources, methods and procedures of the statistics.
4. The statistical agencies are entitled to comment on erroneous interpretation and misuse of statistics.
5. Data for statistical purposes may be drawn from all types of sources, be they statistical surveys or administrative records. Statistical agencies are to choose the source with regard to quality, timeliness, costs and the burden on respondents.

140 Statistics South Africa, "Purpose of official statistics, and statistical principles". Available at http://www.statssa.gov.za/?page_id=750 (accessed 8 January 2019).

141 Nicholas Eberstadt and others, "'In Order That They Might Rest Their Arguments on Facts': The Vital Role of Government-Collected Data", *Hamilton Project and American Enterprise Institute*, March 2017, p. 4. Available at https://www.brookings.edu/wp-content/uploads/2017/02/thp_20170227_govt_collected_data_report.pdf.

142 Walter J. Radermacher, "The Future Role of Official Statistics", Power from Statistics Outlook Report, p. 2. Available at https://www.researchgate.net/publication/320616460_The_Future_Role_of_Official_Statistics.

6. Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes.
7. The laws, regulations and measures under which the statistical systems operate are to be made public.
8. Coordination among statistical agencies within countries is essential to achieve consistency and efficiency in the statistical system.
9. The use by statistical agencies in each country of international concepts, classifications and methods promotes the consistency and efficiency of statistical systems at all official levels.
10. Bilateral and multilateral cooperation in statistics contributes to the improvement of systems of official statistics in all countries.

Extracted (with modifications) from: Resolution adopted by the Economic and Social Council on 24 July 2013 (E/RES/2013/21). Available at https://unstats.un.org/unsd/dnss/gp/FP-Rev2013-E.pdf.

Traditionally, official statistics are comprised of census data, survey data and administrative data.

A census is: "A count for official purposes, especially one to count the number of people living in a country and to obtain information such as age, sex, race, etc."[143] A population census is: "The operation that produces at regular intervals the official counting (or benchmark) of the population in the territory of a country and in its smallest geographical sub-territories together with information on a selected number of demographic and social characteristics of the total population."[144]

The essential features that make a census unique are:

- **Individual enumeration** – Information on each enumerated person is obtained so that their characteristics can be separately recorded;
- **Simultaneity** – Information obtained should refer to a well-defined and unique reference period (or specific moment in time);
- **Universality** – The counting (or benchmarking) of the population should include every person residing and/or present in the defined territory of a country at a defined singular point in time;
- **Small area data** – The census should produce data on the number and characteristics of the population related to the smallest geographic areas of the country, and to small population subgroups, consistent with the overriding requirement to protect individual confidentiality; and
- **Defined periodicity** – The census should be taken at regular intervals so that comparable information is made available in a fixed sequence. It is recommended that census data be produced at least every ten years.[145]

143 Cambridge Dictionary, "Census". Available at https://dictionary.cambridge.org/us/dictionary/english/census (accessed on 8 January 2019).

144 United Nations Economic Commission for Europe, *Conference of European Statisticians Recommendations for the 2020 Censuses of Population and Housing* (New York and Geneva, 2015), p. 5. Available at https://www.unece.org/fileadmin/DAM/stats/publications/2015/ECECES41_EN.pdf.

145 Ibid., pp. 6-7.

The following determine the content of the census:

- The demand for data at national and local levels;
- The availability of data from other statistical sources; and
- The constraints of a census for data collection where (for traditional censuses at least) only a limited number of questions can be asked on single topics, and where sensitive or more complex topics that require extended modules and specialized training of interviewers can be covered only to a limited extent.[146]

Survey data comes from a sample survey—"a study of a subset of a population, in order to estimate population attributes".[147] A "survey usually enumerates a smaller proportion of the total population",[148] while a census is a full count. Survey data provides a subnational estimate of a given characteristic of the population, and is collected because of rapid changes in the size and other characteristics of populations, and increasing demand for additional detailed data on social, economic and housing characteristics that are not appropriate for census collection.[149]

Administrative data can be defined as "data collected by sources external to statistical offices",[150] and is usually collected routinely on all individuals affected by a particular programme. Administrative data offers longitudinal study opportunities.[151] Compared to survey results, administrative data has fewer problems associated with attrition, non-response and underreporting in survey information.

Administrative data sources include:

- Tax data (personal income tax, value-added tax, business/profits tax, property taxes, import/export duties);
- Social security data (contributions, benefits, pensions);
- Health/education records;
- Registration systems for persons/businesses/property/vehicles;
- Identity cards/passports/driving licenses;
- Electoral registers;
- Register of farms;
- Local council registers;
- Building permits;
- Licensing systems (television, sale of restricted goods);

---

146 Ibid., pp. 7-8.

147 Stat Trek, "Statistics Dictionary: Sample Survey". Available at http://stattrek.com/statistics/dictionary.aspx?definition=sample%20survey (accessed on 8 January 2019).

148 United Nations Department of Economic and Social Affairs, *Handbook on Population and Housing Census Editing: Revision 1* (New York, 2010) p. 1. Available at https://unstats.un.org/unsd/publication/SeriesF/seriesf_82rev1e.pdf.

149 United Nations Economic Commission for Europe, *Conference of European Statisticians Recommendations for the 2020 Censuses of Population and Housing* (New York and Geneva, 2015), p. 11. Available at https://www.unece.org/fileadmin/DAM/stats/publications/2015/ECECES41_EN.pdf.

150 United Nations Economic Commission for Europe, *Using Administrative and Secondary Sources for Official Statistics: A Handbook of Principles and Practices* (New York and Geneva, 2011) p. 2.

151 Robert Doar and Linda Gibbs, "Unleashing the Power of Administrative Data: A Guide for Federal, State, and Local Policymakers", *Results for America and American Enterprise Institute*, October 2017, p. 5. Available at https://results4america.org/wp-content/uploads/2017/10/Unleashing-the-Power-of-Administrative-Data.pdf.

- Published business accounts;
- Internal accounting data held by businesses; and
- Private businesses with data holdings (credit agencies, business analysts, utility companies, telephone directories, retailers with store cards).[152]

Statistical uses of administrative data include:

- Use for survey frames, directly as the frame or to supplement/update an existing frame;
- Replacement of data collection (use of taxation data for small businesses in lieu of seeking survey data for them);
- Use in editing and imputation;
- Direct tabulation;
- Indirect use in estimation (as auxiliary information in calibration estimation, benchmarking or calendarization); and
- Survey evaluation, including data confrontation (comparison of survey estimates with estimates from a related administrative programme).[153]

The advantages of using administrative data are: cost reduction; reduced response burden on data suppliers; frequency; coverage; timeliness; and enhanced public image.[154]

It is important to highlight that administrative data is not meant to replace survey data in official statistics. Both have value for data-driven governance. As noted by Doar and Gibbs: "The ability to link administrative data sets with each other and survey data offers significant potential to answer important questions that neither type of data can do by themselves."[155]

Official statistics remain important even in the midst of the data revolution. AidData's 2017 study of how leaders use data showed that "leaders use national statistics and evaluation data most frequently and also find them to be the most helpful sources of development data".[156]

Despite attempts at being comprehensive there is a lack of official data on important concerns like gender.

---

152 United Nations Economic Commission for Europe, *Using Administrative and Secondary Sources for Official Statistics: A Handbook of Principles and Practices* (New York and Geneva, 2011) p. 4.

153 Statistics Canada, "Use of administrative data", 16 June 2017. Available at http://www.statcan.gc.ca/pub/12-539-x/2009001/administrative-administratives-eng.htm.

154 United Nations Economic Commission for Europe, *Using Administrative and Secondary Sources for Official Statistics: A Handbook of Principles and Practices* (New York and Geneva, 2011), pp. 7-10.

155 Robert Doar and Linda Gibbs, "Unleashing the Power of Administrative Data: A Guide for Federal, State, and Local Policymakers", *Results for America and American Enterprise Institute*, October 2017, p. 4. Available at https://results4america.org/wp-content/uploads/2017/10/Unleashing-the-Power-of-Administrative-Data.pdf.

156 Takaaki Masaki and others, "Decoding Data Use: How do leaders source data and use it to accelerate development? – Executive Summary", *AidData*, n.d., p.1. Available at http://docs.aiddata.org/ad4/pdfs/Decoding_data_use--Executive_summary.pdf.

A World that Counts notes that "gender inequality and the undervaluing of women's activities and priorities in every sphere has been replicated in the statistical record".[157]

The lack of statistics on gender is unfortunate as "it is through the collection, production, analysis and use of gender statistics that policymakers and development practitioners can begin to properly address the specific issues of women and men".[158] More so, as noted by Emily Courey Pryor: "When we fail to measure critical dimensions of women's and girls' lives—everything from their economic contributions to their degrees of access to vital services—we undervalue their roles and experiences in society, entrench biases, and ultimately, leave women and girls behind."[159]

Box 11. The Challenges of Quantifying Women's Work

Women are often not paid for the work they do. Every day they chalk up an average of 4.5 hours of free labour in household chores and childcare regardless of where they are in the world. In developing countries, women can do 10 times as much as men.

If the hours of unpaid labour done by women worldwide were paid at minimum wage, they would be worth at least USD 10 trillion more than the gross domestic product of China, according to a conservative estimate by McKinsey.

But these are just estimates. In truth, we do not have sufficient data on much of the work done by women around the world. The same society and infrastructure that collects and applies data from every facet of our lives cannot efficiently track the hours women work and get paid outside official workplaces.

Papa Seck, Chief Statistician of UN Women who leads a project on evidence and data for gender equality, said in an interview that labour surveys are often based on old economic theories that assume only the husband can provide information about the household.

When questions are asked about property, or work, only his answers are taken into account, so the work of a woman may not be accounted for. Even when both husband and wife respond to a labour survey, questions may not be nuanced enough to capture work that doesn't earn a wage. Other times, data is aggregated without keeping gender into account.

Extracted (with modifications) from: Annalisa Merelli, "There's a mind-boggling amount of work women do that we literally can't quantify", Quartz, 18 May 2016. Available at https://qz.com/686075/we-still-have-literally-no-way-to-quantify-exactly-how-much-work-women-do/.

157 United Nations Secretary-General's Independent Expert Advisory Group on the Data Revolution for Sustainable Development, "A World that Counts: Mobilizing the Data Revolution for Sustainable Development", November 2014, p. 14. Available at http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf.

158 United Nations Department of Economic and Social Affairs, "Using data to measure gender equality", 4 November 2014. Available at http://www.un.org/en/development/desa/news/gender/using-data-to-measure-gender-equality.html.

159 Emily Courey Pryor, "'A World That Counts' Everyone, Including Women and Girls", *United Nations World Data Forum*, 27 December 2016. Available at https://undataforum.org/WorldDataForum/a-world-that-counts-everyone-including-women-and-girls.

Despite its shortcomings, official statistics remain important for people's empowerment:

> High-quality statistics strengthen democracy by allowing citizens access to key information that enhances accountability. Access to solid statistics is a fundamental 'right' that permits choices and decisions based on information. Without statistics there cannot be a well-grounded and participated democracy.[160]

3.2 Big Data

A popular definition of big data is: "Data sets whose size is beyond the ability of typical database software tools to capture, store, manage and analyse."[161] Some suggest that this is a moving definition (e.g., what is typical today may be atypical tomorrow).

Another influential definition is: "An information asset characterized by the three Vs (volume, variety and velocity)."[162] In this definition "volume" refers to the amount of data, "variety" refers to the number of types of data and "velocity" refers to the speed of data processing. In this view, big data is not only about the incredible amount of data, but also the wide diversity and rapid speed that data is generated.

Others suggest that there are not only three but five Vs of big data,[163] adding "veracity" (the messiness or trustworthiness of the data) and "value" (benefit or worth) to volume, variety and velocity.

The hype over big data has given rise to a critical definition: "A cultural, technological and scholarly phenomenon that rests on the interplay of – (1) Technology: Maximizing computation power and algorithmic accuracy to gather, analyse, link and compare large data sets; (2) Analysis: Drawing on large data sets to identify patterns in order to make economic, social, technical and legal claims; and (3) Mythology: The widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity and accuracy."[164]

Another way of looking at big data is through its features. The 10 common characteristics of big data are:

1. **Big** – Large data sets are a means to an end, they are not an end in themselves;
2. **Always on** – Always-on big data enables the study of unexpected events and real-time measurements;
3. **Non-reactive** – Measurements in big data sources are much less likely to change behaviour;
4. **Incomplete** – No matter how big your data, it probably does not have the information you want;
5. **Inaccessible** – Data held by companies and governments is difficult for researchers to access;

---

160 Walter J. Radermacher, "The Future Role of Official Statistics", Power from Statistics Outlook Report, p. 3. Available at https://www.researchgate.net/publication/320616460_The_Future_Role_of_Official_Statistics.

161 James Manyika and others, "Big Data: The next frontier for innovation, competition, and productivity", *McKinsey Global Institute*, May 2011. Available at https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_exec_summary.ashx.

162 Margaret Rouse, "Definition: 3Vs (volume, variety and velocity)", *WhatIs.com*, February 2013. Available at http://whatis.techtarget.com/definition/3Vs.

163 Bernard Marr, "Big Data: The 5 Vs Everyone Must Know", *LinkedIn*, 6 March 2014. Available at https://www.linkedin.com/pulse/20140306073407-64875646-big-data-the-5-vs-everyone-must-know.

164 danah boyd and Kate Crawford "Critical Questions For Big Data: Provocations for a cultural, technological, and scholarly phenomenon", *Information, Communication & Society*, vol. 15, no. 5 (June 2012), p. 663. Available at https://people.cs.kuleuven.be/~bettina.berendt/teaching/ViennaDH15/boyd_crawford_2012.pdf.

6. **Non-representative** – Non-representative data is bad for out-of-sample generalizations, but can be quite useful for within-sample comparisons;
7. **Drifting** – Populations drift, usage drift and system drift make it hard to use big data sources to study long-term trends;
8. **Algorithmically confounded** – Behaviour in big data systems is not natural, it is driven by the engineering goals of the systems;
9. **Dirty** – Big data sources can be loaded with junk and spam; and
10. **Sensitive** – Some of the information that companies and governments have is sensitive.[165]

Among the sources of big data are web searches, credit card transactions, emails (Gmail uses algorithms to scan and then analyse the content of emails), social media posts (Facebook harvests and uses all information that is created while using its platform) and smart watches and activity trackers (that measure your biometrics such as heart rate, heart rate variability and body temperature). Governments, as well as universities and research institutions, also produce data sets that can be used for analysis.

However, some proponents limit big data to:

- **Exhaust data** – Passively collected data from people's use of digital services such as mobile phones, financial transactions or web searches.
- **Sensing data** – Actively collected data from sensors, e.g., in smart cities or from wearables, and through remote sensing and satellite images.
- **Digital content** – Open web content actively produced by people such as social media interactions, news articles, blogs or job postings. Unlike exhaust and sensing data, this is digital content intentionally edited by somebody, i.e., subjective or even deceptive, depending on the intentions of the author.[166]

Big data is important because it allows us to "harness information in novel ways to produce useful insights or goods and services of significant value".[167]

---

165 Matthew J. Salganik, *Bit By Bit: Social Research in the Digital Age* (Princeton and Oxford, Princeton University Press, 2018), pp. 17-41.

166 Soenke Ziesche, *Innovative Big Data Approaches for Capturing and Analyzing Data to Monitor and Achieve the SDGs* (Bangkok, ESCAP, 2017), p. 18. Available at https://reliefweb.int/sites/reliefweb.int/files/resources/Innovative%20Big%20Data%20Approaches%20for%20Capturing%20and%20Analyzing%20Data%20to%20Monitor%20and%20Achieve%20the%20SDGs.pdf.

167 Viktor Mayer-Schonberger and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work and Think* (London, John Murray, 2013).

Box 12. Big Data in Business

Below are some examples of how big data is used in business:

Profiling models and ad targeting – Every time you log on to Google or Facebook and see ads, they are based on your preferences, browsing history, Facebook likes, groups joined, what your friends liked and so on.

Revenue management – Every time you try to buy an air ticket online, the prices vary on the basis of the route, demand, expected last-minute demand, how early you book and so on.

Recommendation engines – Every time you log on to e-retail sites and look at a product, you will start getting recommendations for other products also considered by other visitors. If you end up buying something, you will get recommendations for other products that bundle with it. For example, buy a phone and it will recommend a case or screen protector. If you make international calls, you may get a recommendation for a call package. The idea is to convert what is unguaranteed future income (no guarantee when you make your next international call) to guaranteed monthly revenue (by getting you to buy a call package that allows a certain number of international calls at a discounted rate).

Extracted (with modifications) from: Adityavijay Rathore's response to, "What are real-life examples of the application of big data analytics?" Quora, 9 January 2015. Available at https://www.quora.com/What-are-real-life-examples-of-the-application-of-big-data-analytics.

Big data can be used for development and governance.

The United Nations Global Pulse suggests that the potential applications of big data for development fall under three categories:

- **Early warning** – Early detection of anomalies in how populations use digital devices and services can enable faster response in times of crisis;
- **Real-time awareness** – Big data can paint a fine-grained and current representation of reality that can inform the design and targeting of programmes and policies; and
- **Real-time feedback** – The ability to monitor a population in real time makes it possible to understand where policies and programmes are failing, and make the necessary adjustments.[168]

Emmanuel Letouzé proposes an alternative taxonomy of the uses of big data for development.[169] Big data can be:

- **Descriptive** – Early detection of anomalies in how populations use digital devices and services can enable faster response in times of crisis;
- **Predictive** – Big data can paint a fine-grained and current representation of reality that can inform the design and targeting of programmes and policies; and
- **Prescriptive or diagnostic** – The ability to monitor a population in real time makes it possible to understand where policies and programmes are failing, and make the necessary adjustments.

168 United Nations Global Pulse, "Big Data for Development: Challenges and Opportunities", May 2012. Available at http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseMay2012.pdf.

169 Emmanuel Letouzé, "Big data for development: Facts and figures", *SciDevNet*, 15 April 2014. Available at https://scidev.net/global/data/feature/big-data-for-development-facts-and-figures.html.

Box 13. How Thailand is Using Big Data to Power the Government

After setting up Internet access for all, the Thai government is turning its attention to another facet of Thailand 4.0—data.

Putting big data first

As a first step, the Thai government will collate data from all 20 ministries into a centralized big data management system. All ministries will have three initial tasks: checking lists of their data sets, identifying data sets, and defining focal points of usage for public benefit.

As most of the data sets in Thailand are "traditional and unstructured", simply converting them into electronic data will be a big challenge. But when the system is fully integrated, all government agencies will have access to the data, enabling them to better implement policies and facilitate the country's digital transformation.

In addition, the data sets will be shared with the public, so startups and investors can use the government's data to develop solutions.

Big data in the government

Of course, the government also plans to use big data to make itself more efficient.

Sak Segkhoonthod, President of Thailand's Digital Government Development Agency, thinks that data will have an impact on important government functions like budgeting, planning and solving citizens' problems.

He explains: "First of all, the government will be able to ensure the budget goes to the appropriate avenues. Secondly, there will be better transparency throughout the government, as data becomes more easily accessible. Thirdly, citizens will have an opportunity to participate in certain things with the government as they can now get the meaningful information quicker."

Nuttapon Nimmanphatcharin, President and CEO of Thailand's Digital Economy Promotion Agency, agreed that an open system of big data could help combat corruption—something that's "in the pipeline for the near future".

Extracted (with modifications) from: Wen Chuan Tan, "How Thailand is using big data to power the government", Tech in Asia, 18 October 2018. Available at https://www.techinasia.com/thailand-big-data-government.

There are definite advantages in analysing enormous amounts of data.

Small samples are not ideal for studying rare events (incidents that occur with low frequency but may have widespread impact), heterogeneity (dissimilar or diverse elements), and small (but significant) differences.[170]

Big data can lead to better targeted policies and refinement of existing measures, and development of new indicators. By minimizing the need for survey, big data can also reduce the costs of statistical production.[171]

But "big" is not necessarily always the best.

---

170 Matthew J. Salganik, *Bit By Bit: Social Research in the Digital Age* (Princeton and Oxford, Princeton University Press, 2018), pp. 17-21.

171 Walter J. Radermacher, "The Future Role of Official Statistics", Power from Statistics Outlook Report, p. 3. Available at https://www.researchgate.net/publication/320616460_The_Future_Role_of_Official_Statistics.

For one, big data may not be representative of the group or category being studied. In developing countries, big data will only generate insights about sections of society that are connected. The insights generated will likely be true only of a small proportion and more privileged section of society.

There are other challenges in using big data: (1) they are not the result of a process consistent with standard practice; (2) they do not fit the methodologies, classifications and definitions of official statistics, and are therefore difficult to harmonize and convey in statistical structures; and (3) they raise security, privacy, data ownership and sustainability of access issues.[172]

It is also important to remember that big data is not a silver bullet.

Take the SDG indicators. An analysis showed that 70 SDG indicators (or almost one-third of all indicators) are not suitable for big data calculations.[173] To be more specific:

- Over 45 per cent of the Tier 3 indicators (i.e., no internationally-established methodology or standards are yet available for the indicator, but methodology/standards are being developed or tested), and almost 11 per cent of Tier 2 indicators (i.e., indicators are conceptually clear, internationally-established methodology or standards are available, but data is not regularly produced by countries) are not suitable for big data analysis.
- The SDGs 12 (Responsible Consumption and Production), 13 (Climate Action) and 17 (Partnership for the Goals) have the highest number of indicators that are not suitable for big data.
- The SDGs 12 and 13 also happen to have the highest numbers of Tier 2 and Tier 3 indicators.[174]

3.3 Real-Time Data

Real-time data refers to data that is delivered and used immediately after collection. However, in the development field, it also refers to: "Information which is produced and made available in a relatively short and relevant period of time, and information which is made available within a time frame that allows action to be taken in response."[175]

Real-time data includes social media feeds, satellite imagery, sensor-monitored rainfall and flood levels, and smartphone location data. To enable timely decisions, real-time data needs digital technology for its collection, sharing, management, analysis and reporting.[176]

Real-time data can increase the capacity of actors to respond to changes in operating contexts, learn from constant evaluations of the effectiveness of actions, uncover anomalies, respond to issues as they arise, improve internal coordination, optimize resource allocation, react to citizen feedback, and anticipate trends and future events.[177]

---

172 Ibid.

173 Soenke Ziesche, *Innovative Big Data Approaches for Capturing and Analyzing Data to Monitor and Achieve the SDGs* (Bangkok, ESCAP, 2017), p. 98. Available at https://reliefweb.int/sites/reliefweb.int/files/resources/Innovative%20Big%20Data%20Approaches%20for%20Capturing%20and%20Analyzing%20Data%20to%20Monitor%20and%20Achieve%20the%20SDGs.pdf.

174 Ibid.

175 United Nations Global Pulse, "Big Data for Development: Challenges and Opportunities", May 2012, p. 15. Available at http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseMay2012.pdf.

176 Andreas Pawelke and others, *Data for development: What's next? – Concepts, trends and recommendations for German development cooperation* (Bonn and Eschborn, Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH, 2017), p. 36. Available at http://webfoundation.org/docs/2017/12/Final_Data-for-development_Whats-next_Studie_EN-1.pdf.

177 Ibid., pp. 38-39.

At present, very few governments use real-time data to track programme implementation. Governments track activities and progress on a quarterly basis (if not more infrequently) using indicators as the benchmark to assess progress. With real-time data, progress can be monitored on a daily or weekly basis.

In the United States of America, real-time data has been successfully used to prevent health care fraud (see Box 14).

Box 14. Real-Time Data Analytics in Government

In 2016, the United States Department of Justice and the Office of Inspector General for the United States Department of Health and Human Services (HHS-OIG) issued the annual Health Care Fraud and Abuse Control (HCFAC) Programme Report showing that for every dollar spent on health care-related fraud and abuse investigations through this and other programmes in the last three years, the government recovered USD 6.10. "These impressive recoveries," said former HHS Secretary Kathleen Sebelius, "are due in large part to the new computer analytics system that detects and stops fraudulent billing before money ever goes out the door".

