Module 2 – Fundamentals
Submodule 2.4
Data
[beta version]

Thus far, you...

- learned about the background and objectives of the E-Government Survey
- learned about the scope and purpose of the E-Government Survey
- learned about the structure and methodology of the E-Government Survey
- learned about how the E-Government Survey can help in achieving SDGs

In this section you will learn...

Background

- Definitions
- Data-Driving Decision-making in public sectors
- Why Data Matters
- Open Government Data

Data Governance

- Definition
- Data Governance Roadmap
- Data Governance Framework

Data Types & Applications

- Big Data and SDGs
- Challenges and Opportunities
- Other Data Driven Technology

Objective

- By the end of this submodule, you will be able to:
  - ✓Understand the definition and characteristics of open data
  - ✓Understand the framework of data governance
  - ✓Understand data-driven technologies and applications

Completion time

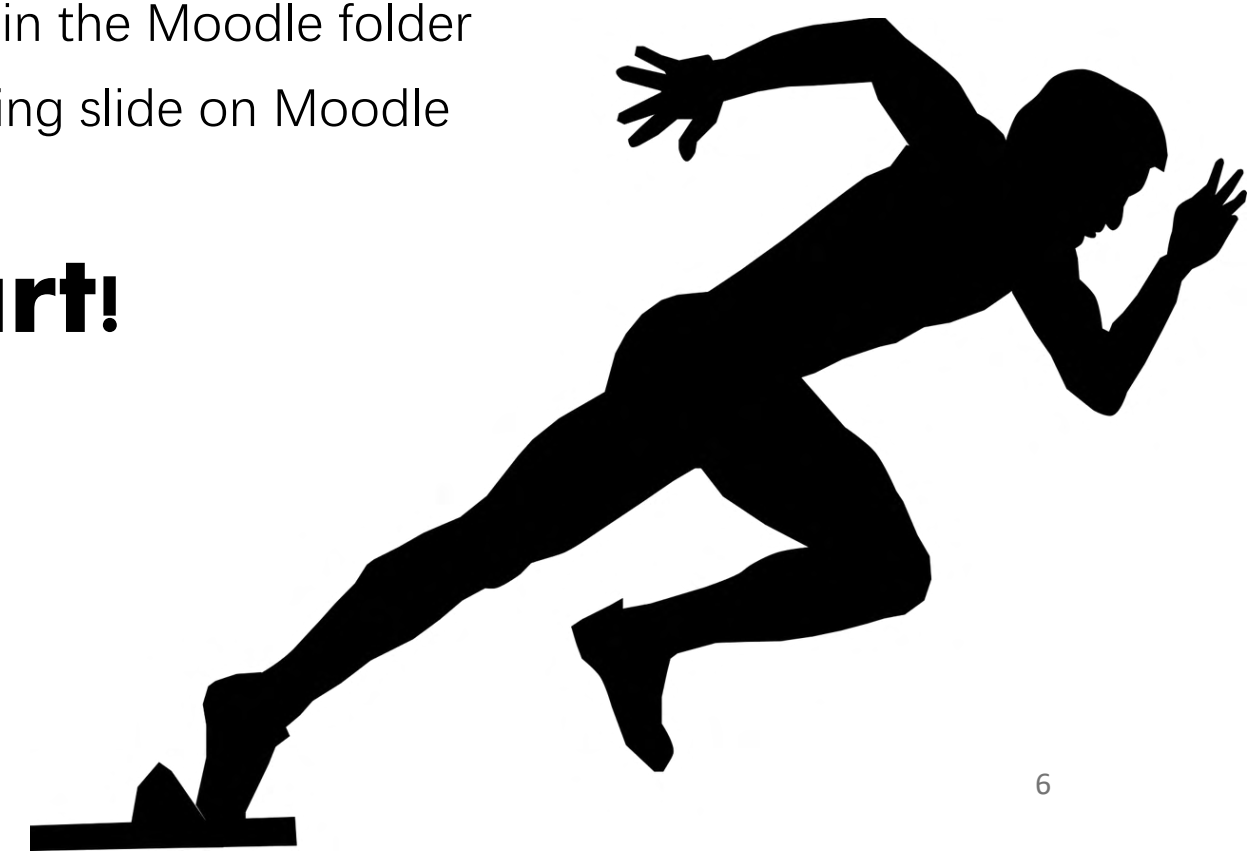
- In total there are around **30 pages** for this submodule. It will take approximately **30 to 40 minutes** for each user to complete. This is an indication and can differ per user.

- Feel free to skip some parts of this submodule if you are already familiar with the content.

Other Information

- You can read along (PDF) as well as listen to the content (audio) while taking this course;
  - Course material (PDF) can be downloaded in the Moodle folder
  - Audio can be streamed on the corresponding slide on Moodle

Let's start!



Definitions

- ***Open Data*** :

Open information in terms of access, redistribution, reuse, absence of technological restriction, attribution, integrity,