The fiscal year 2015 HCFAC Programme Report explains that HHS-OIG's complex data analysis tools include data mining, predictive analytics, trend evaluation and modelling approaches that better analyse and target the oversight of HHS programmes. The report further details that the Health Care Fraud Prevention and Enforcement Action Teams use near-time data to determine fraud patterns, identify suspected fraud trends and calculate ratios of allowed services as compared with national averages, as well as other assessments. These advanced technologies and techniques, taken from private industry, have led to historic efforts to bring innovation to the fight against health-care fraud.

Extracted (with modifications) from: John M. LeBlanc, "Real-Time Data Analytics in Government Investigations", Manatt, 23 June 2016. Available at https://www.manatt.com/Insights/Articles/2016/Real-Time-Data-Analytics-in-Government-Investigations.

In Africa, Pulse Lab Kampala is collaborating closely with the Ugandan government on various projects to monitor the quality of public service delivery in real time as Uganda has prioritized monitoring and evaluation in its National Development Plan.[178]

However, the wider use of real-time data in government has been hindered by competing data sources, distrust in data quality, lack of awareness of available data, data not adequately turned into information, lack of visualization, and information that is not tailored to the end user's needs.[179]

178 Pulse Lab Kampala, "Analytics in Real Time Can Help Monitor the Quality of Public Service Delivery", *United Nations Global Pulse*, 15 June 2016. Available at https://www.unglobalpulse.org/real-time-analytics-for-public-service-delivery.

179 Andreas Pawelke and others, *Data for development: What's next? – Concepts, trends and recommendations for German development cooperation* (Bonn and Eschborn, Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH, 2017), p. 39. Available at http://webfoundation.org/docs/2017/12/Final_Data-for-development_Whats-next_Studie_EN-1.pdf.

3.4 Small Data

Small data is promoted as "the human-centric alternative to big data".[180] While both big and small data are collecting and processing data to gain insights, the latter relies on a mix of keen observation of small samples and applied intuition.

For Martin Lindstrom, small data:

> Could be contained in a toothbrush holder in a bathroom in Tel Aviv; or in how a roll of toilet paper presses up against a bathroom wall in northern Brazil. It could show up in how a family's shoe collection is arrayed in a hallway, or in the scrambled letters and numbers that make up a person's computer password.[181]

Small data is about clues that leads to comprehension: "A lone piece of small data is almost never meaningful enough to build a case or create a hypothesis but blended with other insights and observations ... that data eventually comes together to create a solution that forms the foundation of a future brand or business".[182]

Collecting small data is part of what Lindstrom calls "subtext research"—a process that includes:

- **Collecting** – Capturing as many different perspectives from as many trustworthy sources as possible, and establishing navigation points to help frame initial observations and create a hypothesis;
- **Clues** – Artefacts (small data) that help create a narrative—a cohesive story that supports or disproves the hypothesis, maybe physical or emotional;
- **Connecting** – Making sense of small data (are the clues beginning to tilt in one direction?);
- **Causation** – "Small mining"—looking for cause and effect;
- **Correlation** – Finding mutual relationship or connection;
- **Compensation** – Identifying unmet desire or concern; and
- **Concept** – Defining the "solution" for the desire or concern that has been identified.[183]

Lindstrom compares his process to ethnography—"the recording and analysis of a culture or society, usually based on participant-observation and resulting in a written account of a people, place or institution".[184]

Another definition of small data is "small data sets that are capable of impacting decisions in the present".[185] A related definition is "data in a volume and format that makes it accessible, informative and actionable".[186]

---

180 Roger Dooley, "Small Data: The Next Big Thing", *Forbes*, 16 February 2016. Available at https://www.forbes.com/sites/rogerdooley/2016/02/16/small-data-lindstrom/#1811e5fb7870.

181 Martin Lindstrom, *Small Data: The Tiny Clues that Uncover Huge Trends* (New York, Picador, 2017), p. 9.

182 Ibid.

183 Ibid., pp. 219-225.

184 Discover Anthropology, "Ethnography", *Royal Anthropological Institute*. Available at https://www.discoveranthropology.org.uk/about-anthropology/fieldwork/ethnography.html (accessed on 8 January 2019).

185 Sahil Miglani, "Big Data and Small Data: What's the Difference?" *Dataversity*. Available at http://www.dataversity.net/big-data-small-data/ (accessed on 8 January 2019).

186 Ahmed Banafa, "Small Data vs. Big Data: Back to the Basics", *OpenMind*, 25 July 2016. Available at https://www.bbvaopenmind.com/en/small-data-vs-big-data-back-to-the-basics/.

Small data "connects people with timely, meaningful insights (derived from big data and/or "local" sources), organized and packaged—often visually—to be accessible, understandable and actionable for everyday tasks".[187] It includes baseball scores, inventory reports, driving records, sales data, search histories, weather forecasts and usage alerts.[188]

Still unclear? Here is John Spacey's take: "Small data was previously simply known as data. The modern term is used to distinguish between traditional data configurations and big data."[189]

Box 15. Why Small Data?

- **Big data is hard**: Doing it at scale and waiting for trickle-down benefits can take time. Not to mention the fact that most marketers and online strategists do not need full-on big data to target their campaigns or deliver personalized experiences.
- **Small data is all around us**: Social channels are rich with small data that is ready to be collected to inform marketing and buyer decisions. At a personal level, we are constantly creating this small data each time we check in, search, browse, post etc., creating a unique signature that provides a glimpse into our digital and physical health.
- **Small data is at the centre of the new customer relationship management**: Small data is the key to building rich profiles that will be the centre of the new customer relationship management solutions.
- **Return on investment**: A focus on the last mile of big data offers to leverage investments in small data (USD 10 billion and counting according to IDC) spent on upstream systems, tools and services.
- **Data-driven marketing is the next wave**: Big (and small) data-driven marketing has the potential to revolutionize the way businesses interact with customers, transform how customers access and consume (and even wear) useful data, and ultimately redefine the relationship between buyers and sellers.
- **Consumer examples abound**: Consumers have seen the potential of small data to streamline their shopping, power their fitness routine, or deliver recommendations about the best price for their next flight. With more smart, wearable data-driven devices on the way, there promises to be even more market demand for packaged data and data-delivery devices that "fit" the needs of everyday consumers.
- **Platform and tool vendors are starting to pay attention**: The promise of operationalizing big data and "turning insight into action" is a major tone from many of the big names in tech including SAP, Oracle and EMC.
- **It's about the end user**: It includes what they need and how they can take action. Focus on the user first and a lot of our technology decisions become clearer.
- **Simple**: Small data is the right data. Some small data will start life as big data, but you should not need to be a data scientist to understand or apply it for everyday tasks.

Extracted (with modifications) from: Ahmed Banafa, "Small Data vs. Big Data: Back to the Basics", OpenMind, 25 July 2016. Available at https://www.bbvaopenmind.com/en/small-data-vs-big-data-back-to-the-basics/.

187 Small Data Group, "Defining Small Data", 18 October 2013. Available at https://smalldatagroup.com/2013/10/18/defining-small-data/.

188 Margaret Rouse, "Small Data", *WhatIs.com*, June 2014. Available at https://whatis.techtarget.com/definition/small-data.

189 John Spacey, "8 Examples of Small Data", *Simplicable*, 30 January 2018. Available at https://simplicable.com/new/small-data.

3.5 Citizen-Generated Data

Citizen-generated data is "data that people or their organizations produces to directly monitor, demand or drive change on issues that affect them".[190] Examples include citizen-created data on air quality in Beijing, a sexual harassment map in Egypt (see Box 16), and updated water point statuses in Tanzania.

Citizen-generated data "gives people agency over decisions on what data is collected, what it is used for and how it is used".[191] It can "highlight issues that are important to people and feed their views up into higher-level policy debates".[192] It can also be used to verify official narratives and data sets".[193] It can bring "accountability to development processes through independent monitoring".[194] Most of all, "data actively produced by citizens with a specific purpose in mind can provide a direct representation of citizens and enable them to directly monitor, demand or drive change on issues that affect them".[195]

Development Initiatives' two case studies in Kenya and Uganda show that citizen-generated data initiatives have contributed to "improvements in the delivery of development-related projects' accessibility and/or quality of public services", and "empowerment and participation of local actors in accountability efforts".[196] They also reveal the potential of citizen-generated data to contribute to SDG monitoring at the local level.[197]

Citizen-generated data complements institutional data, and should not be seen as a replacement or alternative. Citizen-generated data is particularly useful in trying to understand communities where data is lacking, like women, the vulnerable and marginalized sectors of the population.

---

190 DataShift, "What is Citizen-Generated Data and What is the Datashift Doing to Promote it?", n.d. Available at http://civicus.org/images/ER%20cgd_brief.pdf.

191 Andreas Pawelke and others, *Data for development: What's next? – Concepts, trends and recommendations for German development cooperation* (Bonn and Eschborn, Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH, 2017), p. 32. Available at http://webfoundation.org/docs/2017/12/Final_Data-for-development_Whats-next_Studie_EN-1.pdf.

192 Ibid.

193 Ibid.

194 Development Initiatives, "Citizen-generated data and sustainable development: Evidence from case studies in Kenya and Uganda report", March 2017, p. 10. Available at https://hivos.org/sites/default/files/publications/15-citizen-generated-data-and-sustainable-development-evidence-from-case-studies-in-kenya-and-uganda.pdf.

195 Andreas Pawelke and others, *Data for development: What's next? – Concepts, trends and recommendations for German development cooperation* (Bonn and Eschborn, Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH, 2017), p. 32. Available at http://webfoundation.org/docs/2017/12/Final_Data-for-development_Whats-next_Studie_EN-1.pdf.

196 Development Initiatives, "Citizen-generated data and sustainable development: Evidence from case studies in Kenya and Uganda report", March 2017, p. 5. Available at https://hivos.org/sites/default/files/publications/15-citizen-generated-data-and-sustainable-development-evidence-from-case-studies-in-kenya-and-uganda.pdf. The Kenya case examined the School Report Card initiative, an effort to increase parent participation in schooling of their children. The Uganda case examined a process of providing unsolicited citizen feedback to duty bearers and service providers in local communities in Uganda.

197 Andreas Pawelke and others, *Data for development: What's next? – Concepts, trends and recommendations for German development cooperation* (Bonn and Eschborn, Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH, 2017), p. 30. Available at http://webfoundation.org/docs/2017/12/Final_Data-for-development_Whats-next_Studie_EN-1.pdf.

Box 16. Mapping Sexual Harassment in Egypt

HarassMap was launched in December 2010 by a founding group of four women together with tech partners, advisers and volunteers. Below, the four women reflect on the establishment and impact of HarassMap:

"We were all overwhelmed by the sexual harassment that we, and almost everyone we knew, were exposed to on a daily basis.

In 2009, a volunteer introduced us to Frontline SMS and Ushahidi, which are free software that can be linked together to make an anonymous reporting and mapping system that can be used online and through short message service or SMS.

Since about 97 per cent of Egyptians then—half of whom were women—owned a mobile phone, this technology seemed like an opportunity to re-engage the public in this issue. It took a year to develop our model and we analysed the situation and crafted an approach to target the areas we felt were not being addressed by the advocacy-focused programmes of non-governmental organizations that were already active at the time.

We wanted to be sure that HarassMap would never be 'just a map' and it was important to us that it would have a strong community-based component that could make a strong impact on the ground.

The starting point was to use the reporting and mapping technology to support an offline community mobilization effort to break stereotypes, stop making excuses of perpetrators, and convince people to speak out and act against sexual harassment.

It makes us very happy to see that over the last years, sexual harassment has evolved from being a taboo topic to one that is widely discussed. Our team of volunteers continues to grow and we keep expanding our work on the ground and online."

Extracted (with modifications) from: HarassMap, "Our Story". Available at https://harassmap.org/who-we-are/our-story (accessed on 8 January 2019).

4. DATA GOVERNANCE AND DATA MANAGEMENT

This section aims to:

- Discuss data governance and data management, with a focus on Data Management Association International's Data Management Body of Knowledge.

Data can enhance governance when it is easily accessible, effortlessly shared and smoothly reusable.

But data sharing is extremely difficult when government agencies use different ICT systems and store data in different formats, and when overlapping data sets and information silos exist. The problem is compounded when information security and data quality standards are lax. As Jelani Harper notes:

> The distributed nature of data assets in the public sector is even more complex than that in the private sector. Governmental entities contend with multiple databases with varying access points, locations and architectures, as well as security issues of regional, national and international concern.[198]

Data governance and data management are key to ensuring data accessibility, reliability, quality and timeliness for users.

4.1 Data Governance

Data governance is the "comprehensive process for controlling the integrity, use, availability, usability and security of all data owned by or controlled by an enterprise".[199] It is "a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods".[200]

In both the public and private sectors, the key trends driving the need for data governance include:

- Increasing data volumes from more and more sources, causing data inconsistencies that need to be identified and addressed, before decisions are made using incorrect information;
- More self-service reporting and analytics (data democratization), creating the need for a common understanding of data across the organization;
- The continuing impact of regulatory requirements, making it even more important to have a strong handle on what data is where, and how it is being used; and
- An increasing need for a common business language to enable cross-departmental analysis and decisions.[201]

198 Jelani Harper, "Lessons learned from big interoperability in government", *KMWorld*, 30 October 2017. Available at http://www.kmworld.com/Articles/Editorial/Features/Lessons-learned-from-big-interoperability-in-government-121253.aspx.

199 Nate Lord, "What is Data Governance? Data Protection 101", *Digital Guardian*, 10 September 2018. Available at https://digitalguardian.com/blog/what-data-governance-data-protection-101.

200 Data Governance Institute, "Definitions of Data Governance". Available at http://www.datagovernance.com/adg_data_governance_definition/ (accessed on 8 January 2019).

201 Nancy Couture, "Why data governance?" *CIO*, 3 January 2018. Available at https://www.cio.com/article/3245588/governance/why-data-governance.html.

The goals of data governance are to:

- Allow an organization to manage its data as an asset;
- Define, approve, communicate and implement principles, policies, procedures, metrics, tools and responsibilities for data management; and
- Monitor and guide policy compliance, data usage and management services.[202]

Data governance:

- Enables better decision-making;
- Reduces operational friction;
- Protects the needs of data stakeholders;
- Trains management and staff to adopt common approaches to data issues;
- Define standard and builds repeatable processes;
- Decreases costs;
- Increases effectiveness through coordination of efforts; and
- Ensures transparency of processes.[203]

A body (e.g., data governance council, committee or group) is usually chartered and given the responsibility to set the strategic direction for WHAT the data governance programme needs to accomplish, and WHEN it needs to accomplish it.[204] This body also ensures everyone knows about it and that it gets the support it needs.

Data governance activities include:

1. Defining data governance for the organization –
   - Develop data governance strategy (a plan for maintaining and improving data quality, integrity, security and access);
   - Perform readiness assessment;
   - Perform discovery and business alignment; and
   - Develop organizational touchpoints.
2. Defining the data governance strategy –
   - Define the data governance operating framework;
   - Develop goals, principles and policies;
   - Underwrite data management projects;

---

202 DAMA International, *Data Management Body of Knowledge*, second edition (DAMA-DMBOK 2) (New Jersey, Technics Publications, 2017), Chapter 3, pp. 67-95.

203 Data Governance Institute, "Goals and Principles for Data Governance". Available at http://www.datagovernance.com/adg_data_governance_goals/ (accessed on 8 January 2019).

204 George F., "Data governance council – what is it and why do you need one?" *Lights on Data*, 18 July 2018. Available at http://www.lightsondata.com/data-governance-council/.

- Engage in change management;
- Engage in issue management; and
- Assess regulatory compliance requirements.

3. Implementing data governance –
   - Sponsor data standards and procedures;
   - Develop business glossary;
   - Coordinate with architecture groups; and
   - Sponsor data asset valuation.
4. Embedding data governance.[205]

The primary deliverables are:

- Data governance strategy;
- Data strategy;
- Data strategy road map;
- Data principles;
- Data governance policies and processes;
- Operating framework;
- Road map and implementation strategy;
- Operation plan;
- Business glossary;
- Data governance scorecard;
- Data governance website;
- Communications plan;
- Recognized data value; and
- Maturing data management practices.[206]

NASCIO—an association of the United States of America representing state chief information officers and information technology executives and managers—advocates that governments:

- Place data governance as a priority agenda item to be investigated and addressed as part of an overall ICT governance initiative;
- Understand data governance as part of the larger scope of knowledge asset management;

---

205 DAMA International, *Data Management Body of Knowledge*, second edition (DAMA-DMBOK 2) (New Jersey, Technics Publications, 2017), Chapter 3, pp. 67-95.

206 Ibid.

- Understand data governance must be properly planned and chartered (with policy, organization, process, commitment, communications, frameworks, methods and procedures, valuation, and tools);
- Leverage a maturity model for planning manageable phases in data governance; and
- Make communication of the benefits and intended outcomes of data governance a high priority in order to gain and sustain involvement from stakeholders.[207]

In South-East Asia, Thailand and Indonesia have recognized the importance of data governance.

Box 17. Satu Data Indonesia

Data stewards in the public sector can benefit from easy-to-understand technical guidelines in order to effectively implement new governance procedures. The Draft Presidential Regulation on Satu Data Indonesia proposes such guidance.

The three areas that present opportunities to strengthen data governance in the public sector are:

1. **Organizational structure** – Poor coordination between government entities is one of the difficulties data stewards continue to encounter. To improve this, clear guidelines on the organizational structure can clarify coordination procedures related to the governance of data. These guidelines may include a comprehensive explanation of the data coordination steps, how to match competent data stewards with a particular task, and how to go about organizing a data governance forum among government stakeholders.
2. **Data flow** – Although data stewards in the public sector understand the importance of maintaining data quality, not all of them are fully aware of the necessary processes—from the stage of data collection to the point of data usage in policymaking. Guidelines that can translate this overall process into clear steps can help data stewards to better perform their duties. Specifically, five segments of the data journey have been identified that can help actualize the Satu Data Indonesia policy. These segments include: data collection and distribution flow, standardization flow, validation flow, interoperability flow, and data dissemination flow.
3. **Data format** – Generally, data stewards across agencies and ministries in the public sector do not have a single, common format for cataloguing digital data. As a result, individuals use whatever preferred formats, and this makes it less efficient when it comes to sharing data between government counterparts. The development of a standard data format guideline has the benefit of helping all players meet the interoperability principles outlined in Satu Data Indonesia, while simultaneously improving data management skills related to collecting, processing and distributing data. This guideline can also serve as a glossary of data terminologies, metadata format and data type format.

Extracted (with modifications) from: Pulse Lab Jakarta, "Identifying Opportunities to Strengthen Data Governance in the Public Sector", Medium, 8 October 2017. Available at https://medium.com/pulse-lab-jakarta/identifying-opportunities-to-strengthen-data-governance-in-the-public-sector-92d9bc7db5ad.

Data governance is distinct from data management.

207 NASCIO, "Data Governance – Managing Information as an Enterprise Asset: Part I – An Introduction", April 2008, p. 11. Available at https://www.nascio.org/Portals/0/Publications/Documents/NASCIO-DataGovernance-Part1.pdf.

4.2 Data Management

Data management is: "The development and execution of processes, architectures, policies, practices and procedures in order to manage the information generated by an organization."[208] It includes acquiring, validating, storing, protecting and processing data to ensure its accessibility, reliability and timeliness for its users.[209]

Data management helps ensure the availability of data when and where it is needed. It is necessary particularly when government data is stored in a variety of formats and systems that support specific programmes, departments and organizations.[210]

The benefits of good data management include:

- Improved awareness and understanding of what data is available for current and future use, resulting from better cataloguing and data archiving;
- Improved access to data, free from unnecessary obstacles, safeguarded from disclosure of personal information or infringement of legal and contractual obligations;
- Better quality and more timely information i.e., access to the right information at the right time, resulting from quicker identification of customer needs and the avoidance of wrong or conflicting information, through the use of effective metadata;
- Better value for money, resulting from clear, fair and consistent data charges and conditions of use, which recognize the need for free access by the appropriate customers;
- Better exploitation of data generally, enabled by easier data exchange and integration with other harmonized data; and
- Efficiency gains across government and its agencies, resulting from the use of better quality data.[211]

Data management is implemented "through a cohesive infrastructure of technological resources and a governing framework that defines the administrative processes used throughout the life cycle of data".[212]

Data management ensures an organization gets value out of its data. its data. It is guided by data governance.

Governments interested in data-driven governance should consider using the Data Management Association (DAMA) International's *Guide to the Data Management Body of Knowledge* (DAMA-DMBOK 2). The guide is a collection of processes and knowledge areas that are generally accepted as best practices within the data management discipline.

---

208 Blue-Pencil, "What is Data Management and Why it is Important", 23 November 2015. Available at http://www.blue-pencil.ca/what-is-data-management-and-why-it-is-important/.

209 Molly Galetto, "What is Data Management?" *NGDATA*, 31 March 2016. Available at https://www.ngdata.com/what-is-data-management/.

210 Francesca El-Attrash, "What Government Needs to Know About Data Management", *GovLoop*, 9 August 2017. Available at https://www.govloop.com/government-needs-know-data-management/.

211 Intra-Governmental Group on Geographic Information, *The Principles of Good Data Management*, second edition (London, Office of the Deputy Prime Minister, 2005). Available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/14867/Good_dataMan.pdf.

212 Technopedia, "Data Management". Available at https://www.techopedia.com/definition/5422/data-management (accessed on 8 January 2019).

A key weakness of DAMA-DMBOK 2 is that it considers data governance as an element of data management. However, this does not detract from its value as a useful comprehensive guide to understanding and implementing data management.

For the rest of this section, the focus will be on the following DAMA-DMBOK 2 knowledge areas: data architecture; data modelling and design; data storage and operations; data security; data integration and interoperability; documents and content management; reference and master data; data warehousing and business intelligence; metadata management; and data quality management.

4.2.1 Data Architecture

Data architecture refers to: "An organized arrangement of component elements intended to optimize the function, performance, feasibility, cost and aesthetics of an overall structure or system [...] It identifies the data needs of the enterprise (regardless of structure), and designs and maintains the master blueprint to meet those needs."[213]

The goals of data architecture management are to:

- Identify data storage and processing requirements;
- Design structures and plans to meet the current and long-term data requirements of the enterprise; and
- Strategically prepare organizations to quickly evolve their products, services and data to take advantage of opportunities inherent in emerging technologies.

Activities include:

1. Establishing enterprise data architecture –
   - Evaluate existing data architecture specifications;
   - Develop a road map; and
   - Manage enterprise requirements within projects.
2. Integrating with enterprise architecture. The primary deliverables are:
   - Data architecture design;
   - Data flows;
   - Data value chains;
   - Enterprise data model; and
   - Implementation road map.

213 This subsection is drawn from: DAMA International, *Data Management Body of Knowledge*, second edition (DAMA-DMBOK 2) (New Jersey, Technics Publications, 2017), Chapter 4, pp. 97-120.

Box 18. The Six Principles of Modern Data Architecture

1. **View data as a shared asset**: Enterprises that start with a vision of data as a shared asset ultimately outperform their competition.
2. **Provide the right interfaces for users to consume the data**: Putting data in one place is not enough to achieve the vision of a data-driven organization. In order for people (and systems) to benefit from a shared data asset, you need to provide the interfaces that make it easy for users to consume that data.
3. **Ensure security and access controls**: Look to technologies that allow you to architect for security, and deliver broad self-service access, without compromising control.
4. **Establish a common vocabulary**: Without this shared vocabulary, you will spend more time disputing or reconciling results than driving improved performance.
5. **Curate the data**: Without proper data curation (which includes modelling important relationships, cleansing raw data, and curating key dimensions and measures), end users can have a frustrating experience—which will vastly reduce the perceived and realized value of the underlying data.
6. **Eliminate data copies and movement**: Every time data is moved there is an impact on cost, accuracy and time. By eliminating the need for additional data movement, modern enterprise data architectures can reduce cost (time, effort, accuracy), increase "data freshness" and optimize overall enterprise data agility.

Extracted (with modifications) from: Joshua Klahr, "The 6 Principles Of Modern Data Architecture", AtScale, 19 January 2018. Available at https://www.atscale.com/blog/the-six-principles-of-modern-data-architecture.

4.2.2 Data Modelling and Design

Data modelling and design refers to: "The process of discovering, analysing and scoping data requirements, and then representing and communicating these data requirements in a precise form called the data model."[214]

The goal is to confirm and document an understanding of different perspectives, which leads to applications that more closely align with current and future business requirements. This creates a foundation to successfully complete broad-scoped initiatives such as master data management and data governance programmes.

Activities include:

1. Planning for data modelling.
2. Building data models –
   - Create the conceptual data model;
   - Create the logical data model; and
   - Create the physical data model.

214 This subsection is drawn from: DAMA International, *Data Management Body of Knowledge*, second edition (DAMA-DMBOK 2) (New Jersey, Technics Publications, 2017), Chapter 5, pp. 123-166.

The primary deliverables are:

- Conceptual data model;
- Logical data model; and
- Physical data model.

Box 19. What is Data Modelling?

Data modelling is the process of creating a data model for the data to be stored in a database.

Data modelling helps in the visual representation of data, and enforces business rules, regulatory compliances and government policies on the data. Data models ensure consistency in naming conventions, default values, semantics and security while ensuring quality of the data.

Data models emphasize what data is needed and how it should be organized instead of what operations need to be performed on the data. A data model is like an architect's building plan that helps to build a conceptual model and set the relationship between data items.

Extracted (with modifications) from: Guru99, "What is Data Modelling? Conceptual, Logical, & Physical Data Models". Available at https://www.guru99.com/data-modelling-conceptual-logical.html (accessed on 8 January 2019).

4.2.3 Data Storage and Operations

Data storage and operations include the design, implementation and support of stored data to maximize its value through its life cycle, from creation/acquisition to disposal.[215] Data storage and operations involve two subactivities—database support and data technology support.

The goals are to:

- Manage the availability of data throughout the data life cycle;
- Ensure the integrity of data assets; and
- Manage the performance of data transactions.

Activities include:

1. Managing database technology –
   - Understand and evaluate database technology; and
   - Manage and monitor database technology.

215 This subsection is drawn from: *DAMA International, Data Management Body of Knowledge*, second edition (DAMA-DMBOK 2) (New Jersey, Technics Publications, 2017), Chapter 6, pp. 169-214.

2. Managing database operations –
   - Understand requirements;
   - Plan for business continuity;
   - Develop database instances;
   - Manage database performance;
   - Manage test databases; and
   - Manage data migration.

The primary deliverables are:

- Database technology evaluation criteria;
- Database environments;
- Migrated/replicated/versioned data;
- Business continuity plans; and
- Database performance service-level agreements.

Box 20. Types of Data to Retain

The first step in building an effective data storage policy is to answer the question: Which types of data should I store for an extended period, and which can I delete instantly?

Below is a general hierarchy that outlines which types of data to keep on hand. The data at the top of the list is the most important to store for as long as possible, while the data at the bottom is least important:

- Data that is required to be retained by compliance or regulatory policies. If you are required by law to store a certain type of data, you should definitely keep that data around.
- Data that relates to your customers and helps you engage with them by achieving "customer 360". Understanding your customers is hard, and you do not want to give up the data that helps you with that challenge.
- Business documents, contracts and so on. These are important to store for as long as possible.
- Data that is generated by everyday business operations but is not regulated. This data can be helpful to have on hand for historical reviews or planning purposes, but it is not essential.
- Machine data generated by your networking equipment, servers, sensors or other types of automated sources. Machine data tends to be the least useful type of data to store long term. It is sometimes useful to be able to review machine data when researching a technical incident or planning infrastructure expansions, but for the most part, machine data is only useful in real time, because the state of your infrastructure changes so quickly.

The exact types of data to prioritize for long-term storage will vary from organization to organization, of course. This hierarchy is just a general guide.

Extracted (with modifications) from: Christopher Tozzi, "Best Practices in Data Storage (Part 1): What Types of Data Should be Retained?" Syncsort, 17 July 2017. Available at http://blog.syncsort.com/2017/07/big-data/data-storage-best-practices-data-types/.

4.2.4 Data Security

Data security is the "planning, development and execution of security policies and procedures to provide proper authentication, authorization, access and auditing of data and information assets".[216]

The goals are to:

- Enable appropriate, and prevent inappropriate, access and change to data assets;
- Understand and comply with all relevant regulations and policies for privacy, protection and confidentiality; and
- Ensure that the privacy and confidentiality needs of all stakeholders are enforced and audited.

Activities include:

1. Identifying relevant data security requirements;
2. Defining data security policy;
3. Defining data security standards;
4. Assessing current security risks; and
5. Implementing controls and procedures.

The primary deliverables are:

- Data security architecture;
- Data security policies;
- Data privacy and confidentiality standards;
- Data security access controls;
- Regulatory compliant data access views;
- Documented security classifications;
- Authentication and user access history; and
- Data security audit reports.

---

216 This subsection is drawn from: DAMA International, *Data Management Body of Knowledge*, second edition (DAMA-DMBOK 2) (New Jersey, Technics Publications, 2017), Chapter 7, pp. 217-266.

Box 21. Threat Prevention

Organizations often spend most of their information security resources implementing risk management, incident prevention and anti-virus solutions. These traditional activities are still required, but they only form a single layer within a more complex strategy. To decrease the attack surface across endpoints and networks, and to make a successful attack as difficult as possible, organizations must:

- **Build a foundation of tight controls and processes, unique to your organization** – Only you will understand the characteristics of your organization and your users. Build processes unique to you and ensure user roles are clearly defined.
- **Educate employees to be your first line of defense** – By educating employees, you can greatly reduce the risk of vulnerabilities caused by human error. Ensure all staff members have access to proper training, even on subjects that seem basic to you, such as remembering to log off workstations when they leave their desks or keeping their passwords secret.
- **Learn about industry-specific data breach scenarios** – Study threat intelligence to understand the specific types of attacks and patterns.
- **Take a layered approach to security technology** – A variety of security solutions will reduce the threat landscape and prevent advanced attacks on your network. Data encryption is a must. Supplement endpoint security with network-based controls.

Extracted (with modifications) from: Ryan St. Hilaire, "Data Security Best Practices Not Good Enough", eSecurity Planet, 8 April 2015. Available at https://www.esecurityplanet.com/network-security/data-security-best-practices-not-good-enough.html.

4.2.5 Data Integration and Interoperability

Data integration and interoperability "describe the processes related to the movement and consolidation of data within and between data stores, applications and organizations".[217]

The goals are to:

- Provide data securely, with regulatory compliance, in the format and time frame needed;
- Lower cost and complexity of managing solutions by developing shared models and interfaces;
- Identify meaningful events and automatically trigger alerts and actions; and
- Support business intelligence, analytics, master data management and operational efficiency efforts.

217 This subsection is drawn from: DAMA International, *Data Management Body of Knowledge*, second edition (DAMA-DMBOK 2) (New Jersey, Technics Publications, 2017), Chapter 8, pp. 269-299.

Activities include:

1. Planning and analysing –
   - Define data integration and life cycle requirements;
   - Perform data discovery;
   - Document data lineage;
   - Profile data; and
   - Examine business rules compliance.
2. Designing data integration and interoperability solutions –
   - Design solution components;
   - Map sources to targets; and
   - Design data orchestration.
3. Developing data integration and interoperability solutions –
   - Develop data services;
   - Develop data flow orchestration;
   - Develop data migration approach;
   - Develop complex event processing; and
   - Maintain data integration and interoperability metadata.
4. Implementing and Monitoring.

The primary deliverables are:

- Data integration and interoperability architecture;
- Data exchange specifications;
- Data access agreements;
- Data services;
- Complex event processing; and
- Thresholds and alerts.

Box 22. Data Interoperability and Disasters

Some of the difficulties of the response to, and recovery from, Hurricane Katrina could have been avoided if data communications had been interoperable. For example, if data interoperability had enhanced the interface between the government conducting evacuations, and private voluntary organizations assisting with evacuations and the sheltering facilities, it could have allowed them to more easily match parents with missing children.

Data interoperability does not only impact governments, but private sector entities as well. While most first response is in the hands of government, most of the critical infrastructure in the United States of America is owned by the private sector.

True interoperability would provide for an immediate situational awareness at both ends of the problem: The utility that must repair the problem and the government jurisdiction that must make arrangements for such things as medical response, generators for individuals on life support, water and sewage pumping stations and myriad other functions that are the responsibility of local government.

The impact of disasters in different parts of the world could be greatly improved through the use of international standards. The impact of such a standards process not only pertains to the United States, but will impact the international community as well. The independent organization should have international ties and membership, and the international community should have a place in the standards development process. A good example of the type of benefits we could derive from this process is the improved results we would have gained by the free flow of information on an international basis during the Asian Tsunami in 2004. While the response to that disaster was very good, better data communications could have proven useful in the response and recovery process.

Extracted (with modifications) from: Bill Lent, "Facing the Challenge Of Data Interoperability", Disaster Resource Guide. Available at http://disaster-resource.com/index.php?option=com_content&view=article&id=335%3Afacing-the-challenge-of-data-interoperability&catid=9%3Acrisis-response&Itemid=15 (accessed on 8 January 2019).

4.2.6 Documents and Content Management

Documents and content management means "controlling the capture, storage, access, and use of data and information stored outside relational databases".[218]

The goals are to:

- Comply with legal obligations and customer expectations regarding records management;
- Ensure effective and efficient storage, retrieval and use of documents and content; and
- Ensure integration capabilities between structured and unstructured content.

218 This subsection is drawn from: DAMA International, *Data Management Body of Knowledge*, second edition (DAMA-DMBOK 2) (New Jersey, Technics Publications, 2017), Chapter 9, pp. 303-344.

Activities include:

1. Planning for life cycle management –
   - Plan for records management; and
   - Develop a content strategy.
2. Creating content handling policies, including e-discovery approach.
3. Defining the information architecture and managing the content life cycle –
   - Capture and manage records and content;
   - Retain and dispose records and content; and
   - Archive records and content.
4. Publishing and delivering content.

The primary deliverables are:

- Content and records management strategy;
- Policy and procedure;
- Content repository;
- Managed records in many media formats; and
- Audit trail and log.

Box 23. Document Management among American State Governments

The following are the results of a 2016 Center for Digital Government survey of 203 state officials of the United States of America regarding their use of document management solutions:

1. **Document management dominates** – Over 80 per cent of respondents agree that improving document management is a priority for them personally and for their organization.
2. **Now is the time to modernize** – Sixty-eight per cent of organizations are actively modernizing their approach to document management, and 15 per cent plan to in the future.
3. **Process automation is a natural complement** – A quarter of the respondents plan to automate a process within 12 months.
4. **You can start small or go all in** – Thirty-seven per cent of organizations implement document management agency-wide, 23 per cent in a single department and 22 per cent statewide.
5. **Fast and configurable are key** – When looking to procure, over a third of organizations prefer a heavily configurable commercial-off-the-shelf solution. Only 10 per cent prefer to develop a solution in-house.

6. **Feature-rich is the new currency** – The most important automated document management features are: retrievable records (77 per cent); digitized electronic forms (59 per cent); electronic routing (50 per cent); and error recognition (30 per cent).
7. **Not all solutions are created equal** – Respondents say some unmet needs with current document management solutions are: ease of use (48 per cent); storage (35 per cent); and security (2 per cent).
8. **Procurement roadblocks exist** – Thirty-eight per cent of respondents say integration with other systems is the greatest challenge when justifying the need for a solution.
9. **Department heads seal the deal** – Sixty-four per cent of respondents say department heads, like C-suites and vice presidents, are the decision makers most commonly involved in the selection process.
10. **It's all about the 3Ss—security, simplicity and savings** – Drivers in selecting a solution are: 60 per cent security; 49 per cent simplicity and ease of use; and 45 per cent cost savings.

Extracted (with modifications) from: Laserfiche, "10 Things to Know About Document Management in the Public Sector". Available at https://www.laserfiche.com/ecmblog/infographic-10-things-to-know-about-document-management-in-the-public-sector/ (accessed on 8 January 2019).

4.2.7. Reference and Master Data

Reference and master data means "managing shared data to meet organizational goals, reduce risks associated with data redundancy, ensure higher quality and reduce the cost of data integration".[219]

The goals are to:

- Enable sharing of information assets across business domains and applications within an organization;
- Provide authoritative source of recorded and quality-assessed master and reference data; and
- Lower cost and complexity through the use of standards, common data models and integration patterns.

Activities include:

1. Identifying drivers and requirements;
2. Evaluating and assessing data sources;
3. Defining architectural approach;
4. Modelling data;
5. Defining stewardship and maintenance processes;
6. Establishing governance policies; and
7. Implementing data sharing/integration services.

219 This subsection is are drawn from: DAMA International, *Data Management Body of Knowledge*, second edition (DAMA-DMBOK 2) (New Jersey, Technics Publications, 2017), Chapter 10, pp. 347-379.

The primary deliverables are:

- Master and reference data requirements;
- Data models and documentation;
- Reliable reference and master data; and
- Reusable data services.

Box 24. What is Master Data and Reference Data?

Master Data

Master data is key business information that supports the transactions.

Master data describes the customers, products, parts, employees, materials, suppliers, sites, etc. involved in the transactions. It is commonly referred to as places (locations, geography, sites), parties (persons, customers, suppliers, employees) and things (products, items, material, vehicles).

Master data is usually authored and used in the normal course of operations by existing business processes. Unfortunately, these operational business processes are tailored for an "application-specific" use case of this master data and therefore fail in achieving the overall enterprise requirement that mandates commonly-used master data across applications with high quality standards and common governance.

Reference Data

Reference data is data that is referenced and shared by a number of systems. Most of the reference data refers to concepts that either impact business processes—e.g., order status (CREATED | APPROVED | REJECTED)—or is used as an additional standardized semantic that further clarifies the interpretation of a data record—e.g., employee job position (JUNIOR | SENIOR | VP).

Some of the reference data can be universal and/or standardized (e.g., countries – ISO 3166-1). Other reference data may be "agreed on" within the enterprise (customer status), or within a given business domain (product classifications).

Reference data is frequently considered as a subset of master data. The full name for this data category is "master reference data".

Extracted (with modifications) from: FX Nicolas, "Back to Basics: Transactional, Master, Golden and Reference Data explained", Semarchy, 10 October 2018. Available at https://blog.semarchy.com/backtobasics_data_classification.

4.2.8 Data Warehousing and Business Intelligence

Data warehousing and business intelligence mean "planning, implementation and control processes, to provide decision-support data and support knowledge workers engaged in reporting, query and analysis".[220]

The goals are to:

- Build and maintain the technical environment and technical and business processes needed to deliver integrated data in support of operational functions, compliance requirements and business intelligence activities; and
- Support and enable effective business analysis and decision-making by knowledge workers.

Activities include:

1. Understanding the requirement –
   - Define and maintain the data warehouse and business intelligence architecture;
2. Developing the data warehouse and data marts;
3. Populating the data warehouse;
4. Implementing the business intelligence portfolio; and
5. Maintaining data products.

The primary deliverables are:

- Data warehouse and business intelligence architecture;
- Data products;
- Population process;
- Governance activities;
- Lineage dictionary;
- Learning and adoption plan;
- Release plan;
- Production support process;
- Load tuning activities; and
- Business intelligence activity monitoring.

220 This subsection is drawn from: DAMA International, *Data Management Body of Knowledge*, second edition (DAMA-DMBOK 2) (New Jersey, Technics Publications, 2017), Chapter 11, pp. 381-414.

Box 25. What is Data Warehousing?

Data warehousing is a technique for collecting and managing data from varied sources to provide meaningful business insights. It is a blend of technologies and components that allows the strategic use of data.

It is electronic storage of a large amount of information by a business that is designed for query and analysis instead of transaction processing. It is a process of transforming data into information and making it available to users in a timely manner to make a difference.

Data warehouse is needed for all types of users like:

- Decision makers who rely on mass amount of data;
- Users who use customized, complex processes to obtain information from multiple data sources;
- People who want simple technology to access data;
- People who want a systematic approach for making decisions;
- Users who want fast performance on a huge amount of data, which is a necessity for reports, grids or charts; and
- Users who want to discover "hidden patterns" of data flows and groupings.

Extracted (with modifications) from: Guru99, "What Is Data Warehousing? Types, Definition & Example". Available at https://www.guru99.com/data-warehousing.html (accessed on 8 January 2019).

4.2.9 Metadata Management

Metadata or data about data "includes information about technical and business processes, data rules and constraints, and logical and physical data structures".[221]

The goals are to:

- Provide organizational understanding of terms and usage;
- Collect and integrate metadata from diverse source;
- Provide a standard way to access metadata; and
- Ensure metadata quality and security.

Activities include:

1. Defining metadata strategy.

221 This subsection is drawn from: DAMA International, *Data Management Body of Knowledge*, second edition (DAMA-DMBOK 2) (New Jersey, Technics Publications, 2017), Chapter 12, pp. 417-448.

2. Understanding metadata requirements –
   - Business users requirements; and
   - Technical requirements.
3. Defining metadata architecture –
   - Create metamodel;
   - Apply metadata standards; and
   - Manage metadata stores.
4. Creating and maintaining metadata –
   - Integrate metadata; and
   - Distribute and deliver metadata.
5. Querying, reporting and analysing metadata.

The primary deliverables are:

- Metadata strategy;
- Metadata standards;
- Metadata architecture;
- Metamodel;
- Unified metadata;
- Metadata stores;
- Data lineage;
- Impact analysis;
- Dependency analysis; and
- Metadata control process.

Box 26. Metadata Basics

In simple terms, metadata is "data about data", and if managed properly, it is generated whenever data is created, acquired, added to, deleted from or updated in any data store and data system in scope of the enterprise data architecture.

Metadata provides a number of very important benefits to the enterprise, including:

- **Consistency of definitions** – Metadata contains information about data that helps reconcile the difference in terminology such as "clients" and "customers", "revenues" and "sales", etc.

- **Clarity of relationships** – Metadata helps resolve ambiguity and inconsistencies when determining the associations between entities stored throughout the data environment. For example, if a customer declares a "beneficiary" in one application, and this beneficiary is called a "participant" in another application, metadata definitions would help clarify the situation.
- **Clarity of data lineage** – Metadata contains information about the origins of a particular data set and can be granular enough to define information at the attribute level. Metadata may maintain allowed values for a data attribute, its proper format, location, owner and steward. Operationally, metadata may maintain: auditable information about users, applications and processes that create, delete or change data; the exact timestamp of the change; and the authorization that was used to perform these actions.

Extracted (with modifications) from: Alex Berson and Larry Dubov, "The benefits of metadata and implementing a metadata management strategy", TechTarget. Available at https://searchitchannel.techtarget.com/feature/The-benefits-of-metadata-and-implementing-a-metadata-management-strategy (accessed on 8 January 2019).

4.2.10 Data Quality Management

Data quality management refers to: "Planning, implementation and control of activities that apply quality management techniques to data, in order to assure it is fit for consumption and meets the needs of data consumers."[222]

The goals are to:

- Develop a governed approach to make data fit for purpose based on data consumers' requirements;
- Develop standards, requirements and specifications for data quality controls as part of the data life cycle;
- Define and implement processes to measure, monitor and report on data quality levels; and
- Identify and advocate for opportunities to improve the quality of data through process and systems improvement.

Activities include:

1. Defining high quality data.
2. Defining a data quality strategy.
3. Defining scope of initial assessment –
   - Identify critical data; and
   - Identify existing rules and patterns.

222 This subsection is drawn from: DAMA International, *Data Management Body of Knowledge*, second edition (DAMA-DMBOK 2) (New Jersey, Technics Publications, 2017), Chapter 13, pp. 451-494.

4. Performing initial data quality assessment –
   - Identify and prioritize issues; and
   - Perform root cause analysis of issues.
5. Identifying and prioritizing improvements –
   - Prioritize actions base on business impact;
   - Develop preventive and corrective actions; and
   - Confirm planned actions.
6. Developing and deploying data quality operations –
   - Develop data quality operational procedures; and
   - Correct data quality.

The primary deliverables are:

- Improved quality data;
- Data management operational analysis;
- Data profiles;
- Data quality certification reports; and
- Data quality service-level agreements.

Box 27. The Six Dimensions of Data Quality

On the surface, it is obvious that data quality is about cleaning up bad data—data that is missing, incorrect or invalid in some way. But in order to ensure that data is trustworthy, it is important to understand the following six key dimensions of data quality to assess how the data is "bad" in the first place:

1. **Completeness** is defined as expected comprehensiveness. Data can be complete even if optional data is missing. As long as the data meets the expectations then the data is considered complete.
2. **Consistency** means data across all systems reflects the same information and they are in sync with each other across the enterprise.
3. **Conformity** means the data is following the set of standard data definitions like data type, size and format. For example, date of birth of customer is in the format "dd/mm/yyyy".
4. **Accuracy** is the degree to which data correctly reflects the real-world object or an event being described.
5. **Integrity** means validity of data across the relationships and ensures that all data in a database can be traced and connected to other data.
6. **Timeliness** references whether information is available when it is expected and needed. Timeliness of data is very important.

Extracted (with modifications) from: Somasekhar Thatipamula, "Data Done Right: 6 Dimensions of Data Quality", Smart Bridge, 9 August 2013. Available at https://smartbridge.com/data-done-right-6-dimensions-of-data-quality/.

4.2.11 Measuring Progress

One way of determining the state of data management in an enterprise is through a data management maturity assessment.

Data management maturity assessment is: "A method for ranking practices for handling data within an organization to characterize the current state of data management and its impact on the organization."[223]

Governments seeking to determine their level of data management maturity can use the (generic) six levels of data management maturity as follows:

- **Level 0 – No Capability**: No organized data management practices or formal enterprise processes for managing data.
- **Level 1 – Initial/Ad Hoc:** General purpose data management practices using a limited tool set, with little or no governance.
- **Level 2 – Repeatable:** Emergence of consistent tools and role definition to support process execution. Organization begins to use centralized tools and provide more oversight for data management.
- **Level 3 – Defined:** Emerging data management capability. Introduction and institutionalization of scalable data management processes and a view of data management as an organizational enabler.
- **Level 4 – Managed:** Standardized tools for data management from desktop to infrastructure, coupled with well-formed centralized planning and governance functions. Expressions of this level are a measurable increase in data quality and organization-wide capabilities such as end-to-end data audits.
- **Level 5 – Optimization:** Proliferation of data is controlled to prevent needless duplication. Well understood metrics are used to manage and measure data quality and processes.[224]

Increased data management maturity not only ensures accessible and quality data, but also future-proofs an enterprise. As Matthew Chase argues:

> Those agencies who have optimized data management strategies now will be set up for success in keeping pace with the many advanced applications for data that are coming down the pike, including data visualization, machine learning, artificial intelligence, and so on. Those who find themselves closer to the inactive end of the scale are likely to continue to fall further behind as technological innovations continue to become more sophisticated, and at an even faster rate.[225]

---

223 DAMA International, *Data Management Body of Knowledge*, second edition (DAMA-DMBOK 2) (New Jersey, Technics Publications, 2017), p. 533.

224 Ibid., pp. 534-536.

225 Matthew Chase, "The state of data management in the public sector in 2018", *Experian*, 8 February 2018. Available at https://www.edq.com/blog/the-state-of-data-management-in-the-public-sector-in-2018/.

5. DATA ANALYTICS, BIAS AND INTUITION

This section aims to

- Explain the four types of data analytics—descriptive, diagnostic, predictive and prescriptive;
- Examine algorithmic bias; and
- Explore the role of intuition in decision-making.

5.1 Data Analytics

Data analytics creates new values in data.

It is "the pursuit of extracting meaning from raw data using specialized computer systems ... that transform, organize and model the data to draw conclusions and identify patterns".[226]

There are four types of analytics—descriptive, diagnostic, predictive and prescriptive.[227]

5.1.1 Descriptive Analytics

Descriptive analytics provides insight to answer: *What has happened?*[228]

Data aggregation and data mining are the two main techniques used in descriptive analytics.[229] Outputs include summary of expenditures, tabulation of social metrics in government Facebook sites and Twitter feeds, and reporting of general trends like inflation or employment rate.

226 Informatica, "What is Data Analytics". Available at https://www.informatica.com/services-and-training/glossary-of-terms/data-analytics-definition.html#fbid=goi4oqkVXLo (accessed on 8 January 2019).

227 Alex Bekker, "4 types of data analytics to improve decision-making", *ScienceSoft*, 11 July 2017. Available at https://www.scnsoft.com/blog/4-types-of-data-analytics.

228 Halo, "Descriptive, Predictive, and Prescriptive Analytics Explained". Available at https://halobi.com/blog/descriptive-predictive-and-prescriptive-analytics-explained/ (accessed on 8 January 2019).

229 Anushka Mehta, "Four Types of Business Analytics to Know", *Analytics Insight*, 13 October 2017. Available at https://www.analyticsinsight.net/four-types-of-business-analytics-to-know/.

Box 28. Descriptive Analytics and Business Intelligence

Descriptive analytics is one of the most basic pieces of business intelligence a company will use. Although descriptive analytics can be very industry specific—such as the seasonal variation in shipment completion times—it can also include broadly-accepted measures that are used in finance. Return on invested capital is a descriptive analytic created by taking three data points—net income, dividends and total capital—and turning them into an easy-to-understand percentage that can be used to compare one company's performance to others. Generally speaking, the larger and more complex a company, the more descriptive analytics it will use to measure its performance.

Extracted (with modifications) from: Investopedia, "Descriptive Analytics", 10 March 2018. Available at https://www.investopedia.com/terms/d/descriptive-analytics.asp#ixzz5NYWlugiV.

Descriptive analytics rarely attempts to investigate or establish cause and effect relationships. It cannot provide an answer to important questions such as: *How can we avoid this problem? How can we duplicate this successful solution?* These are covered by diagnostic analytics.

5.1.2 Diagnostic Analytics

Diagnostic analytics measures historical data against other data to answer: *Why something happened?*

Diagnostic analytics focuses on the processes and causes. The techniques used include drill-down, data discovery, data mining and correlations.[230] Training algorithms for classification and regression also fall in this type of analytics.[231]

Diagnostic analytics is used to:

- **Identify anomalies** – Based on the results of descriptive analysis, analysts must identify areas that require further study because they raise questions that cannot be answered simply by looking at the data;
- **Drill into the analytics (discovery)** – Analysts must identify the data sources that will help them explain these anomalies. Often, this step requires analysts to look for patterns outside the existing data sets, and it may require pulling in data from external sources to identify correlations and determine if any of them are causal in nature; and
- **Determine causal relationships** – Hidden relationships are uncovered by looking at events that may have resulted in the identified anomalies[232].

230 CornerStone, "Diagnostic Analytics". Available at https://www.cornerstoneondemand.com/glossary/diagnostic-analytics (accessed on 8 January 2019).

231 Anushka Mehta, "Four Types of Business Analytics to Know", *Analytics Insight*, 13 October 2017. Available at https://www.analyticsinsight.net/four-types-of-business-analytics-to-know/.

232 Dan Vesset, "Diagnostic analytics 101: Why did it happen?" *IBM Business Analytics Blog*, 11 May 2018. Available at https://www.ibm.com/blogs/business-analytics/diagnostic-analytics-101-why-did-it-happen/.

Box 29. How To Do Diagnostic Analytics

1. Identify something worth investigating

The first step in doing diagnostic analytics is to find something that is worth investigating. Typically this is something bad, like a fall in revenue or clicks, but it can also be an unexpected performance boost.

2. Do the analysis

Diagnostic analytics may be as straightforward as finding a single root cause—i.e., revenue dropped last month because new customer sign-ups were down. More complex analyses, however, may require multiple data sets and the search for a correlation using regression analysis.

What you are trying to accomplish in this step is to find a statistically valid relationship between two data sets, where the rise (or fall) in one causes a rise (or fall) in another.

More advanced techniques in this area include data mining and principal component analysis, but straightforward regression analysis is a great place to get started.

3. Selectively filter your diagnoses

While it may be interesting that a variety of factors contributed to a change in performance, it is not helpful to list every possible cause in a report. Instead an analyst should aim to discover the single, or at most two, most influential factor(s) in the issue being diagnosed.

4. State your conclusion clearly

Finally, a diagnostic report must come to a conclusion and make a very clear case for it. It does not have to include all of the background work, but you should:

- Identify the issue you are diagnosing;
- State why you think it happened; and
- Provide your supporting evidence.

Extracted (with modifications) from: Jeff Rajeck, "Analytics approaches every marketer should know #2: Diagnostic analytics", Econsultancy, 31 July 2017. Available at https://www.econsultancy.com/blog/69300-analytics-approaches-every-marketer-should-know-2-diagnostic-analytics.

5.1.3 Predictive Analytics

Predictive analytics uses statistical models and forecasts techniques to answer: *What could happen?*[233]

It provides organizations with actionable insights by combining historical data to identify patterns in the data, and apply statistical models and algorithms to capture relationships between various data sets.

233 Halo, "Descriptive, Predictive, and Prescriptive Analytics Explained". Available at https://halobi.com/blog/descriptive-predictive-and-prescriptive-analytics-explained/ (accessed on 8 January 2019).

The private sector uses predictive analytics to forecast customer behaviour and purchasing patterns, identify trends in sales activities, and anticipate demand for inputs from the supply chain, operations and inventory.[234] This is clearly seen in the experience of Otto—Germany's second largest Internet retailer.[235] Otto allows its ICT system to purchase around 200,000 items a month from third-party partners without human intervention. What accounts for the confidence in the system? Otto's ICT system analyses around three billion past transactions and 200 variables (including past sales, searches on Otto's website and weather information) to predict (with 90 per cent accuracy) what customers will buy a week before they order.

Box 30. Retailer Uses Predictive Analytics to Target Pregnant Women

Andrew Pole was hired by Target (a large American retailer) to identify those unique moments in consumers' lives when their shopping habits become particularly flexible and the right advertisement or coupon would cause them to begin spending in new ways.

Among life events, none are more important than the arrival of a baby. At that moment, new parents' habits are more flexible than at almost any other time in their adult lives. If companies can identify pregnant shoppers, they can earn millions.

As Pole's computers crawled through the data, he was able to identify about 25 products that, when analysed together, allowed him to assign each shopper a "pregnancy prediction" score. More important, he could estimate her due date to within a small window, so Target could send coupons timed to very specific stages of her pregnancy.

About a year after Pole created his pregnancy-prediction model, a man walked into a Target outside Minneapolis and demanded to see the manager. He was clutching coupons that had been sent to his daughter, and he was angry, according to an employee who participated in the conversation.

"My daughter got this in the mail!" he said. "She's still in high school, and you're sending her coupons for baby clothes and cribs? Are you trying to encourage her to get pregnant?"

The manager didn't have any idea what the man was talking about. He looked at the mailer. Sure enough, it was addressed to the man's daughter and contained advertisements for maternity clothing, nursery furniture and pictures of smiling infants. The manager apologized and then called a few days later to apologize again.

On the phone, though, the father was somewhat abashed. "I had a talk with my daughter," he said. "It turns out there's been some activities in my house I haven't been completely aware of. She's due in August. I owe you an apology."

Extracted (with modifications) from: Charles Duhigg, "How Companies Learn Your Secrets", New York Times Magazine, 16 February 2012. Available at https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=1&_r=1&hp.

234 Ibid.

235 Economist, "How Germany's Otto uses artificial intelligence", 12 April 12, 2017. Available at https://www.economist.com/news/business/21720675-firm-using-algorithm-designed-cern-laboratory-how-germanys-otto-uses?etear=sasexpectexceptional.

An interesting use of predictive analytics for governance is in sentiment analysis. For instance, governments can predict citizen sentiment on particular issues as being positive, negative or neutral by analysing social media posts.[236]

Predictive analytics uses machine learning algorithms like random forests and support vector machine, and statistics for learning and testing the data. The most popular tools for predictive analytics include Python, R and RapidMiner.[237]

5.1.4 Prescriptive Analytics

Prescriptive analytics suggests what actions to take.

It uses a combination of techniques and tools such as business rules, machine learning and computational modelling procedures, and applies these to different data sets including historical and transactional data, and real-time data.[238]

Companies use prescriptive analytics to "optimize production, scheduling and inventory in the supply chain to make sure that they are delivering the right products at the right time, and optimizing the customer experience".[239]

Box 31. Improving the Health-Care Industry with Prescriptive Analytics

When health-care providers combine data sets such as patient records, medicine information, economic data, demographical and sociographical data, health trends, and hospital data, they will be able to offer better health care for less money, improve future capital investments for new facilities or hospital equipment, and improve the efficiency of hospitals.

The combination of different data sets can also be used to offer doctors recommendations in the best possible treatment for a patient. Thanks to combining and analysing multiple data sets, the Aurora Health Care Centre was able to improve health care and reduce re-admission rates by 10 per cent, thereby saving USD 6 million annually.

Furthermore, pharmaceutical organizations can benefit from prescriptive analytics by improving their drug development and reduce time-to-market for new medicines. Drugs simulations can improve medicines faster and it becomes easier to find the right patient for clinical trials based on multiple variables.

Extracted (with modifications) from: Mark van Rijmenam, "The Future of Big Data? Three Use Cases of Prescriptive Analytics", Datafloq. Available at https://datafloq.com/read/future-big-data-use-cases-prescriptive-analytics/668 (accessed on 8 January 2019).

---

236 Anushka Mehta, "Four Types of Business Analytics to Know", *Analytics Insight*, 13 October 2017. Available at https://www.analyticsinsight.net/four-types-of-business-analytics-to-know/.

237 Ibid.

238 Halo, "Descriptive, Predictive, and Prescriptive Analytics Explained". Available at https://halobi.com/blog/descriptive-predictive-and-prescriptive-analytics-explained/ (accessed on 8 January 2019).

239 Ibid.

While prescriptive analytics is believed to be "the future of big data, [...] it is still a long way away before it will be common language".[240]

5.1.5 Data Analytics in Government

Governments can use data analytics to enhance public service delivery, spur economic development by improving the efficiency of usage and allocation of resources, and reduce fraud, waste and abuse.[241]

Data analytics in government includes:

- **In administration** – Identifying the least-cost vendor with highest rating, the productivity of individuals based on sick leaves taken, departments that are not performing well, types of skills that will be needed, and whether spending will exceed budget and why;
- **In social services** – Particularly in child welfare, identifying most at-risk children, those least likely to be reunified and the best services for each child;
- **In policing** – Monitoring, tracking and mapping crime statistics on a daily basis. Combined with geospatial data visualizations, data analytics enables police departments to make better decisions about where to dispatch police resources, thereby increasing safety and cost efficiency; and
- **In health care** – Enabling scientific breakthroughs, improving efficiency in health care facilities, advancing disease prevention, diagnosis and treatment, and assisting in disease surveillance (see Box 32).

Box 32. Analytics and Disease Surveillance

Accurate and timely information on global public health issues is key to being able to quickly assess and respond to emerging health risks around the world. The Public Health Agency of Canada has developed the Global Public Health Intelligence Network (GPHIN). Information from GPHIN is provided to the World Health Organization, international governments and non-governmental organizations who can then quickly react to public health incidents.

240 Mark van Rijmenam, "The Future of Big Data? Three Use Cases of Prescriptive Analytics", *Datafloq*. Available at https://datafloq.com/read/future-big-data-use-cases-prescriptive-analytics/668 (accessed on 8 January 2019).

241 Sid Frank and Traci Gusher, "Better data, better government: Effective use of data and analytics at all levels deliver improved citizen services and outcomes", *KPMG Government Institute*, June 2016, p. 5. Available at https://assets.kpmg/content/dam/kpmg/pdf/2016/06/co-gv-6-better-data,-better-government.pdf.

GPHIN is a secure Internet-based early warning system that gathers preliminary reports of public health significance on a real-time basis, 24 hours a day, 7 days a week. This unique multilingual system gathers and disseminates relevant information on disease outbreaks and other public health events by monitoring global media sources such as news wires and websites. This monitoring is done in eight languages with machine translation being used to translate non-English articles into English and English articles into other languages. The information is filtered for relevancy by an automated process, which is then complemented by human analysis. The output is categorized and made accessible to users. Notifications about public health events that may have serious public health consequences are immediately forwarded to users.

GPHIN has a broad scope. It tracks events such as disease outbreaks, infectious diseases, contaminated food and water, bio-terrorism and exposure to chemicals, natural disasters, and issues related to the safety of products, drugs and medical devices.

Extracted (with modifications) from: Abla Mawudeku and Michael Blench, "Global Public Health Intelligence Network (GPHIN)", n.d. Available at http://mt-archive.info/MTS-2005-Mawudeku.pdf (accessed on 8 January 2019).

There are policy, organizational, cultural, resource and technical barriers preventing the more widespread use of analytics in governance.[242]

Governments face technical challenges that include: data in various formats; different degrees of data quality; multiple, distributed data silos; antiquated data management infrastructures; lack of access to private sector held big data; and unavailable advanced analytics tools.[243]

Organizational challenges include: lack of data governance organizations, processes or tools; absence of sponsors/champions to implement data governance; reluctance among personnel to share data, either based on valid legal/policy reasons or on rigid historical management methods; limited understanding of the "art of the possible" in the use of advanced analytics and big data to drive valuable insights; and changing priorities.[244]

5.2 Algorithmic Bias

Using analytics involves "applying an algorithmic or mechanical process to derive insights".[245]

An algorithm is: "A set of mathematical instructions or rules that, especially given to a computer, will help to calculate an answer to a problem."[246] Its efficiency depends on data—data quality "influences the value of the analytics and the confidence to make decisions based on the outputs".[247]

---

242 Ibid.

243 Ibid.

244 Ibid.

245 Avantika Monnappa, "Data Science vs. Big Data vs. Data Analytics", *simplilearn*, 2 March 2018. Available at https://www.simplilearn.com/data-science-vs-big-data-vs-data-analytics-article.

246 Cambridge Dictionary, "Algorithm". Available at https://dictionary.cambridge.org/dictionary/english/algorithm (accessed on 8 January 2019).

247 Royal Academy of Engineering, "Algorithms in decision-making: A response to the House of Commons Science and Technology Committee inquiry into the use of algorithms in decision-making", April 2017, p. 3. Available at https://www.raeng.org.uk/publications/responses/algorithms-in-decision-making.

Algorithms are used to process (make sense of) big data. Organizations—private and public—rely on algorithms to solve problems and make decisions because they "are able to process a far greater range of inputs and variables to make decisions, and can do so with speed and reliability that far exceed human capabilities".[248]

The United Kingdom Royal Academy of Engineering recommends that:

> Government, businesses and public bodies will need to consider their use of algorithms in decision-making, consulting widely, and ensuring that mechanisms are in place to detect and address any mistakes or unintended consequences of decisions made.[249]

This is because algorithms are not value free or neutral. Algorithmic bias occurs when human prejudice and partiality are incorporated in the design.[250] Consequently, discrimination is embedded into the model.

The more organizations use algorithms to support (or replace) humans in decision-making, the more algorithm bias becomes a concern. As noted by Will Knight: "If the bias lurking inside the algorithms that make ever-more-important decisions goes unrecognized and unchecked, it could have serious negative consequences, especially for poorer communities and minorities."[251]

Box 33. Biased Algorithms

In 2015, Google's photo app mistakenly tagged a photo of two black people as gorillas because its algorithm had not been trained with enough images of dark-skinned persons. In another case, the AI judge of a beauty contest mostly chose white participants as winners because its training was done on images of white people.

These are trivial cases that can be easily remedied by providing the AI with more samples in areas where it does not have enough data. In other cases where AI is working with vast amounts of existing data in the endless sea of online information, finding and countering bias become much more difficult.

Algorithmic bias can have even more damaging effect in other areas such as law enforcement. In 2016, a ProPublica investigation found that an AI-powered tool used by law enforcement was more likely to declare black people as under the high risk of recidivism than white people. In some states, judges rely on such tools to decide who stays in jail and who walks free, sometimes without doing further investigation themselves.

---

248 Keith Kirkpatrick, "Battling Algorithmic Bias", *Communications of the ACM*, vol. 59, no. 10 (October 2016), pp. 16-17. Available at https://cacm.acm.org/magazines/2016/10/207759-battling-algorithmic-bias/abstract.

249 Royal Academy of Engineering, "Algorithms in decision-making: A response to the House of Commons Science and Technology Committee inquiry into the use of algorithms in decision-making", April 2017, p. 4. Available at https://www.raeng.org.uk/publications/responses/algorithms-in-decision-making.

250 Keith Kirkpatrick, "Battling Algorithmic Bias," *Communications of the ACM*, vol. 59, no. 10 (October 2016), pp. 16-17. Available at https://cacm.acm.org/magazines/2016/10/207759-battling-algorithmic-bias/abstract.

251 Will Knight, "Biased Algorithms Are Everywhere, and No One Seems to Care", *MIT Technology Review*, 12 July 2017. Available at https://www.technologyreview.com/s/608248/biased-algorithms-are-everywhere-and-no-one-seems-to-care/.

Similar cases can happen in other areas such as loan approval, where people who are underrepresented will be further marginalized and deprived of service. In health care, where AI is making great inroads in diagnosing and curing diseases, algorithms can harm populations whose data has not been included in the training sets.

In fact, if not addressed, algorithmic bias can lead to the amplification of human biases. Under the illusion that software is not biased, humans tend to trust the judgement of AI algorithms, oblivious that those judgments are already reflecting their own prejudices. As a result, we will accept AI-driven decisions without doubting them and create more biased data for those algorithms to further "enhance" themselves on.

Extracted (with modifications) from: Ben Dickson, "What is algorithmic bias?" TechTalks, 26 March 2018. Available at https://bdtechtalks.com/2018/03/26/racist-sexist-ai-deep-learning-algorithms/.

Algorithmic bias can materialize as the result of problems at different stages.[252] These include:

- **Biased or otherwise poor quality input data** – The data may be biased, incomplete or of poor quality, potentially leading an algorithm to produce poor and perhaps discriminatory outcomes.
- **Poorly defined rules** – The data used as an input for algorithmic decisions may be poorly weighted.
- **Lack of contextual awareness** – The definition of quality of the training data and the robustness of the rules and weights are often context specific. Algorithms that work well within the context for which they were designed may discriminate if rolled out in a different context.
- **Feedback loops** – Algorithms do not operate in a vacuum. Their activity affects the environment from which they extract the data they use as input. A biased algorithm may reinforce its biases, in what could be deemed a self-fulfilling prophecy loop.[253]

Algorithm bias, according to Kate Crawford, can lead to two type of harms: allocative harm and representational harm.[254]

Allocative harm occurs when the algorithm apportions or withholds certain opportunities or resources based on prejudiced assumptions. An example is a bank loan risk assessment algorithm that systematically denies loan applications to women.

Representational harms "occur when systems reinforce the subordination of some groups along the lines of identity".[255] Here technology reinforces stereotypes. For instance, an algorithm that identifies East Asians are blinking when they are actually smiling.

---

252 World Wide Web Foundation, *Algorithmic Accountability: Applying the concept to different country contexts* (Washington D.C., 2017), p. 9. Available at http://webfoundation.org/docs/2017/07/Algorithms_Report_WF.pdf.

253 Ibid.

254 Sidney Fussell, "AI Professor Details Real-World Dangers of Algorithm Bias [Corrected]", *Gizmodo*, 8 December 2017. Available at https://gizmodo.com/microsoft-researcher-details-real-world-dangers-of-algo-1821129334.

255 Ibid.

Algorithmic accountability has been proposed to address algorithmic bias and prevent the harms they create. Accountability is not just transparency but includes "an obligation to report and justify algorithmic decision-making, and to mitigate any negative social impacts or potential harms".[256]

In 2016, the Fairness, Accountability and Transparency in Machine Learning community—a group of computer scientists, developers and researchers—released five guiding principles for accountable algorithms. The principles (see Box 33) aims to help developers design and implement algorithmic systems in publicly accountable ways.

Box 34. The Principles for Accountable Algorithms: Fairness, Accountability and Transparency in Machine Learning

1. **Fairness** – Ensure that algorithmic decisions do not create discriminatory or unjust impacts when comparing across different demographics.
2. **Explainability** – Ensure that algorithmic decisions as well as any data driving those decisions can be explained to end users and other stakeholders in non-technical terms.
3. **Auditability** – Enable interested third parties to probe, understand and review the behaviour of the algorithm through disclosure of information that enables monitoring, checking or criticism, including through provision of detailed documentation, technically-suitable application programming interfaces (APIs) and permissive terms of use.
4. **Responsibility** – Make available externally-visible avenues of redress for adverse individual or societal effects of an algorithmic decision system, and designate an internal role for the person who is responsible for the timely remedy of such issues.
5. **Accuracy** – Identify, log and articulate sources of error and uncertainty throughout the algorithm and its data sources so that expected and worst-case implications can be understood and inform mitigation procedures.

Extracted (with modifications) from: World Wide Web Foundation, Algorithmic Accountability: Applying the concept to different country contexts (Washington D.C., 2017), p. 11. Available at http://webfoundation.org/docs/2017/07/Algorithms_Report_WF.pdf.

Public authorities have started to take action to promote algorithmic accountability.

At the local level, the New York City Council passed an algorithmic accountability bill in December 2017.[257] This bill requires the creation of a task force that will study: (1) how city agencies use algorithms to make decisions that affect lives; (2) whether any of the systems appear to discriminate against people based on age, race, religion, gender, sexual orientation or citizenship status; and (3) explore how to make these decision-making processes understandable to the public.[258]

256 Nicholas Diakopoulos and Sorelle Friedler, "How to Hold Algorithms Accountable", *MIT Technology Review*, 17 November 2016. Available at https://www.technologyreview.com/s/602933/how-to-hold-algorithms-accountable/.

257 Elizabeth Zima, "Could New York City's AI Transparency Bill Be a Model for the Country?" *Government Technology*, 4 January 2018. Available at http://www.govtech.com/policy/Could-New-York-Citys-AI-Transparency-Bill-Be-a-Model-for-the-Country.html.

258 Lauren Kirchner, "New York City moves to create accountability for algorithms", *Ars Technica*, 19 December 2017. Available at https://arstechnica.com/tech-policy/2017/12/new-york-city-moves-to-create-accountability-for-algorithms/.

In a 2015 report entitled, *Big Data: Seizing Opportunities, Preserving Values*, the Obama administration recognized "the potential for big data technology to be used to discriminate against individuals, whether intentionally or inadvertently, potentially enabling discriminating outcomes, reducing opportunities and choices available to them".[259] The report recommended several actions to mitigate this risk including expanding technical expertise to stop discrimination and deepening understanding of differential pricing. A subsequent report, *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights* cited the need for "equal opportunity by design" safeguards to address big data's potential harms.[260]

The United Kingdom has adopted a Data Ethics Framework with the following principles:

- Start with clear user need and public benefit;
- Be aware of relevant legislation and codes of practice;
- Use data that is proportionate to the user need;
- Understand the limitations of the data;
- Ensure robust practices and work within your skill set;
- Make your work transparent and be accountable; and
- Embed data use responsibly[261].

The General Data Protection Regulation (GDPR) of the European Union (EU), which came into force on 25 May 2018, also promotes algorithmic accountability through transparency.[262]

The GDPR mandates that organizations give data subjects sufficient information about the automated systems they use for processing personal data to enable them to make informed decision as to whether to opt out from such data processing. Given that the automated systems that process citizens' personal data often rely on machine learning, a commentator noted that "machine learning processes must become transparent—if not truly transparent, then at least much less black box-like—for companies that fall under the GDPR to be able to become compliant".[263]

---

259 Executive Office of the President of the United States, "Big Data: Seizing Opportunities, Preserving Values –Interim Progress Report", February 2015. Available at https://obamawhitehouse.archives.gov/sites/default/files/docs/20150204_Big_Data_Seizing_Opportunities_Preserving_Values_Memo.pdf.

260 Executive Office of the President of the United States, "Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights", May 2016. Available at https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf.

261 United Kingdom Department for Digital, Culture, Media and Sport, "Guidance: Data Ethics Framework (updated 30 August 2018)". Available at https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework.

262 Andre Burt, "Is there a 'right to explanation' for machine learning in the GDPR?" *International Association of Privacy Professionals*, 1 June 2017. Available at https://iapp.org/news/a/is-there-a-right-to-explanation-for-machine-learning-in-the-gdpr/.

263 Juraj Jánošík, "Transparency of machine-learning algorithms is a double-edged sword", *welivesecurity,* 13 November 2017. Available at https://www.welivesecurity.com/2017/11/13/transparency-machine-learning-algorithms/.

5.3 Intuition and Decision-Making

Will the appropriate use of algorithms and analytics mean the end of intuition in decision-making?

Malcom Gladwell in his 2005 book, *Blink: The Power of Thinking Without Thinking*, defends the continuing importance of intuition. In this book, Gladwell highlights the theory of "thin slicing"—"the ability of our unconscious to find patterns in situations and behaviour based on very narrow slices of experience".[264] So, how thin is thin? Gladwell and other scholars believe that we are able to accurately draw conclusions in the emotions and attitudes of people who are interacting by observing only a few seconds of their interaction.[265]

Based on research by behavioural scientists, Gladwell argues that: (1) we all have the ability to extract an enormous amount of meaningful information from the very thinnest slice of experience; and (2) "if we are to learn to improve the quality of decisions we make [...] we need to respect the fact that it is possible to know without knowing why we know and accept that—sometimes—we're better off that way".[266]

Research has shown thin slicing's accuracy in: the first impressions of strangers with self-ratings; being able to identify sexual orientation, telephone operator's job performance, teacher ratings and trustworthiness of sales people; and judgements made between medical students and patients, interviewers and job applicants, and students and supervisors.[267]

Scholarly research on thin slicing is also changing the way we understand decision-making. Researchers found that "when we ask people to deliberate before they make a decision they tend not to be as good as they are if they do it non-consciously".[268] Further, under certain conditions more information is not helpful and can lead to bad decisions.[269]

But can policymakers rely on intuition alone?

Research has identified the following conditions when one can rely on intuition: presence of expertise, nature of the problem and time available.[270]

The more experience one has in the domain the better the odds of using intuition to make domain-specific decisions. For instance, an experienced judge can usually detect when a witness is lying. How much experience is needed? According to research, it takes 10 years of domain-specific experience to develop accurate intuitive judgments (and during those 10 years, repetition and feedback are essential).[271]

---

264 Malcom Gladwell, *Blink: The Power of Thinking Without Thinking* (New York and Boston, Blackbay Books, 2005), p. 24.

265 Jeff Thompson, "Thin Slices & First Impressions", *Psychology Today*, 24 March 2012. Available at https://www.psychologytoday.com/us/blog/beyond-words/201203/thin-slices-first-impressions.

266 Malcom Gladwell, *Blink: The Power of Thinking Without Thinking* (New York and Boston, Blackbay Books, 2005), p. 53.

267 Jeff Thompson, "Thin Slices & First Impressions", *Psychology Today*, 24 March 2012. Available at https://www.psychologytoday.com/us/blog/beyond-words/201203/thin-slices-first-impressions.

268 Rosie Ifould, "Acting on impulse", *The Guardian*, 7 March 2009. Available at https://www.theguardian.com/lifeandstyle/2009/mar/07/first-impressions-snap-decisions-impulse.

269 Malcom Gladwell, *Blink: The Power of Thinking Without Thinking* (New York and Boston, Blackbay Books, 2005), p. 140.

270 Connson Chou Locke, "When It's Safe to Rely on Intuition (and When It's Not)", *Harvard Business* Review, 30 April 2015. Available at https://hbr.org/2015/04/when-its-safe-to-rely-on-intuition-and-when-its-not.

271 Ibid.

Intuition can be used on unstructured problems—those that lack clear decision rules or has few objective criteria with which to make the decision.[272] On the other hand, it would be unwise to use intuition to solve problems that have clear decision rules, objective criteria and abundant data with which to perform an analysis.

Intuitive judgement also works well in situations that demand quick decision and there is little time for detailed analysis. Behavioural science tells us: "When information and time are scarce, using heuristics such as intuition can often be as effective as a rational approach."[273]

A key take from this discussion is that intuition still plays a role in data-driven decision-making.

However, Gary Klein, in *Sources of Power*, rejects the idea that intuition and gut feelings alone can be relied upon for decision-making.[274] He argues that real-world decision-making is a two-stage process, starting with intuition, as decision makers recognize how they need to respond, followed by deliberate evaluation, as they mentally simulate a possible response to see if it will work.[275]

Andrew McAffee and Erik Brynjolfsson go further:

> The evidence is overwhelming that, whenever the option is available, relying on data and algorithms alone usually lead to better decisions and forecasters than relying on the judgement of even experienced and "experts" humans.[276]

272 Ibid.

273 Ibid.

274 Gary Klein, *Sources of Power: How People Make Decisions – 20th Anniversary Edition* (Cambridge and London, MIT Press, 2017).

275 Ibid., p. xvii.

276 Andrew McAffee and Erik Brynjolfsson, *Machine, Platform, Crowd: Harnessing Our Digital Future* (New York and London, W. W. Norton & Company, 2017), p. 64.

6. ENABLING POLICIES

This section aims to discuss enabling policies and programmes to realize data-driven governance. They include policies and programmes that:

- Strengthen data collection, including improving national statistics system, enhancing gender statistics and spurring citizen-generated data;
- Institutionalize data governance;
- Heighten public access and confidence in data by adopting open government data and strengthening data privacy; and
- Gear up for big data.

6.1 Strengthening Data Collection

The activities to strengthen data collection include, improving national statistical systems, enhancing gender statistics and spurring citizen-generated data.

6.1.1 Improving the National Statistics System

The need to improve national statistics systems is well recognized.

Towards this, an initiative within the United Nations umbrella was launched to provide a "framework for discussion on, and planning and implementation of statistical capacity building necessary to achieve the scope and intent" of the SDGs.[277]

The *Cape Town Global Action Plan for Sustainable Development Data* aims to: "Outline the necessary actions to generate quality and timely data on a routine basis to inform sustainable development at the requested level of disaggregation and population coverage, including for the most vulnerable and hard-to-reach groups."[278]

The key principles of the Cape Town Global Action Plan are:

- **Completeness of scope** – The Plan shall address all aspects of coordination, production and use of data for sustainable development;

277 High-level Group for Partnership, Coordination and Capacity-Building for Statistics for the 2030 Agenda for Sustainable Development, "Cape Town Global Action Plan for Sustainable Development Data", adopted by the United Nations Statistical Commission at its 48th Session, March 2017, p. 2. Available at https://unstats.un.org/sdgs/hlg/Cape_Town_Global_Action_Plan_for_Sustainable_Development_Data.pdf.

278 Ibid., p. 3.

- **Accountability** – The modern production of statistics requires comprehensive interaction among data providers, producers and users; and
- **Cooperation** – The Plan recognizes the crucial role of cooperation among countries, regional organizations, and other international organizations and stakeholders in supporting countries' plans and efforts in capacity building.[279]

The Cape Town Global Action Plan identifies six strategic areas for action.

The first area is "coordination and strategic leadership on data for sustainable development".[280] The objectives are to:

- Strengthen national statistical systems and the coordination role of national statistical offices; and
- Enhance coordination among national statistical systems and regional and international organizations active in the production of data and statistics for sustainable development.

The key actions recommended include the following:

- Conduct a needs assessment of national statistical capacities and an assessment of available resources to address those needs; and
- Establish and/or improve the coordination mechanism for collecting, sharing and communicating sustainable development statistics among national statistical systems, and among national, regional and international statistical systems.

The second strategic area for improvement is "innovation and modernization of national statistical systems".[281] Specifically:

- Modernize governance and institutional frameworks to allow national statistical systems to meet the demands and opportunities of constantly evolving data ecosystems;
- Modernize statistical standards, particularly those aimed to facilitate data integration and automation of data exchange across different stages of the statistical production process; and
- Facilitate the application of new technologies and new data sources into mainstream statistical activities.

Among the proposed actions for the second area are:

- Promote the revision of statistical laws and regulatory frameworks, where necessary, consistent with the Fundamental Principles of Official Statistics, to:
  (a) enhance the status, independence and coordination role of national statistics offices;
  (b) strengthen their access to data, including enhanced data sharing across the national statistical system, and thereby their ability to more efficiently respond to emerging data and statistical needs;
  (c) develop a mechanism for the use of data from alternative and innovative sources within official statistics;
  (d) improve transparency of, and public access to, official statistics; and
  (e) strengthen the availability of sustainable funding for national statistical systems;

279 Ibid.

280 Ibid., p. 4.

281 Ibid.

- Define and implement standardized structures for the exchange and integration of data and metadata on the social, economic and environmental pillars of sustainable development and at all levels (global, regional, national and subnational), following the Statistical Data and Metadata eXchange[282] and related standards; and
- Identify specifications for interoperable, open source technologies to incorporate the flexibility in information systems needed to allow the strategic use of new and emerging technologies for official data collection, processing, dissemination and analysis.

The third strategic area is "strengthening of basic statistical activities and programmes, with particular focus on addressing the monitoring needs of the 2030 Agenda".[283] In particular:

- Strengthen and expand household survey programmes, integrated survey systems, business and other economic survey programmes, population and housing census programmes, civil and vital statistics programmes, and the International Comparison Programme taking into account the needs posed by the 2030 Agenda;
- Improve the quality of national statistical registers and expand the use of administrative records integrating them with data from surveys and other new data sources, for the compilation of integrated social, economic and environmental statistics and in relation to follow up on the 2030 Agenda;
- Strengthen and expand the System of National Accounts and the System of Environmental Economic Accounts;
- Integrate geospatial data into statistical production programmes at all levels;
- Strengthen and expand data on all groups of population to ensure that no one is left behind; and
- Strengthen and expand data on domains that are currently not well developed within the scope of official statistics.

Some of the key actions for the third strategic area are:

- Increase harmonization and ensure country ownership of internationally-sponsored household surveys programmes (such as DHS, MICS, LSMS, Child Labour Survey, WHS, CWIQ, etc.) by strengthening the existing Household Survey Network and the Intersecretariat Working Group on Household Surveys;
- Develop, standardize and improve the coverage of registers of persons, property and businesses for statistical purposes;
- Support the implementation of the System of National Accounts and the System of Environmental Economic Accounts, taking into account country experiences to date and current capacity needs to improve implementation;
- Promote the integration of modern geospatial information management systems within mainstream statistical production programmes by highlighting synergies between the two systems;
- Improve the production of high-quality, accessible, timely, reliable and disaggregated data by all characteristics relevant in national contexts to ensure that no one is left behind; and

---

282 See the official site for the SDMX community: A global initiative to improve Statistical Data and Metadata eXchange. Available at https://sdmx.org/.

283 High-level Group for Partnership, Coordination and Capacity-Building for Statistics for the 2030 Agenda for Sustainable Development, "Cape Town Global Action Plan for Sustainable Development Data", adopted by the United Nations Statistical Commission at its 48th Session, March 2017, pp. 5-6. Available at https://unstats.un.org/sdgs/hlg/Cape_Town_Global_Action_Plan_for_Sustainable_Development_Data.pdf.

- Develop, standardize and improve coverage and quality of data that today are beyond the scope of official statistics.

The fourth area for improvement is "dissemination and use of sustainable development data".[284] In particular, develop and promote innovative strategies to ensure proper dissemination and use of data for sustainable development.

Among the top recommended actions are:

- Promote the development of technological infrastructure for better data dissemination; and
- Develop effective communication and data dissemination strategies and guidelines for public and private dialogue oriented to policymakers, legislators, the media, the general public, the economy, etc.

The fifth area is "multi-stakeholder partnerships for sustainable development data".[285] Specifically, to develop and strengthen partnerships of national and international statistical systems with governments, academia, civil society, private sector and other stakeholders involved in the production and use of data for sustainable development.

The key actions are:

- Improve the transparency and accessibility of official statistics to the public; and
- Create frequent and periodic opportunities to consult with all stakeholders on the production and use of statistics for sustainable development.

The sixth area is "mobilize resources and coordinate efforts for statistical capacity building".[286] Particularly, to ensure that resources are available to implement the necessary programmes and actions as outlined in this global action plan (both domestic and from international cooperation).

The top activities proposed are:

- Provide an overview of capacity needs based on the implemented or existing needs assessments, and consider appropriate matches between types of support and types of needs;
- Identify and coordinate existing resources, including South-South and triangular cooperation mechanisms, to strategically address these needs and identify resource gaps; and
- Develop a programme for statistical capacity building on the basis of capacity needs.

Improving national statistical systems will advance data-driven governance and enhance gender statistics.

6.1.2 Enhancing Gender Statistics

The Chief Statistician of UN Women observes that while the world is producing huge amounts of data, "there are glaring blind spots and gaping holes" when it comes to data on women and girls.[287]

---

284 Ibid., p. 7.

285 Ibid.

286 Ibid.

287 UN Women, "Take five with Papa Seck: Getting better at gender data—why does it matter?" 21 September 2016. Available at http://www.unwomen.org/en/news/stories/2016/9/feature-story-take-five-with-papa-seck-on-gender-data.

Gender statistics is: "The scientific notation and interpretation of statistics that in an adequate and complete way are reflecting the living conditions and situations of women and men with respect to all policy fields and areas."[288] Specifically, they are:

- Data that are collected and presented disaggregated by sex as a primary and overall classification;
- Data that are reflecting gender issues;
- Data that are based on concepts and definitions that adequately reflect the diversity of women and men and capture all aspects of their lives; and
- Using data collection methods that take into account stereotypes and social and cultural factors that may induce gender biases (including, underreporting of women's economic activity, undercounting of girls, their births or their deaths, or underreporting of violence against women[289]).

In 2013, the United Nations Statistical Commission released the Minimum Set of Gender Indicators.[290] This set includes 52 quantitative indicators and 11 qualitative indicators organized into five domains: (1) economic structures and access to resources; (2) education; (3) health and related services; (4) public life and decision-making; and (5) human rights of women and child.

Similar to the SDG indicators, the Minimum Set of Gender Indicators are also categorized into three tiers that reflect challenges on data and methodology:

- Tier 1 – The indicator is conceptually clear, with internationally-established methodology and standards, and data regularly produced by countries with sufficient coverage to allow tracking of progress over time;
- Tier 2 – The indicator is conceptually clear, with internationally-established methodology and standards, but data is not regularly produced by countries; and
- Tier 3 – The indicator has no internationally-established methodology or standards, and data is not regularly produced by countries.

Dwelling deeper, the United Nations Economic and Social Commission for Asia and the Pacific (ESCAP) in 2015 promulgated a "core set of gender indicators for Asia and the Pacific" in addition to the global Minimum Set of Gender Indicators.[291]

The Asia-Pacific core set of gender indicators is composed of the following:

- **Six basic domains** – These are designed to provide information on region-wide issues on gender equality and women's empowerment;
- **Five supplementary domains** – These are related to issues of gender equality and women's empowerment that are considered as priorities by certain subgroups of countries in the region;
- **Priority target group (rural women)** – These are a subset of indicators in the basic domains that reflect issues

288 Gender Stats, "Gender Statistics in 5 Charts". Available at http://genderstats.org/ (accessed on 8 January 2019).

289 European Institute for Gender Equality, "Gender Statistics". Available at http://eige.europa.eu/gender-mainstreaming/methods-tools/gender-statistics (accessed on 8 January 2019).

290 United Nations Statistics Division, "Minimum Set of Gender Indicators". Available at https://genderstats.un.org/#/home (accessed on 8 January 2019).

291 ESCAP, "Core set of gender indicators for Asia and the Pacific: Note by the secretariat", Committee on Statistics, Fourth Session, 23 January 2015 (E/ESCAP/CST(4)/10). Available at http://www.asiapacificgender.org/sites/default/files/pdf/statstics_documents/Core_Set_Gender_Indicators_Asia_Pacific.pdf.

of particular concern to rural women—a generally disadvantaged population group in many countries in the Asia-Pacific region; and

- **Qualitative indicators related to national norms** – These monitor how national legislation works in ensuring gender equality through ratification of relevant international conventions, and the institutionalization of specific measures and policies to eliminate discrimination against women and promote gender equality.[292]

To ensure gender-fair data-driven governance, governments need to evaluate their data and data gathering mechanism in light of the recommendations above.

6.1.3 Spurring Citizen-Generated Data

Citizen-generated data complements official statistics. It also directly addresses data justice as citizen-produced data can make visible individuals and groups who are invisible from government due to absence of data.

An important element in spurring citizen-generated data and data-driven solutions is building community support.

There are at least six factors that are essential to building community support for data-driven solutions. These are:

1. **Organizing for ownership** – In many cases, efforts to engage affected communities take place after leaders have designed and launched data-driven initiatives. But engagement should begin earlier so that community members will have an incentive to support the initiative;
2. **Allowing for complexity** – Leaders must adapt to the complex system of influences that bear on the success of any data-driven solution;
3. **Working with local institutions** – Local organizations have built up social capital that creates an enabling environment for data-driven interventions to succeed;
4. **Applying an equity lens** – Too often, social change efforts do not engage the right mix of people;
5. **Building momentum** – The work of engaging communities requires a sense of patient urgency. Early wins will help a community build a narrative of success that can replace existing narratives that dwell on the apparent intractability of social problems. Likewise, quick wins will enable community members to see that their engagement matters; and
6. **Managing constituencies through change** – Leaders who shift to a new data-driven framework need to manage how various constituencies react to that change. A good way to start is by distinguishing between technical challenges and adaptive challenges (relating to people's behaviour and beliefs).[293]

Those committed to spurring citizen-generated data should:

- **Align interests among key stakeholders to encourage partnerships** – Successful citizen-generated data projects bring together actors with different interests in the same data. Data serves as common ground for actors and is the focal point of collaborations. There is often a difference between production, use and uptake, and the benefits associated with each stage can be different. Different actors can value different aspects of the data, and understanding how actors perceive this value is key to building multi-stakeholder partnerships.

---

292 Ibid.

293 Melody Barnes and Paul Schmitz, "Community Engagement Matters (Now More Than Ever)", *Stanford Social Innovation Review* (Spring 2016). Available at https://ssir.org/articles/entry/community_engagement_matters_now_more_than_ever#.

- **Ensure that citizen-generated data is usable in multiple ways to maximize uptake and impact** – The more ways a data set can be used, the more different types of actors will become interested in the data. To facilitate different use cases by different actors, data needs to be accessible and presented in an interoperable format.
- **Tap into existing resources and processes in order to make it easier to produce and use citizen-generated data effectively** – This includes using technology that citizens are familiar with and are using, as well as building on established routines and group dynamics, such as existing bureaucratic processes or community forums.
- **Consider the specific incentives that depend on the context and the goal** – Key dimensions to consider include whether the project aims to link up with government directly or not, and the socio-political and governance environment. This covers, amongst others, whether the government is responsive, whether there is a strong legal framework and high levels of trust, or whether there is adequate information about the issue.[294]

For government, actions to spur citizen-generated data include:

- Recognizing that citizen-generated data and civil society data can play a different and complementary role to public sector data. Government should look for ways to engage with, acknowledge and support these initiatives.
- Investigating funding and sustainability models for citizen-generated data and civil society data. In many cases, citizen-generated data should be considered complementary to institutional data collection, rather than as a pretext to stop or reduce funding for the latter.
- Supporting consultation and participation processes to enable citizen and civil society input regarding institutional data collection, including through events, responsive communication channels and participatory design processes.
- Supporting further research and the development of resources in this area that can be used to make public data infrastructures more responsive to the interests and concerns of civil society.[295]

6.2 Institutionalizing Data Governance

To recall, data governance is: "A quality control discipline for adding new rigour and discipline to the process of managing, using, improving and protecting organizational information."[296] It sets the "parameters for data management and usage, creating processes for resolving data issues, and enabling business users to make decisions based on high-quality data and well-managed information assets".[297]

---

294 Danny Lammerhirt, Shazade Jaeson and Eko Presetyo, "Making Citizen Generated Data Work: Towards a Framework Strengthening Collaborations Between Citizens, Civil Society Organisations, and Others", *Data Shift*, March 2017. Available at http://civicus.org/thedatashift/wp-content/uploads/2017/03/Making-Citizen-Generated-Data-Work_short-report_.pdf.

295 Jonathan Gray, Danny Lammerhirt and Liliana Bounegru, "Changing What Counts: How Can Citizen-Generated and Civil Society Data be Used as an Advocacy Tool to Change Official Data Collection?" *Data Shift*, March 2016. Available at http://civicus.org/thedatashift/wp-content/uploads/2016/03/changing-what-counts-2.pdf.

296 IBM, "The IBM Data Governance Council Maturity Model: Building a roadmap for effective data governance", October 2007. Available at https://www-935.ibm.com/services/uk/cio/pdf/leverage_wp_data_gov_council_maturity_model.pdf.

297 TechTarget, "Building an effective data governance framework", October 2013. Available at https://searchdatamanagement.techtarget.com/essentialguide/Building-an-effective-data-governance-framework.

Without good data governance, organizations cannot be more efficient and effective, and cannot become more transparent.[298]

To institutionalize data governance, governments must first understand its enablers:

- **Strategic intent** – WHY data governance is of value, the end state that government is trying to reach, and the foundational policies that describe the motivation of executive leadership;
- **Data governance maturity model** – The journey from the AS IS to the SHOULD BE regarding the management of data, information and knowledge assets. Data governance maturity models provide the means for gauging progress. By presenting intermediate milestones, as well as the desired end state, maturity models assist in planning HOW government will reach the next level of effectiveness, as well as WHEN and WHERE within government;
- **Organizational models, roles and responsibility matrices** – WHO should be involved in decision-making, implementing, monitoring and sustaining;
- **Framework** – WHAT is governed including related concepts, components and the interrelationships among them;
- **Methodology for navigating the framework** – The methods and procedures for HOW to navigate the framework, create the artefacts that describe the enterprise, and sustain the effort over time;
- **Performance metrics** – Measure and evaluate progress and efficacy of the initiative. These are traceable back to strategic intent and related maturity models. These metrics need to be continually evaluated for relevancy; and
- **Valuation and security of government information assets** – Proper valuation of data and information will determine the level of investment to ensure quality and appropriate security throughout the information asset life cycle.[299]

To succeed, data governance policies and procedures must overcome the following barriers:

- Determining the rules and requirements, including interpreting and understanding the rules concerning data sources;
- Gaining agreement of all parties on policies;
- Developing new tools and software to enable data governance;
- Cost of implementing policies;
- Incompatible systems;
- Competing priorities within the organization;
- Getting management to understand what is necessary; and
- Building the project process.[300]

---

298 Colin Wood, "Data Governance: The Public Sector's Next Big Frontier", *Government Technology*, 29 April 2014. Available at http://www.govtech.com/data/Data-Governance.html.

299 NASCIO, "Data Governance Part II: Maturity Models – A Path to Progress", March 2009, pp. 2-3. Available at https://nascio.org/Portals/0/Publications/Documents/NASCIO-DataGovernancePTII.pdf.

300 NASCIO, "Data Governance – Managing Information as an Enterprise Asset: Part I – An Introduction", April 2008, p. 5. Available at https://www.nascio.org/Portals/0/Publications/Documents/NASCIO-DataGovernance-Part1.pdf.

The next step is to develop a data governance strategy. This includes:

- **Charter** – Identify the drivers, vision, mission and principles for data governance, including readiness assessment, internal process discovery, and current issues or success criteria;
- **Operating framework and accountabilities** – Define structures and responsibilities for data governance activities;
- **Implementation road map** – Determine time frames for the roll out of policies and directives, business glossary, architecture, asset valuation, standards and procedures, expected changes to business and technology processes, and deliverables to support auditing activities and regulatory compliance; and
- **Plan for operational success** – Describe a target state of sustainable data governance activities.[301]

Finally, governments should consider adopting data governance principles, such as those developed by the Data Governance Institute, as follows:

- **Integrity** – Data governance participants will practice integrity with their dealings with each other; they will be truthful and forthcoming when discussing drivers, constraints, options and impacts for data-related decisions;
- **Transparency** – Data governance and stewardship processes will exhibit transparency; it should be clear to all participants and auditors how and when data-related decisions and controls were introduced into the processes;
- **Auditability** – Data-related decisions, processes and controls subject to data governance will be auditable; they will be accompanied by documentation to support compliance-based and operational auditing requirements;
- **Accountability** – Data governance will define accountabilities for cross-functional data-related decisions, processes and controls;
- **Stewardship** – Data governance will define accountabilities for stewardship activities that are the responsibilities of individual contributors, as well as accountabilities for groups of data stewards;
- **Checks and balances** – Data governance will define accountabilities in a manner that introduces checks-and-balances between business and technology teams, as well as between those who create/collect data, those who manage it, those who use it, and those who introduce standards and compliance requirements;
- **Standardization** – Data governance will introduce and support standardization of enterprise data; and
- **Change management** – Data governance will support proactive and reactive change management activities for reference data values and the structure/use of master data and metadata.[302]

---

301 DAMA International, *Data Management Body of Knowledge*, second edition (DAMA-DMBOK 2) (New Jersey, Technics Publications, 2017), p. 82.

302 The Data Governance Institute, "Goals and Principles for Data Governance". Available at http://www.datagovernance.com/adg_data_governance_goals/ (accessed on 8 January 2019).

6.3 Heightening Public Access and Confidence in Data

The activities to consider in order to heighten public access and confidence in data include opening government data and strengthening data privacy.

6.3.1 Opening Government Data

Open government data is defined as: "Data produced or commissioned by government or government-controlled entities that can be freely used, reused and redistributed by anyone."[303] It is also: "A philosophy—and increasingly a set of policies—that promotes transparency, accountability and value creation by making government data available to all."[304]

The eight principles of open government data are:

1. **Complete** – All public data is made available. Public data is data that is not subject to valid privacy, security or privilege limitations;
2. **Primary** – Data is collected at the source, with the highest possible level of granularity, not in aggregate or modified forms;
3. **Timely** – Data is made available as quickly as necessary to preserve the value of the data;
4. **Accessible** – Data is available to the widest range of users for the widest range of purposes. Data must be made available on the Internet so as to accommodate the widest practical range of users and uses;
5. **Machine processable** – Data is reasonably structured to allow automated processing;
6. **Non-discriminatory** – Data is available to anyone, with no requirement of registration. Anonymous access to the data must be allowed for public data, including access through anonymous proxies. Data should not be hidden behind "walled gardens";
7. **Non-proprietary** – Data is available in a format over which no entity has exclusive control. Proprietary formats add unnecessary restrictions over who can use the data, how it can be used and shared, and whether the data will be usable in the future; and
8. **License free** – Data is not subject to any copyright, patent, trademark or trade secret regulation. Reasonable privacy, security and privilege restrictions may be allowed.[305]

Open data promotes: transparency and democratic control; participation; self-empowerment; improved or new products and services; innovation; more efficient and effective government services; impact measurement of policies; and new knowledge from combined data sources and patterns in large data volumes.[306]

---

303 Open Knowledge Foundation, "What is Open Government Data". Available at https://opengovernmentdata.org/ (accessed on 8 January 2019).

304 OECD, "Open Government Data". Available at http://www.oecd.org/gov/digital-government/open-government-data.htm (accessed on 8 January 2019).

305 Joshua Tauberer, "The Annotated 8 Principles of Open Government Data". Available at https://opengovdata.org/ (accessed on 8 January 2019).

306 Open Knowledge International, "Why Open Data?" *Open Data Handbook*. Available at http://opendatahandbook.org/guide/en/why-open-data/ (accessed on 8 January 2019).

The three building blocks for open data to achieve its promise are: (1) the publication of open data by governments; (2) the conversion of data to actionable information by intermediaries; and (3) the use of data by citizens, government officials and other stakeholders to achieve development outcomes.[307]

The Sunlight Foundation provides governments a useful guide to developing and implementing an open data policy.[308]

The first part of the guideline deals with what data should be public. The recommendations are to:

- **Proactively release government data online** – Proactive disclosure is the release of public data before an individual requests it. In the 21st century, it means proactively putting new data online.
- **Reference and build on existing public accountability and access policies** – A strong open data policy builds upon the principles embodied by existing laws and policies that defend and establish public access, often defining standards for data quality, disclosure and publishing.
- **Build on the values, goals and mission of the community and government** – An open data policy can be pursued with the intent of realizing many different varieties of public good, including greater government transparency, honesty, accountability, efficiency, civic engagement and economic growth. An explicit statement of goals, values or intention can help clarify the outcomes that a government hopes an open data policy will help achieve.
- **Create a public and comprehensive list of all data holdings** – Governments should conduct an inventory of existing data early in the process of open data policy development in order for the government and other stakeholders to be aware of the full potential dimensions of data release.
- **Specify methods of determining the prioritization of data release** – While an open data policy ideally enables the online release of all public government data, the release of data may end up being a staggered process for practical reasons, such as insufficient funding or staffing. Governments should be clear about the range of potential methods that could be used in determining the priority order of data release.
- **Stipulate that open data applies to contractors and quasi-governmental agencies** – Data that is gathered from the public using public funds should remain publicly-accessible, regardless of government decisions to delegate its management.
- **Appropriately safeguard sensitive data** – A well-crafted open data policy is complementary to pre-existing legislation and directives about access to public data, which means that it can integrate pre-existing public access law exemptions for data that is sensitive for privacy, security or other reasons.[309]

The second part is on how to make data public. The recommendations are to:

- **Mandate data formats for maximal technical access** – Data must be released in formats that lend themselves to easy and efficient reuse via technology. This means releasing data in open formats (or open standards) and in machine-readable formats that are appropriately structured (or machine-processable).

---

307 Andreas Pawelke and others, *Data for development: What's next? – Concepts, trends and recommendations for German development cooperation* (Bonn and Eschborn, Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH, 2017), p. 29. Available at http://webfoundation.org/docs/2017/12/Final_Data-for-development_Whats-next_Studie_EN-1.pdf.

308 Sunlight Foundation, "Open Data Policy Guidelines", *Open Data Policy Hub*. Available at https://opendatapolicyhub.sunlightfoundation.com/guidelines/ (accessed on 8 January 2019).

309 Ibid.

- **Provide comprehensive and appropriate formats for varied uses** – Appropriate methods of distribution should be considered, to maximize the degree of access, use and quality of published data.
- **Remove restrictions for accessing data** – To provide truly open access, there must be the right to reuse government data with no technical restrictions such as registration requirements, access fees and usage limitations, among others.
- **Mandate data be explicitly license-free** – If data is to be truly public and maximally reusable, there should be no license-related barrier to the reuse of public data.
- **Charge data-creating agencies with recommending an appropriate citation form** – The practice of citing government data can be encouraged by having direct data managers develop model citations for their data sets. These model citations should list key elements of the source's identity that would be required to effectively identify an individual data source and identify the unit of government that created or maintains the data.
- **Require publishing of metadata** – Providing a common and fully-described core metadata scheme (as well as other documentation) can be useful for the public and government alike. A strong metadata scheme takes its lead from common international meta-attributes (such as DCAT), and allows data publishers to classify contextual fields or elements within their data sets.
- **Require publishing of data creation processes** – A summary of the processes that were used to create a specific data set provides valuable context that may not be discernible via metadata alone and should accompany the data set's release.
- **Mandate the use of unique identifiers** – The use of unique identifiers within and across data sets improves the quality and accuracy of data analysis. Without unique identifiers, some analyses can become difficult or impossible, since similar names may or may not refer to the same entities. Importantly, identifiers should be non-proprietary and public.
- **Require code sharing or publishing as open source** – In addition to data, the code used to create government websites, portals, tools and other online resources can provide benefits as valuable open data itself.
- **Require digitization and distribution of archival materials** – Questions about what archival material should be digitized and what timelines are realistic for digitizing archival material can be informed by the same kind of prioritization process used for general data release.
- **Create a central location devoted to data publication and policies** – Data portals and similar websites can facilitate the distribution of open data by providing an easy-to-access, searchable hub for multiple data sets.
- **Publish bulk data** – Bulk access provides a simple but effective means of publishing data sets in full by enabling the public to download all of the data stored in a database at once.
- **Create public APIs for accessing data** – Government bodies can develop APIs that allow third parties to automatically search, retrieve or submit data directly from databases online.
- **Optimize methods of data collection** – To optimize data quality and timeliness, disclosure regulations should take advantage of online data-collection methods. Electronic filing, also known as "e-filing", is one method of optimizing the quality and timeliness of data collection.

- **Mandate ongoing data publication and updates** – The ideal of online data is real-time access—data should be made available as close as possible to the time that it is collected. It is not enough to mandate the one-time release of a data set, because it becomes incomplete as soon as additional data is created but not published.
- **Create permanent, lasting access to data** – Once released, digitized government data should remain permanently available, "findable" at a stable online location or through archives in perpetuity.[310]

The third part is how to implement an open data policy. The recommendations are to:

- **Create or appoint an oversight authority** – Some questions may defy easy treatment in the process of creating an open data policy, so it is appropriate to define a single authority empowered to resolve conflicts and ensure compliance with new open data measures.
- **Create guidance or other binding regulations for implementation** – An open data policy should be practically aspirational, meaning that it should define a vision for why the policy is being implemented, but also be able to provide actionable steps for the government and oversight authorities to follow to see the policy through to implementation. Creating regulations or guidance can ensure a strong, reliable policy and usually mean the difference between policy passed for show versus policy passed for substance.
- **Incorporate public perspectives into policy implementation** – The public should be involved in the ongoing assessment and review of the policy's implementation. Governments should create meaningful opportunities for public feedback about data quality, quantity, selection and format, as well as the user-friendliness of the point of access.
- **Set appropriately ambitious timelines for implementation** – Setting clear deadlines can demonstrate the strength of a commitment and will help translate commitments into results. Deadlines can also help identify failures clearly, opening the door to public oversight.
- **Create processes to ensure data quality** – Data quality will not be ensured through data release alone. Efforts need to be made to keep the data up-to-date, accurate and accessible.
- **Ensure sufficient funding for implementation** – Like any other initiative, implementing an open data policy should be done with an eye on long-term sustainability. One way to do this is to consider funding sources for the implementation of the policy, as well as its future maintenance. Sufficient funding can mean the difference between successful and unsuccessful policies.
- **Create or explore potential partnerships** – Partnerships can be useful in a variety of important efforts related to data release, such as: increasing the availability of open data, identifying constituent priorities for data release, and connecting government data to that held by non-profit organizations, think tanks, academic institutions and nearby governments.
- **Mandate future review for potential changes to this policy** – In order to keep up with current best practices and feedback from existing policy oversight, the open data policy should mandate future review of the policy itself, as well as of any guidance created by the policy or other implementation processes.[311]

The Open Data Barometer recommends the following key steps "to ensure the 'data revolution' will lead to a genuine revolution in the transparency and performance of governments":

---

310 Ibid.

311 Ibid.

- High-level political commitment to proactive disclosure of public sector data, particularly the data most critical to accountability;
- Sustained investment in supporting and training a broad cross-section of civil society and entrepreneurs to understand and use data effectively;
- Contextualizing open data tools and approaches to local needs, for example by making data visually accessible in countries with lower literacy levels;
- Support for city-level open data initiatives as a complement to national-level programmes; and
- Legal reform to ensure that guarantees of the right to information and the right to privacy underpin open data initiatives.[312]

Despite public, high-level commitment by governments to open data and a number of multi-stakeholder global initiatives, open data adoption has been sluggish. The *2017 Open Data Inventory*—an assessment of the coverage and openness of official statistics published on the websites of national statistical offices—reported that "progress to date has been slow".[313]

6.3.2 Strengthening Data Privacy

Surveillance—or "purposeful, routine, systematic and focused attention paid to personal details, for the sake of control, entitlement, management, influence or protection"—has become the norm.[314]

An August 2018 study revealed that unknown to smartphone owners, Google is tracking Android device location even when the phone is stationary. In a 24-hour period, an Android device sends about 4.4MB to Google while an iPhone sends 0.76MB. The same study showed that "iPhones send data 10 times less frequently to Apple's servers than the Android device to Google's servers. Apple is also collecting location data just once per day on average".[315]

In the public sector, "law enforcement agencies and other government entities have found Facebook, Twitter, Instagram and other social media sites to be rich sources of data to mine for a variety of purposes".[316]

In this age, the data privacy challenge is: "To determine when and how it is ethically responsible to analyse the information, what to look for in the data, which questions to ask of the data, and the scale to which it is reasonable to make predictions about future events and actions based on that data."[317]

---

312 Web Foundation, "Key Findings", *Open Data Barometer*. Available at https://opendatabarometer.org/2ndEdition/summary/index.html (accessed on 8 January 2019).

313 Open Data Watch, "Open Data Inventory 2017 Annual Report: A Progress Report on Open Data", 2018. Available at https://opendatawatch.com/publications/open-data-inventory-2017-annual-report/.

314 David Lyon, "Surveillance Society", presentation made at Piacenza, Italy, 28 September 2008. Available at http://www.festivaldeldiritto.it/2008/pdf/interventi/david_lyon.pdf.

315 Liam Tung, "Want Google to track you less? Get an iPhone, ditch the Android", *ZDNet*, 23 August 2018. Available at https://www.zdnet.com/article/want-google-to-track-you-less-get-an-iphone-ditch-the-android/.

316 Rachel Levinson-Waldman, "Government Access to and Manipulation of Social Media: Legal and Policy Challenges", *Howard Law Journal*, vol. 61, no. 3 (August 2018), p. 560. Available at https://www.brennancenter.org/analysis/government-monitoring-social-media-legal-and-policy-challenges.

317 Jens-Erik Mai, "Big data privacy: The datafication of personal information", *The Information Society*, vol. 32, no. 3 (2016), p. 194. Available at http://download.xuebalib.com/xuebalib.com.22693.pdf.

Since the 1970s, data privacy legislation has been the traditional response to growing surveillance. Data privacy (or information privacy) "is the right to have some control over how your personal information is collected and used".[318]

The first privacy legislation—The Data Protection Act—came into force in the German state of Hesse in 1971.[319] Sweden's Data Act of 1973 is considered the first national data protection law. The United States Congress passed the Privacy Act of 1974, which regulates the collection and use of records by federal agencies, and gives individuals right to access and correct their personal information.[320] France, a pioneer in privacy legislation, passed its Data Protection Act in 1978 (which was amended in 2005).[321]

In Asia, Japan formulated its Act on the Protection of Computer Processed Data Held by Administrative Organs in 1988.[322] Taiwan passed a Computer Processed Personal Data Protection Act in 1995. The Republic of Korea's first data protection law (Public Agency Data Protection Act) was passed in 1995. The same year, Hong Kong's colonial government enacted the Personal Data (Privacy) Ordinance, which is considered as Asia's first comprehensive data privacy law. Thailand's Official Information Act was passed in 1997. The Philippines enacted its Data Privacy Act in 2012.

At the regional level, the Organisation for Economic Co-operation and Development (OECD) released its data protection principles in 1980, which are as follows:

- **Collection limitation principle** – There should be limits to the collection of personal data and any such data should be obtained by lawful and fair means and, where appropriate, with the knowledge or consent of the data subject.
- **Data quality principle** – Personal data should be relevant to the purposes for which they are to be used, and, to the extent necessary for those purposes, should be accurate, complete and kept up-to-date.
- **Purpose specification principle** – The purposes for which personal data is collected should be specified not later than at the time of data collection and the subsequent use limited to the fulfillment of those purposes or such others as are not incompatible with those purposes and as are specified on each occasion of change of purpose.
- **Use limitation principle** – Personal data should not be disclosed, made available or otherwise used for purposes other than those specified, except with the consent of the data subject or by the authority of law.
- **Security safeguards principle** – Personal data should be protected by reasonable security safeguards against such risks as loss or unauthorized access, destruction, use, modification or disclosure of data.
- **Openness principle** – There should be a general policy of openness about developments, practices and policies with respect to personal data. Means should be readily available of establishing the existence and nature of personal data, and the main purposes of their use, as well as the identity and usual residence of the data controller.

---

318 International Association of Privacy Professionals, "What does privacy mean?" Available at https://iapp.org/about/what-is-privacy/ (accessed on 8 January 2019).

319 Jan Holvast, "History of Privacy", in *The Future of Identity*, V. Matyáš and others, eds. (International Federation for Information Processing, 2009), p. 28. Available at https://link.springer.com/content/pdf/10.1007/978-3-642-03315-5_2.pdf.

320 Daniel J. Solove, "A Brief History of Information Privacy Law", *GW Law*, 2006. Available at https://scholarship.law.gwu.edu/cgi/viewcontent.cgi?article=2076&context=faculty_publications.

321 Deutsche Welle, "France maintains long tradition of data protection", 26 January 2011. Available at http://www.dw.com/en/france-maintains-long-tradition-of-data-protection/a-14797711.

322 Graham Greenleaf, "A Brief History of Data Privacy Laws in Asia", *Tiki-Toki*. Available at http://www.tiki-toki.com/timeline/entry/381411/A-Brief-History-of-Data-Privacy-Laws-in-Asia/ (accessed on 8 January 2019). Subsequent discussion on Asian data privacy laws are based on this website.

- **Individual participation principle** – An individual should have the right to:
  a. Obtain from a data controller, or otherwise, confirmation of whether or not the data controller has data relating to him/her;
  b. Have communicated to him/her, data relating to him/her: (i) within a reasonable time; (ii) at a charge, if any, that is not excessive; (iii) in a reasonable manner; and (iv) in a form that is readily intelligible to him/her;
  c. Be given reasons if a request is denied, and be able to challenge such denial; and
  d. Challenge data relating to him/her and, if the challenge is successful, have the data erased, rectified, completed or amended.
- **Accountability principle** – A data controller should be accountable for complying with measures that give effect to the principles stated above.[323]

This document, and its principles, had influenced the development of national legislations and model codes of not only the OECD but other countries as well.

In 2005, the Asia-Pacific Economic Cooperation (APEC) adopted the APEC Privacy Framework, which "is consistent with the core values of the *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*, and reaffirms the value of privacy to individuals and to the information society".[324] A decade later, APEC issued an updated framework that draws on the concepts introduced in the *OECD Privacy Guidelines* of 2013,[325] with due consideration for the different legal features and context of the APEC region.[326]

The most important recent development in data privacy is the coming into effect of the GDPR—a new regulatory framework that unifies data protection laws across the EU.

The GDPR has seven key principles:

1. Lawfulness, fairness and transparency;
2. Purpose limitation;
3. Data minimization;
4. Accuracy;
5. Storage limitation;
6. Integrity and confidentiality (security); and
7. Accountability.[327]

323 Ben Gerber, "OECD Privacy Principles", 9 August 2010. Available at http://oecdprivacy.org/.

324 APEC, *APEC Privacy Framework* (Singapore, 2015). Available at https://www.apec.org/Publications/2017/08/APEC-Privacy-Framework-(2015).

325 OECD, *OECD Privacy Guidelines* (Paris, 2013). Available at http://www.oecd.org/internet/ieconomy/privacy-guidelines.htm.

326 APEC, *APEC Privacy Framework* (Singapore, 2015). Available at https://www.apec.org/Publications/2017/08/APEC-Privacy-Framework-(2015).

327 Information Commissioner's Office, United Kingdom, "The Principles". Available at https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/principles/ (accessed on 8 January 2019).

Among its key features that are salient to organizations are:

- Appointment of a data protection officer;
- Mandatory conduct of privacy impact assessments;
- Enhanced data subjects rights;
- Data breach notification within 72 hours; and
- New compliance obligations.[328]

For non-EU states, GDPR is relevant because of its extraterritorial applicability. The GDPR applies to:

- All companies processing the personal data of data subjects residing in the EU, regardless of the company's location; and
- The processing of personal data of data subjects in the EU by a controller or processor not established in the EU, where the activities relate to—offering goods or services to EU citizens (irrespective of whether payment is required), and the monitoring of behaviour that takes place within the EU.[329]

Furthermore, non-EU businesses processing the data of EU citizens will have to appoint a representative in the EU.[330]

A recent and interesting development is the use of ethics in data privacy protection. Pioneered by the EU Data Privacy Supervisor and followed by Hong Kong and the Philippines, ethics is promoted as an important complement to legal obligations to achieve privacy goals.[331] Further, ethics is seen as a guide to behaviour in an environment where law is unable to keep up with technological development. In this approach, compliance with the law is simply the floor (baseline) for data protection initiatives.

But do data privacy legislations and ethics-based regulation adequately protect us in the age of big data?

Nobel laureate Joseph Stiglitz believes that there is a need for regulations on: "What data the tech firms can store; what data they can use; whether they can merge different data sets; the purposes for which they can use that data; and what degree of transparency they must provide about what they do with the data."[332]

---

328 Aditya Vats, "10 Key Issues Of General Data Protection Regulation (GDPR)", *Medium*, 25 May 2017. Available at https://medium.com/@adityavats/10-key-issues-of-general-data-protection-regulation-gdpr-d70e3875b59e.

329 EUGDPR.org, "GDPR Key Changes". Available at https://www.eugdpr.org/key-changes.html (accessed on 8 January 2019).

330 Ibid.

331 European Data Privacy Supervisor, "Towards a new digital ethics: Data, dignity and technology", Opinion 4, 2015, p. 6. Available at https://edps.europa.eu/sites/edp/files/publication/15-09-11_data_ethics_en.pdf. For Hong Kong, see Information Accountability Foundation, "Ethical Accountability Framework for Hong Kong, China: A Report prepared for the Office of the Privacy Commissioner for Personal Data", n.d. Available at https://www.pcpd.org.hk/misc/files/Ethical_Accountability_Framework.pdf. For the Philippines, see National Privacy Commission, Philippines, "NPC launches DPO ACE Program, sets benchmark for data privacy training in PH", 12 December 2018. Available at https://www.privacy.gov.ph/2018/12/npc-launches-dpo-ace-program-sets-benchmark-for-data-privacy-training-in-ph/.

332 Ian Sample, "Joseph Stiglitz on artificial intelligence: 'We're going towards a more divided society'", *The Guardian*, 8 September 2018. Available at https://www.theguardian.com/technology/2018/sep/08/joseph-stiglitz-on-artificial-intelligence-were-going-towards-a-more-divided-society.

Jens-Erik Mai argues that existing privacy models (and by extension laws) "share a common focus on the collection of data and ... are concerned with how and for what purposes the collected data are later used".[333] He believes that in the era of datafication and big data we need to shift the focus from data collection to data processing and analysis. This is because "the privacy concerns at play are the construction of new knowledge, insights or realities based on the available data".[334] Mai suggests that we need to combine the existing privacy approaches with the "datafication" model of privacy that "focuses on the anonymous creation of new personal information, the reinterpretation and statistical analysis of data, and the commoditized nature of personal information".

How this is reflected in legislation or regulation is something that still needs fleshing out.

6.4 Gearing Up for Big Data

The potential of big data in providing timely and relevant statistics to help resolve development challenges is well recognized.[335] Yet, a number of issues related to methodology and technology, legislation, privacy, management, and finance remain to be resolved. There is also the issue of creating "an environment where the public trust in the use of big data for official statistics is established, and where privacy and confidentiality of personal information can be assured".[336]

Creating an enabling environment for big data requires leadership.

Ben Rossi believes that "government leaders must commit themselves to leading (big data initiatives) at all levels and across all domains".[337] Rossi adds that to lead effectively, leaders need to have a "fundamental mastery" of data, analytics and algorithms. Competency can influence whether big data initiatives make citizens' lives better or these same citizens will be victimized by data, analytics and algorithms.

Aside from leadership, the British government highlights two main challenges to overcome in order to have a successful big data initiative: (1) winning and retaining public trust; and (2) building civil service capacity in collecting, storing, analysing, sharing and using data.[338]

---

333 Jens-Erik Mai, "Big data privacy: The datafication of personal information", *The Information Society*, vol. 32, no. 3 (2016), p. 198. Available at http://download.xuebalib.com/xuebalib.com.22693.pdf.

334 Ibid.

335 United Nations Global Working Group for Big Data, "Using Big Data for the Sustainable Development Goals". Available at https://unstats.un.org/bigdata/taskteams/sdgs/ (accessed on 8 January 2019).

336 Ibid.

337 Ben Rossi, "Why the Governments' Data Science Ethical Framework is a Recipe for Disaster", *Information Age*, 2 June 2016. Available at https://www.information-age.com/why-governments-data-science-ethical-framework-recipe-disaster-123461541/.

338 John Manzoni, "Big data in government: The challenges and opportunities", *Gov.UK*, 21 February 2017. Available at https://www.gov.uk/government/speeches/big-data-in-government-the-challenges-and-opportunities.

For the British government: "Public trust is absolutely critical to achieving our ambition for a data-driven government."[339] To earn public trust means: (1) personal data collected and held by an agency is used appropriately and effectively; and (2) personal data is secure, particularly when it is being shared by different agencies. A specific measure implemented to gain public trust is the Office for National Statistics' framework called the "Five Safes" for building and maintaining trust and confidence in the use of government data.[340]

As for the second challenge—building capacity—the British government has launched programmes aimed at developing a specialist data science community, as well as enhancing data literacy for non-data specialists in the civil service. Towards this, a Data Science Campus within the United Kingdom Office for National Statistics was created. The goal of the Campus is: "To build a new generation of tools and technologies to exploit the growth and availability of innovative data sources, and to provide rich informed measurement and analyses on the economy, the global environment and wider society".[341] Created in 2017, the Data Science Campus has a target of producing "500 qualified data analysts for government over the coming years".[342] Among its programmes to accomplish this are: apprenticeships in data analytics; MSc in data analytics for government; continuous professional development modules; co-funded and/or co-supervised MSc and PhD placements and programmes; support and funding for MSc students; data science seminar series; the "Art of the Possible" training; data science accelerator and Office for National Statistics' Data Science Academy mentoring; and STEM Ambassadors.[343]

There are also operational issues that governments seeking to harness the power of big data must address.

Researchers at the University of Manchester developed a "design-reality gap model for big data for development". To understand the barriers to using big data in development, the model measures the "gap between the design requirements or assumptions of big data vs. the current reality on the ground".[344] It examines seven dimensions:

1. **Information** – Includes both information and data, as the precursor to information;
2. **Technology** – Mainly focuses on the ICTs that handle data;
3. **Processes** – The activities undertaken in generating, capturing, analysing, presenting and using data;
4. **Objectives and values** – The "objectives" component covers issues of self-interest and politics, and informal and formal strategies; and the "values" component covers culture (e.g., what stakeholders feel are the right and wrong ways to do things);
5. **Skills and knowledge** – Covers both qualitative and quantitative aspects of human competencies for undertaking data-related processes;

339 Ibid.

340 Peter Stokes, "The 'Five Safes' – Data Privacy at ONS", *Office for National Statistics*, United Kingdom, 27 January 2017. Available at https://blog.ons.gov.uk/2017/01/27/the-five-safes-data-privacy-at-ons/. This is an initiative to protect private information collected by government, which may run counter to an open data policy.

341 Office for National Statistics, United Kingdom, "Data Science Campus: Data science for public good". Available at https://www.ons.gov.uk/aboutus/whatwedo/datasciencecampus (accessed on 8 January 2019).

342 Data Science Campus, "Building Capability". Available at https://datasciencecampus.ons.gov.uk/capability (accessed on 8 January 2019).

343 Ibid. STEM stands for science, technology, engineering and mathematics.

344 Richard Heeks, "Measuring Barriers to Big Data for Development", *ICTs for Development*, 9 August 2016. Available at https://ict4dblog.wordpress.com/2016/08/09/measuring-barriers-to-big-data-for-development/.

6. **Management systems and structures** – The broader management systems required to organize within and between data-related organizations and networks, plus the way in which those systems are structured, both formally and informally; and
7. **Other resources** – Time and money.

Using the model to analyse the barriers to using big data for development in Colombia, the University of Manchester researchers noted that serious gaps exist in all dimensions.

These researchers generated recommendations from Colombia that could be useful for other developing countries seeking to launch big-data-for-development projects, and they include the following:

1. **Information** – Make greater use of existing big data sets, such as those available via social media and mobile call detail records, and "dark data" already within the public sector;
2. **Technology** – Make incremental investments in systems to digitally capture data from existing sources;
3. **Processes** – Prioritize enactment of value-generation rather than data-generation processes;
4. **Objectives and values** – Enact or amend legislation to deal specifically with big data issues such as data privacy, intellectual property rights and commercialization;
5. **Skills and knowledge** – Incorporate big data into curricula for public administration training and degree programmes;
6. **Management systems and structures** – Appoint big data champions and task forces in each main public agency, and incorporate champions into a proposed centre of excellence to enable cross-agency sharing of good practice; and
7. **Other resources** – Develop public-private partnership funding for big data initiatives.

Also useful is UNDP and United Nations Global Pulse's *A Guide to Data Innovation for Development: From Idea to Proof-Of-Concept*, which provides step-by-step guidance on leveraging new sources of data for development.[345]

345 UNDP and United Nations Global Pulse, *A Guide to Data Innovation for Development: From Idea to Proof-Of-Concept*, version 1 (New York, 2016). Available at http://www.undp.org/content/undp/en/home/librarypage/development-impact/a-guide-to-data-innovation-for-development---from-idea-to-proof-.html.

7. CREATING A DATA CULTURE IN THE PUBLIC SECTOR

This section aims to discuss how to create a data culture in the public service.

Data culture, according to Elizabeth Dunlea, means "using data in a pervasive way" in an organization.[346] In the public sector, data culture means "a deep, organization-wide comfort level with using metrics to maximize social impact".[347]

The six principles that underpin a healthy data culture are:

1. **Data culture is decision culture** – The fundamental objective in collecting, analysing and deploying data is to make better decisions.
2. **Data culture and leadership** – Commitment from the very top decision makers is essential, but that commitment must be manifested by more than occasional high-level pronouncements; there must be an ongoing, informed conversation with top decision makers and those who lead data initiatives throughout the organization.
3. **The democratization of data** – Get data in front of people and they get excited, but building cool experiments or imposing tools top-down are not sufficient. To create a competitive advantage, stimulate demand for data from the grass roots.
4. **Data culture and risk** – An effective data culture puts risk at its core. Although companies must identify their limits and honour them, risk management should operate as a smart accelerator, by introducing analytics into key processes and interactions in a responsible manner.
5. **Culture catalysts** – To really ensure buy-in, someone has to take charge and make a strong effort to bring about change. It requires people who can bridge both worlds—data science and on-the-ground operations. And, usually, the most effective change agents are not digital natives.
6. **Marrying talent and culture** – This involves striking the appropriate balance for your institution between injecting new employees and transforming existing ones. Take a broader view in sourcing and a sharper look at the skills your data team requires.[348]

To build a data culture, it is important to know the barriers that you need to overcome.

Rahul Bhargava identifies the key barriers to data culture:

346 Elizabeth Dunlea, "The Key to Establishing a Data-Driven Culture", *Gartner*, 30 November 2015. Available at https://www.gartner.com/smarterwithgartner/the-key-to-establishing-a-data-driven-culture/.

347 Kathleen Kelly Janus, "Creating a Data Culture", *Stanford Social Innovation Review*, 2 March 2018. Available at https://ssir.org/articles/entry/creating_a_data_culture.

348 Alejandro Díaz, Kayvaun Rowshankish and Tamim Saleh, "Why data culture matters", *McKinsey Quarterly*, September 2018. Available at https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/why-data-culture-matters.

- **Confusion** – Most introductions to data are confusing and overly technical. The goal is not to turn everyone in the organization into data scientists. A data culture means people recognize data and are able to pinpoint new opportunities for deriving knowledge and insight from it.
- **Not knowing your data** – Sometimes you do not even know the data you have. It is hard to keep track of data sets within your organization that may be related to each other. Identify a person, (office) and a technology that can be a central clearinghouse for data.
- **Organizational silos** – Most organizations suffer from silos—independent functional units that take pains to control a slice of the overall work. You have to acknowledge these walls in order to break them down.
- **Technology - centric thinking** – People should not have to go to the ICT department to pull out the latest numbers they need.[349]

Building data culture will require a plan that includes: identifying internal advocates/experts; spotting key exemplars; building external relationships; leading from the top and from below; and taking small ("baby") steps.[350]

As in other transformation initiatives, leadership is key in establishing data culture. Below are ways that leaders can lead by example with data:

- **Daily usage** – One of the most impactful ways of sending a message to your organization that data matters is for executives to actively use the data.
- **Decisions** – If data is truly important to your organization, all of your leaders' decisions should be based on data—with no exceptions. When executives request data in order to make key decisions, they reinforce data's role as an important strategic asset that serves an integral part of the decision-making process.
- **Communications** – Each email, presentation or meeting discussion represents an opportunity to share insights on business performance, promote data-driven wins and emphasize data's importance to the organization.
- **Meetings** – If a greater emphasis is placed on reviewing key metrics and developing action plans based on the results, data can then guide the meeting agenda and make meetings more focused, productive and useful for everyone involved.
- **Training** – Where busy executives choose to spend their time can indicate how important something is to them. If executives carve out time to participate in data skills training, it sends a powerful message to their team that these skills will be critical to their team's success.
- **Digital displays** – When executives choose to display key metrics in prominent public locations via digital displays, they send a message to employees that the metrics and targets are collectively owned by their team.[351]

---

349 Rahul Bhargava, "You Don't Need a Data Scientist, You Need a Data Culture", *Data Therapy*, 6 December 2017. Available at https://datatherapy.org/2017/12/06/building-a-data-culture/.

350 Rahul Bhargava, "Architectures for Building a Data Culture", *Data Therapy*, 20 July 2015. Available at https://datatherapy.org/2015/07/20/architectures-for-building-a-data-culture/.

351 Brent Dykes, "Creating a Data-Driven Culture: Why Leading By Example is Essential", *Forbes*, 26 October 2017. Available at https://www.forbes.com/sites/brentdykes/2017/10/26/creating-a-data-driven-culture-why-leading-by-example-is-essential/#58e6672b6737.

NOTES FOR TRAINERS

PART 1

Tips on Delivering the Module:

Key Points, Exercises and Readings for each Section

SECTION 1 DATA REVOLUTION AND DATA JUSTICE

This introductory section provides the context for societies' intensive and extensive use of data to define and achieve their common future.

Key Points

1. The information and communication technology (ICT) revolution has led to:

- **A data revolution** – an explosion in the volume of data, the speed with which data is produced, the number of producers of data, the dissemination of data and the range of things on which there is data.
- **The shift from individuals as "information consumers" to "information producers". As information producers, we create digital footprints** – the sum of all data that we produce as a result of our online activities. This includes data that we create when we use social media (such as Facebook and Twitter) and when we Google. We also produce data exhaust—the by-product of our online activities. Our data exhaust provides significant information about our online behaviour, and when processed can lead to valuable insights about our habits and preferences.
- **The Internet of Things (IoT)** – Internet-connected devices that create new data without human intervention. By 2025, the IoT will produce over 2 zettabytes of data, mostly generated by consumer electronics devices.

2. We have also witnessed the evolution from "digitization" to "digitalization" to "digital transformation" to "datafication":

- **Digitization** – The process of converting analog into digital;
- **Digitalization** – Turning interactions, communications, business functions and business models into (more) digital ones. Also, the integration of digital technology into everyday life;
- **Digital transformation** – The use of digital technology in all aspects of an enterprise to fundamentally change how it creates and delivers value; and

- **Datafication** - Turning previously invisible process/activity into data that can be monitored, tracked, analysed and optimized. To "datafy" is to render into data many aspects of the world that have never been quantified before.

3. Data justice is the key challenge in a data-driven world.

Today almost half of the worlds' population is still "invisible"—they do not produce digital footprints because they do not have access to the Internet and other digital technologies. Among those who are already visible, some are seen only as consumers and clients who do not have "agency".

Data justice means fairness in the way people are made visible, represented and treated as a result of their production of digital data.

The three pillars of data justice are:

- **Visibility** – Privacy and representation of those in the margins;
- **Engagement with technology** – (1) freedom not to use specific digital technologies; (2) how not to become part of commercial databases (as a by-product of development interventions); and (3) the freedom to control the terms of one's engagement with data markets; and
- **Non-discrimination** – The power to identify and challenge bias in data use, as well as, the freedom from prejudicial treatment.

Data justice is necessary for an inclusive data-driven governance.

Suggested Exercises

1. Digitization and Digitalization in the Public Sector

- Divide participants into four groups.
- Two groups discuss digitization initiatives in their agencies.
- Other two groups discuss digitalization initiatives in their agencies.
- Each group presents a summary of their discussion.

2. The Pillars of Data Justice

- Divide participants into three groups.
- Group 1 to discuss and give example of data justice pillar of visibility.
- Group 2 to discuss and give example of data justice pillar of engagement with technologies.
- Group 3 to discuss and give example of data justice pillar of non-discrimination.

Selected Readings

United Nations Secretary-General's Independent Expert Advisory Group on the Data Revolution for Sustainable Development, "A World that Counts: Mobilizing the Data Revolution for Sustainable Development", November 2014, p. 11. Available at http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf.

Margaret Rouse, "Data Exhaust", WhatIs.com, April 2015. Available at https://whatis.techtarget.com/definition/data-exhaust.

I-Scoop, "Digitization, digitalization and digital transformation: the differences". Available at https://www.i-scoop.eu/digitization-digitalization-digital-transformation-disruption/ (accessed on 8 January 2019).

Margarita Shilova, "The Concept Of Datafication; Definition & Examples", Apiumhub, 15 June 2017. Available at https://apiumhub.com/tech-blog-barcelona/datafication-examples/.

Linnet Taylor, "What is data justice? The case for connecting digital rights and freedoms globally", 16 February 2017. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2918779.

SECTION 2 DATA AND GOVERNANCE

This section begins with a discussion of governance and then focuses on data-driven governance and data-driven government.

Key Points

1. Governance is defined as how societies organize to define and achieve their common future.

- Governance includes "demand politics" and "supply politics".
- **It occurs at different levels** – local, national and global—and has many actors, including government, civil society and business groups.
- **Governance is closely related to government** – the formal structures of the public sector and the set of actors exercising state power, or a material entity specialized to the provision of governance. Government is the most important governance actor.

2. Data-driven governance is societies' intensive and extensive use of data to define and achieve their common future.

- Data-driven government is one where, for all critical decisions, actionable information is available when and where needed.
- In the public sector, greater use of data can lead to: sounder governance and control; optimized fraud and error detection, mitigation and prevention; improved services based on insights gained from those being served; and improved efficiency that can lead to reduced costs.

The various forms of data-driven governance include:

- **Data-driven decision-making** – The process that involves collecting data, extracting patterns and facts from that data, and utilizing those facts to make inferences that influence decision-making;
- **Evidence-based policymaking** – The use of the best available research and information on programme results to guide decisions at all stages of the policy process and in each branch of government; and
- **Results-based management** – A management strategy by which all actors, contributing directly or indirectly to achieving a set of results, ensure that their processes, products and services contribute to the achievement of desired results (outputs, outcomes and higher level goals or impact).

3. Challenges in data-driven governance can be clearly seen in the monitoring of the United Nations Sustainable Development Goals (SDGs).

To ensure the effective implementation of the SDGs, the United Nations General Assembly adopted the SDG indicators—a solid framework of indicators and statistical data to monitor progress, inform policy and ensure accountability of all stakeholders. The SDGs have 232 indicators to measure progress.

The SDG indicators are categorized using a three-tier system based on methodology and data issues:

- Tier 1 – An established methodology exists, and data is widely available;
- Tier 2 – An established methodology exists, but data is not readily available; and
- Tier 3 – An internationally-agreed methodology is yet to be developed.

As of 2017:

- A third of the indicators have available data that can be used for monitoring of the SDG;
- Close to a quarter have a methodology but no data; and
- Thirty-eight per cent do not have an agreed upon methodology and no data.

In Asia, the data challenges faced by governments include:

- The need to strengthen national statistical systems to report on the SDGs and encourage innovations in data collection to simplify the tasks at hand;
- The need for countries to focus on the indicators that are most meaningful for their implementation priorities (all 232 will not be relevant in all cases); and
- Overcoming monitoring that is driven by the availability of data rather than its relevance for national priorities.

Suggested Exercise

Gauging Government's Data Maturity Model

- Participants are divided into groups.
- Each group will rate the data maturity of government (as a whole), using NESTA's scale:
- **Nascent** – Rich in data, poor in intelligence. Data is not a key part of decision-making processes.

- **Basic** – Data is used in reports but usually in a cursory way and that little reference to decisions that have to be made.
- **Intermediate** – Data analysis is usually requested for decision making, but can be inadequate because analysis is not of high quality, not targeted at the decision to be made or the right data is not available.
- **Advanced** – Some decisions are informed by data on both the frontline and at senior levels, but it is not consistent across the organization.
- **Datavore** – Data is analysed on specifically for the purposes of key decisions that have to be made, consistently across the organization. Data is available in a timely fashion to support decision-making.

All groups present results.

Selected Readings

B. Guy Peters and Jon Pierre, Comparative Governance: Rediscovering the Functional Dimension of Governing (Cambridge, Cambridge University Press, 2016).

Terence Lutes, "Data-driven government: Challenges and a path forward", IBM Analytics White Paper, 2015. Available at https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=GQW03008USEN.

Tom Symons, "Wise Council: Insights From the Cutting Edge of Data-Driven Local Government", NESTA, November 2016. Available at https://media.nesta.org.uk/documents/wise_council.pdf.

Joel Schwartz, "Data-Driven Decision Making: A Primer For Beginners", Northeastern University Blogs, 14 December 2017. Available at https://www.northeastern.edu/graduate/blog/data-driven-decision-making/.

Justin Parkhurst, The Politics of evidence: From evidence-based policy to the good governance of evidence (Oxford, Routledge, 2017).

UNDG, "Results-Based Management Handbook: Harmonizing RBM concepts and approaches for improved development results at country level", October 2011. Available at https://undg.org/wp-content/uploads/2016/10/UNDG-RBM-Handbook-2012.pdf.

UNDP, "Guidance Note: Data for Implementation and Monitoring of the 2030 Agenda for Sustainable Development", September 2017. Available at http://www.undp.org/content/undp/en/home/librarypage/poverty-reduction/guidance-note--data-for-implementation-and-monitoring-of-the-203.html.

SECTION 3 A WORLD AWASH IN DATA

Despite the data revolution, many governments in developing countries still face the challenge of lack of quality data. It is thus important to look at improving traditional data sources and harnessing new data sources for governance. This section discusses official statistics, big data, real-time data, small data and citizen-generated data.

Key Points

1. Official statistics – Numerical data sets, produced by official governmental agencies mainly for administrative purposes.

Its purpose is to produce and disseminate authoritative results designed to reliably reflect economically and socially relevant phenomena of a complex and dynamic reality in a given country.

Unlike private sector data, official statistics are based on common principles, standards, methodologies and technologies established in accordance with a professional code of ethics.

Traditionally, official statistics are comprised of census data, survey data and administrative data.

Despite attempts at being comprehensive there is a lack of official data on important concerns like gender. *A World that Counts* notes that gender inequality and the undervaluing of women's activities and priorities in every sphere has been replicated in the statistical record.

2. Big data – Data sets whose size is beyond the ability of typical database software tools to capture, store, manage and analyse. It is characterized by the 3Vs (volume, variety and velocity), where "volume" refers to the amount of data, "variety" refers to the number of types of data and "velocity" refers to the speed of data processing.

Others suggest that there are not only three but five Vs of big data, adding "veracity" (the messiness or trustworthiness of the data) and "value" (benefit or worth) to volume, variety and velocity.

Some proponents limit big data to:

- **Exhaust data** – Passively collected data from people's use of digital services such as mobile phones, financial transactions or web searches.
- **Sensing data** – Actively collected data from sensors, e.g., in smart cities or from wearables, and through remote sensing and satellite images.
- **Digital content** – Open web content actively produced by people such as social media interactions, news articles, blogs or job postings. Unlike exhaust and sensing data, this is digital content intentionally edited by somebody, i.e., subjective or even deceptive, depending on the intentions of the author.

3. Real-time data – Data that is delivered and used immediately after collection. However, in the development work, it also refers to information that is produced and made available in a relatively short and relevant period of time, and information that is made available within a timeframe that allows action to be taken in response.

Real-time data includes social media feeds, satellites imagery, sensor-monitored rainfall and flood levels as well as smartphone location data.

Real-time data enables timely decisions as well as prevent fraud.

4. Small data – The human-centric alternative to big data. It is also defined as small data sets that are capable of impacting decisions in the present.

Small data connects people with timely, meaningful insights (derived from big data and/or local sources), organized and packaged—often visually—to be accessible, understandable and actionable for everyday tasks.

According to John Spacey, Small data was previously simply known as data. The modern term is used to distinguish between traditional data configurations and big data.

5. Citizen-generated data – Data that people or their organizations produce to directly monitor, demand or drive change on issues that affect them.

Citizen-generated data can highlight issues that are important to people and feed their views up into higher-level policy debates.

Uses:

- Verify official narratives and data sets;
- Bring accountability to development processes through independent monitoring; and
- Provide a direct representation of citizens and enable them to directly monitor, demand or drive change on issues that affect them.

Examples include citizen-created data on air quality in Beijing, a sexual harassment map in Egypt, and updated water point statuses in Tanzania.

Suggested Exercise

Citizen-Generated Data

- Participants are divided into groups.
- Each group to identify and discuss possible citizen-generated data initiative in their country to help monitor the Sustainable Development Goals.
- All groups present results of discussion.

Selected Readings

Walter J. Radermacher, "The Future Role of Official Statistics", Power from Statistics Outlook Report. Available at https://www.researchgate.net/publication/320616460_The_Future_Role_of_Official_Statistics.

Soenke Ziesche, Innovative Big Data Approaches for Capturing and Analyzing Data to Monitor and Achieve the SDGs (Bangkok, ESCAP, 2017). Available at https://reliefweb.int/sites/reliefweb.int/files/resources/Innovative%20Big%20Data%20Approaches%20for%20Capturing%20and%20Analyzing%20Data%20to%20Monitor%20and%20Achieve%20the%20SDGs.pdf.

United Nations Global Pulse, "Big Data for Development: Challenges and Opportunities", May 2012. Available at http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseMay2012.pdf.

Andreas Pawelke and others, Data for development: What's next? – Concepts, trends and recommendations for German development cooperation (Bonn and Eschborn, Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH, 2017). Available at http://webfoundation.org/docs/2017/12/Final_Data-for-development_Whats-next_Studie_EN-1.pdf.

Roger Dooley, "Small Data: The Next Big Thing", Forbes, 16 February 2016. Available at https://www.forbes.com/sites/rogerdooley/2016/02/16/small-data-lindstrom/#1811e5fb7870.

Development Initiatives, "Citizen-generated data and sustainable development: Evidence from case studies in Kenya and Uganda report", March 2017. Available at https://hivos.org/sites/default/files/publications/15-citizen-generated-data-and-sustainable-development-evidence-from-case-studies-in-kenya-and-uganda.pdf.

SECTION 4 DATA GOVERNANCE AND DATA MANAGEMENT

Data governance and data management ensure that data is readily accessible, effortlessly shared and smoothly reusable. This is the intermediate activity/stage between collecting/generating data and analysing data for patterns.

Key Points

1. Data governance – The comprehensive process for controlling the integrity, use, availability, usability and security of all data owned by or controlled by an enterprise.

It is a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models that describe who can take what actions with what information, and when, under what circumstances, using what methods.

The goals of data governance are to:

- Enable an organization to manage its data as an asset;
- Define, approve, communicate and implement principles, policies, procedures, metrics, tools and responsibilities for data management; and
- Monitor and guide policy compliance, data usage and management services.

2. Data management – The development and execution of processes, architectures, policies, practices and procedures in order to manage the information generated by an organization.

It includes acquiring, validating, storing, protecting and processing data to ensure its accessibility, reliability and timeliness for its users.

Data management helps ensure the availability of data when and where it is needed. It is necessary particularly when government data is stored in a variety of formats and systems that support specific programmes, departments and organizations.

Data management focuses on the following knowledge areas:

- **Data architecture** – Identifies the data needs of the enterprise (regardless of structure), and designs and maintains the master blueprint to meet those needs;
- **Data modelling and design** – The process of discovering, analysing and scoping data requirements, and then representing and communicating these data requirements in a precise form called the data model;
- **Data storage and operations** – Includes the design, implementation and support of stored data to maximize its value through its lifecycle, from creation/acquisition to disposal. Its two sub-activities are database support and data technology support;
- **Data security** – Planning, development and execution of security policies and procedures to provide proper authentication, authorization, access and auditing of data and information assets;
- **Data integration and interoperability** – The processes related to the movement and consolidation of data within and between data stores, applications and organizations;
- **Documents and content management** – Controlling the capture, storage, access, and use of data and information stored outside relational databases;
- **Reference and master data** – Managing shared data to meet organizational goals, reduce risks associated with data redundancy, ensure higher quality and reduce the cost of data integration;
- **Data warehousing and business intelligence** – Planning, implementation and control processes to provide decision-support data and support knowledge workers engaged in reporting, query and analysis;
- **Metadata** - Includes information about technical and business processes, data rules and constraints, and logical and physical data structures; and
- **Data quality management** – Planning, implementation and control of activities that apply quality management techniques to data, in order to assure it is fit for consumption and meets the needs of data consumers.

3. Progress in data management can be measured through a data management maturity model. The following are the stages of said model:

- **Level 0** – No Capability. No organized data management practices or formal enterprise processes for managing data.
- **Level 1** – Initial/Ad Hoc. General purpose data management practices using a limited tool set, with little or no governance.
- **Level 2** – Repeatable. Emergence of consistent tools and role definition to support process execution. Organization begins to use centralized tools and provide more oversight for data management.
- **Level 3** – Defined. Emerging data management capability. Introduction and institutionalization of scalable data management processes and a view of data management as an organizational enabler.
- **Level 4** – Managed. Standardized tools for data management from desktop to infrastructure, coupled with a well-formed centralized planning and governance functions. Expressions of this level are a measurable increase in data quality and organization-wide capabilities such as end-to-end data audits.

- **Level 5** – Optimization. Proliferation of data is controlled to prevent needless duplication. Well understood metrics are used to manage and measure data quality and processes

Suggested Exercise

Gauging Data Management Maturity

- Participants are divided into groups.
- Each group discusses and identifies government's data management maturity level (from 0-5).
- All groups present results of discussion.

Selected Readings

DAMA International, Data Management Body of Knowledge, second edition (DAMA-DMBOK 2) (New Jersey, Technics Publications, 2017).

Data Governance Institute, "Definitions of Data Governance". Available at http://www.datagovernance.com/adg_data_governance_definition/.

Pulse Lab Jakarta, "Identifying Opportunities to Strengthen Data Governance in the Public Sector", Medium, 8 October 2017. Available at https://medium.com/pulse-lab-jakarta/identifying-opportunities-to-strengthen-data-governance-in-the-public-sector-92d9bc7db5ad.

Matthew Chase, "The state of data management in the public sector in 2018", Experian, 8 February 2018. Available at https://www.edq.com/blog/the-state-of-data-management-in-the-public-sector-in-2018/.

SECTION 5 DATA ANALYTICS, BIAS AND INTUITION

To gain insight from data they must be analyzed.

There is also need to look out for bias and determine the proper role of intuition in decision-making.

Key Points

1. Data Analytics – The pursuit of extracting meaning from raw data using specialized computer systems that transform, organize and model the data to draw conclusions and identify patterns.

There are four types of analytics:

- **Descriptive analytics** – Provide insight to answer: What has happened?
- **Diagnostic analytics** – Measure historical data against other data to answer: Why something happened?
- **Predictive analytics** – Use statistical models and forecasts techniques to answer: What could happen?
- **Prescriptive analytics** – Suggests: What action to take?

Using analytics involves applying an algorithmic or mechanical process to derive insights.

Algorithms are effective and efficient tools for analysis and problem-solving. However, they are not necessarily neutral or bias-free.

2. Algorithmic Bias

An algorithm is a set of mathematical instructions or rules that, especially given to a computer, will help to calculate an answer to a problem.

Algorithmic bias occurs when human prejudice and partiality are incorporated in the design. Consequently, discrimination is embedded into the model.

Algorithmic bias can be the result of problems at different stages. These include biased or otherwise poor quality input data, poorly defined rules, lack of contextual awareness, and feedback loops.

Algorithm bias can lead to two type of harms:

- **Allocative harm** – When the algorithm apportions or withholds certain opportunities or resources based on prejudiced assumptions; and
- **Representational harm** – When systems reinforce the subordination of some groups along the lines of identity.

Another issue that the use of analytics and algorithms raises is the role of intuition in decision-making and governance.

3. Intuition and Decision-Making

Malcom Gladwell in his 2005 book, *Blink: The Power of Thinking Without Thinking*, defends the continuing importance of intuition.

Research has identified the following conditions when one can rely on intuition:

- **Presence of expertise** – The more experience one has in the domain, the better the odds of using intuition to make domain-specific decisions;
- **Nature of the problem** – Intuition can be used on unstructured problems—those that lack clear decision rules or has few objective criteria with which to make the decision; and
- **Time available** – In situations that demand quick decision and there is little time for detailed analysis.

Intuition should NOT be used on situations that have clear decision rules and objective criteria, and abundant data with which to perform an analysis.

It is important to note that:

- The evidence is overwhelming that, whenever the option is available, relying on data and algorithms alone usually lead to better decisions and forecasters than relying on the judgement of even experienced and "experts" humans.

Suggested Exercises

Data Analytics

- Divide participants into four groups.
- Topic Assignment: Group 1 – descriptive analytics, Group 2 – diagnostic analytics, Group 3 – predictive analytics, and Group 4 – prescriptive analytics.
- All groups discuss actual or potential use for government of the type of analytics assigned to them.
- All groups present.

Limits of Intuition

- Divide the participants into groups.
- Each group identify and discuss specific/concrete instances/cases where the three conditions (presence of expertise, nature of the problem, time available) to justify the use of intuition exists.
- All groups present results of discussion.

Selected Readings

Anushka Mehta, "Four Types of Business Analytics to Know", Analytics Insight, 13 October 2017. Available at https://www.analyticsinsight.net/four-types-of-business-analytics-to-know/.

Sid Frank and Traci Gusher, "Better data, better government: Effective use of data and analytics at all levels deliver improved citizen services and outcomes", KPMG Government Institute, June 2016. Available at https://assets.kpmg/content/dam/kpmg/pdf/2016/06/co-gv-6-better-data,-better-government.pdf.

Keith Kirkpatrick, "Battling Algorithmic Bias" Communications of the ACM, vol. 59, no. 10 (October 2016). Available at https://cacm.acm.org/magazines/2016/10/207759-battling-algorithmic-bias/abstract.

World Wide Web Foundation, Algorithmic Accountability: Applying the concept to different country contexts (Washington D.C., 2017). Available at http://webfoundation.org/docs/2017/07/Algorithms_Report_WF.pdf.

United Kingdom Department for Digital, Culture, Media and Sport, "Guidance: Data Ethics Framework (updated 30 August 2018)". Available at https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework.

Malcom Gladwell, Blink: The Power of Thinking Without Thinking (New York and Boston, Blackbay Books, 2005).

Connson Chou Locke, "When It's Safe to Rely on Intuition (and When It's Not)", Harvard Business Review, 30 April 2015. Available at https://hbr.org/2015/04/when-its-safe-to-rely-on-intuition-and-when-its-not.

SECTION 6 ENABLING POLICIES

Data-driven governance will not spontaneously emerge in the wake of the data revolution. Policies and programmes have to be enacted/or implemented to realize it. These includes strengthening data collection; institutionalizing data governance; heightening public access and trust, and gearing up for big data.

Key Points

1. Strengthening Data Collection

There are three aspects to strengthening data collection: improving national statistical systems, enhancing gender statistics and spurring citizen-generated data.

The *Cape Town Global Action Plan for Sustainable Development Data* is a useful guide to improving data collection at the national level.

It provides the necessary actions to generate quality and timely data on a routine basis to inform sustainable development at the requested level of disaggregation and population coverage, including for the most vulnerable and hard-to-reach groups.

It focuses on six strategic areas for action:

- Coordination and strategic leadership on data for sustainable development;
- Innovation and modernization of national statistical systems;
- Strengthening of basic statistical activities and programmes, with particular focus on addressing the monitoring needs of the 2030 Agenda for Sustainable Development;
- Dissemination and use of sustainable development data;
- Multi-stakeholder partnerships for sustainable development data; and
- Mobilize resources and coordinate efforts for statistical capacity building.

There is also need to improve gender statistics—the scientific notation and interpretation of statistics that in an adequate and complete way are reflecting the living conditions and situations of women and men with respect to all policy fields and areas.

A key resource for this action area is the United Nations Statistical Commission's Minimum Set of Gender Indicators. This set includes 52 quantitative indicators and 11 qualitative indicators organized into five domains: (1) economic structures and access to resources; (2) education; (3) health and related services; (4) public life and decision-making; and (5) human rights of women and child.

Another key resource is ESCAP's core set of gender indicators for Asia and the Pacific, which is an addition to the global Minimum Set of Gender Indicators.

The Asia-Pacific core set of gender indicators is comprised of:

- Six basic domains:
- Five supplementary domains;
- Priority target group (rural women); and
- Qualitative indicators related to national norms.

The third component in strengthening data collection is citizen-generated data.

Citizen-generated data complements official statistics. It also directly addresses data justice as citizen-produced data can make visible individuals and groups who are invisible from government due to absence of data.

For government, actions to spur citizen-generated data include:

- Looking for ways to engage with, acknowledge and support these initiatives;
- Investigating funding and sustainability models for citizen-generated data and civil society data;
- Supporting consultation and participation processes to enable citizen and civil society input regarding institutional data collection, including through events, responsive communication channels and participatory design processes; and
- Supporting further research and the development of resources in this area that can be used to make public data infrastructures more responsive to the interests and concerns of civil society.

2. Institutionalizing Data Governance

Governments should consider developing a data governance strategy that includes:

- The drivers, vision, mission and principles for data governance, including readiness assessment, internal process discovery, and current issues or success criteria;
- The structures and responsibilities for data governance activities;
- The time frame for the roll out of policies and directives, business glossary, architecture, asset valuation, standards and procedures, expected changes to business and technology processes, and deliverables to support auditing activities and regulatory compliance; and
- The target state of sustainable data governance activities.

3. Heightening Public Access and Trust

Includes two activities—adopting an open data policy and strengthening data privacy.

- **Adopting Open Data**

Open data (or open government data) is data produced or commissioned by government or government-controlled entities that can be freely used, reused and redistributed by anyone. It is also a philosophy—and increasingly a set of policies—that promotes transparency, accountability and value creation by making government data available to all.

The three building blocks for open data to achieve its promise of transparency, citizen empowerment and innovation are:
*1. The publication of open data by governments;

*2. The conversion of data to actionable information by intermediaries; and

*3. The use of data by citizens, government officials and other stakeholders to achieve development outcomes.

The Sunlight Foundation provides governments a useful guide to developing and implementing an open data policy. The guideline is divided into three parts:

- What Data Should Be Public (7 recommendations)
- How To Make Data Public (16 recommendations)
- How to Implement an Open Data Policy (8 recommendations)

- **Strengthening Data Privacy**

This includes passing national data privacy legislation that enhances the individual's control over how personal information is collected, used, shared and disposed.

In 1980, the Organisation for Economic Co-operation and Development (OECD) released the following Data Protection Principles:

- Collection Limitation
- Data Quality
- Purpose Specification
- Use Limitations
- Security Safeguards
- Openness
- Individual Participation
- Accountability

These principles were in adopted in various national legislations and model codes globally.

A recent development is the use of ethics as an important complement to legal obligations to achieve privacy goals.

Nobel laureate Joseph Stiglitz believes that there is a need for regulations on what data the tech firms can store; what data they can use; whether they can merge different data sets; the purposes for which they can use that data; and what degree of transparency they must provide about what they do with the data.

4. Gearing up for Big Data

Creating an enabling environment for big data requires:

- Leadership – Commitment to lead at all levels and across all domains;
- Winning and retaining public trust – This means: (1) personal data collected and held by an agency is used appropriately and effectively; and (2) that personal data is secure, particularly when it is being shared by different agencies; and
- Building civil service capability in collecting, storing, analysing, sharing and using data.

There is also need to address the seven dimensions of big data for development:

1. Information – Includes both information and data, as the precursor to information;
2. Technology – Mainly focuses on the information and communication technologies (ICTs) that handle data;
3. Processes – The activities undertaken in generating, capturing, analysing, presenting and using data;
4. Objectives and values – The "objectives" component covers issues of self-interest and politics, and informal and formal strategies; and the "values" component covers culture (e.g., what stakeholders feel are the right and wrong ways to do things);
5. Skills and knowledge – Covers both qualitative and quantitative aspects of human competencies for undertaking data-related processes;
6. Management systems and structures – The broader management systems required to organize within and between data-related organizations and networks, plus the way in which those systems are structured, both formally and informally; and
7. Other resources – Time and money.

Suggested Exercise

Data-Driven Governance Snapshot

- Divide participants into four groups.
- Each group should assess, using a three-point scale, government's readiness for data-driven governance by rating the progress of the following: (1) strengthening data collection; (2) institutionalizing data governance; (3) heightening public access and trust; and (4) gearing up for big data.
- To facilitate discussion and reporting, use the matrix below.

	0 Not Started	1 Some Progress	2 Completed
1) Data Collection • Improve National Statistical System • Enhance Gender Statistics • Minimum Set of Gender Indicators • Asia-Pacific Core Set • Spurring Citizen-Generated Data			

	0 Not Started	1 Some Progress	2 Completed
2) Institutionalizing Data Governance			
3) Heightening Public Access and Trust • Open Government Data • Strengthening Data Privacy			
4) Gearing up for Big Data • Leadership • Public Trust • Civil Service Capacity Building			

All groups present result of discussion.

Selected Readings

High-level Group for Partnership, Coordination and Capacity-Building for Statistics for the 2030 Agenda for Sustainable Development, "Cape Town Global Action Plan for Sustainable Development Data", adopted by the United Nations Statistical Commission at its 48th Session, March 2017, p. 2. Available at https://unstats.un.org/sdgs/hlg/Cape_Town_Global_Action_Plan_for_Sustainable_Development_Data.pdf.

United Nations Statistics Division, "Minimum Set of Gender Indicators". Available at https://genderstats.un.org/#/home.

ESCAP, "Core set of gender indicators for Asia and the Pacific: Note by the secretariat", Committee on Statistics, Fourth Session, 23 January 2015 (E/ESCAP/CST(4)/10). Available at http://www.asiapacificgender.org/sites/default/files/pdf/statstics_documents/Core_Set_Gender_Indicators_Asia_Pacific.pdf.

Danny Lammerhirt, Shazade Jaeson and Eko Presetyo, "Making Citizen Generated Data Work: Towards a Framework Strengthening Collaborations Between Citizens, Civil Society Organisations, and Others", Data Shift, March 2017. Available at http://civicus.org/thedatashift/wp-content/uploads/2017/03/Making-Citizen-Generated-Data-Work_short-report_.pdf.

NASCIO, "Data Governance Part II: Maturity Models – A Path to Progress", March 2009. Available at https://nascio.org/Portals/0/Publications/Documents/NASCIO-DataGovernancePTII.pdf.

Sunlight Foundation, "Open Data Policy Guidelines", Open Data Policy Hub. Available at https://opendatapolicyhub.sunlightfoundation.com/guidelines/.

Holvast, "History of Privacy", in The Future of Identity, V. Matyáš and others, eds. (International Federation for Information Processing, 2009). Available at https://link.springer.com/content/pdf/10.1007/978-3-642-03315-5_2.pdf.

United Nations Global Working Group for Big Data, "Using Big Data for the Sustainable Development Goals". Available at https://unstats.un.org/bigdata/taskteams/sdgs/.

Richard Heeks, "Measuring Barriers to Big Data for Development", ICTs for Development, 9 August 2016. Available at https://ict4dblog.wordpress.com/2016/08/09/measuring-barriers-to-big-data-for-development/.

UNDP and United Nations Global Pulse, A Guide to Data Innovation for Development: From Idea to Proof-Of-Concept, version 1 (New York, 2016). Available at http://www.undp.org/content/undp/en/home/librarypage/development-impact/a-guide-to-data-innovation-for-development---from-idea-to-proof-.html.

SECTION 7 CREATING A DATA CULTURE IN THE PUBLIC SECTOR

Data culture can accelerate that adoption of data-driven governance and amplify its power.

Key Points

Data culture means using data in a pervasive way in an organization.

In the public sector, data culture means a deep, organization-wide comfort level with using metrics to maximize social impact.

Its six principles are as follows:

- Data culture is decision culture – The fundamental objective in collecting, analysing and deploying data is to make better decisions;
- Data culture and leadership – Commitment from the highest levels must go beyond high-level pronouncements;
- Democratization of data – Get data in front of people by stimulating demand from the grass roots;
- Data culture and risk – An effective data culture puts risk at its core. Risk management should operate as a smart accelerator, by introducing analytics into key processes and interactions in a responsible manner;
- Culture catalysts – Someone has to lead the charge—someone who can bridge both worlds—data science and on-the-ground operations; and
- Marrying talent and culture – Striking the appropriate balance between injecting new employees and transforming existing ones.

Building data culture would require a plan that includes: identifying internal advocates / experts; spotting key exemplars; building external relationships; leading from the top and from below; and taking small ("baby") steps.

Selected Readings

Kathleen Kelly Janus, "Creating a Data Culture", Stanford Social Innovation Review, 2 March 2018. Available at https://ssir.org/articles/entry/creating_a_data_culture.

Rahul Bhargava, "You Don't Need a Data Scientist, You Need a Data Culture", Data Therapy, 6 December 2017. Available at https://datatherapy.org/2017/12/06/building-a-data-culture/.

NOTES FOR TRAINERS

PART 2
Delivery Options

Depending on the audience, time constraint and focus, there are at least five ways to deliver this module.

Data-Driven Governance Training of Trainers Course
2.5 days

Day 1

08:00 - 9:00	Opening
09:30 - 11:30	Section 1: Data Revolution and Data Justice
01:00 - 03:00	Section 2: Data and Governance
03:30 - 05:30	Section 3: A World Awash in Data

Day 2

08:00 - 10:00	Section 4: Data Governance and Data Management
10:30 - 12:30	Section 5: Data Analytics, Algorithmic Bias and Intuition
01:00 - 02:30	Section 6: Enabling Policies - 1
03:00 - 04:30	Section 6: Enabling Policies - 2
04:30 - 05:30	Section 7: Data Culture

Day 3

08:30 - 10:00	Developing National Delivery Plans for the Realizing Data-Driven Governance Module
10:30 - 11:30	(Country/Participant) Presentation
11:30	Closing

Data-Driven Governance (Regular) Course
2 days

Day 1

08:30 - 09:00	Opening
09:30 - 10:30	Section 1: Data Revolution and Data Justice
10:30 - 11:30	Section 2: Data and Governance
01:00 - 02:30	Section 3: A World Awash in Data
03:00 - 04:30	Section 4: Data Governance and Data Management

Day 2

08:30 - 10:00	Section 5: Data Analytics, Algorithmic Bias and Intuition
10:30 - 12:00	Section 6: Enabling Policies - 1
01:00 - 02:30	Section 6: Enabling Policies - 2
03:00 - 04:00	Section 7: Data Culture
04:00 - 05:00	Closing

Data-Driven Governance Executive Course
1 day

08:30 - 09:00	Opening
09:30 - 10:30	Section 2: Data and Governance
10:30 - 11:30	Section 4: Data Governance and Data Management
01:00 - 2:30	Section 6: Enabling Policies - 1
03:00 - 04:30	Section 6: Enabling Policies - 2
04:30 - 05:00	Closing

Data-Driven Governance – Big Data Course
1 day

08:30 - 09:00	Opening
09:30 – 10:30	Section 1 & 2: Data Revolution and Data Justice, Data and Governance
11:00 – 12:00	Section 3: Big Data (A World Awash in Data)
01:00 – 02:30	Section 5: Data Analytics, Algorithmic Bias and Intuition
03:00 - 04:00	Section 6: Gearing Up for Big Data (Enabling Policies)
04:00 - 05:00	Section 7: Data Culture
05:00	Closing

Data-Driven Governance – Data Governance and Management Course
1 day

8:30 - 9:00	Opening
9:30 - 10:30	Section 1: Data Revolution and Data Justice
11:00 – 12:00	Section 2: Data and Governance
01:00 - 02:30	Section 4: Data Management - 1
03:00- 04:30	Section 4: Data Management – 2
04:30- 05:00	Section 7: Data Culture
05:00	Closing

APCICT/ESCAP

The Asian and Pacific Training Centre for Information and Communication Technology for Development (APCICT) is a regional institute of the Economic and Social Commission for Asia and the Pacific (ESCAP). APCICT aims to strengthen the efforts of the member countries of ESCAP to use ICT in their socioeconomic development through human and institutional capacity building. APCICT's work is focused on three pillars: training, knowledge sharing, and multi-stakeholder dialogue and partnership. Together they form an integrated approach to ICT human capacity building.

APCICT is located at Incheon, Republic of Korea.

http://www.unapcict.org

ESCAP

The Economic and Social Commission for Asia and Pacific (ESCAP) is the regional development arm of the United Nations and serves as the main economic and social development centre for the United Nations in Asia and the Pacific. Its mandate is to foster cooperation between its 53 members and 9 associate members. ESCAP provides the strategic link between global and country-level programmes and issues. It supports governments of countries in the region in consolidating regional positions and advocates regional approaches to meeting the region's unique socioeconomic challenges in a globalizing world.

The ESCAP office is located at Bangkok, Thailand.

http://www.unescap.org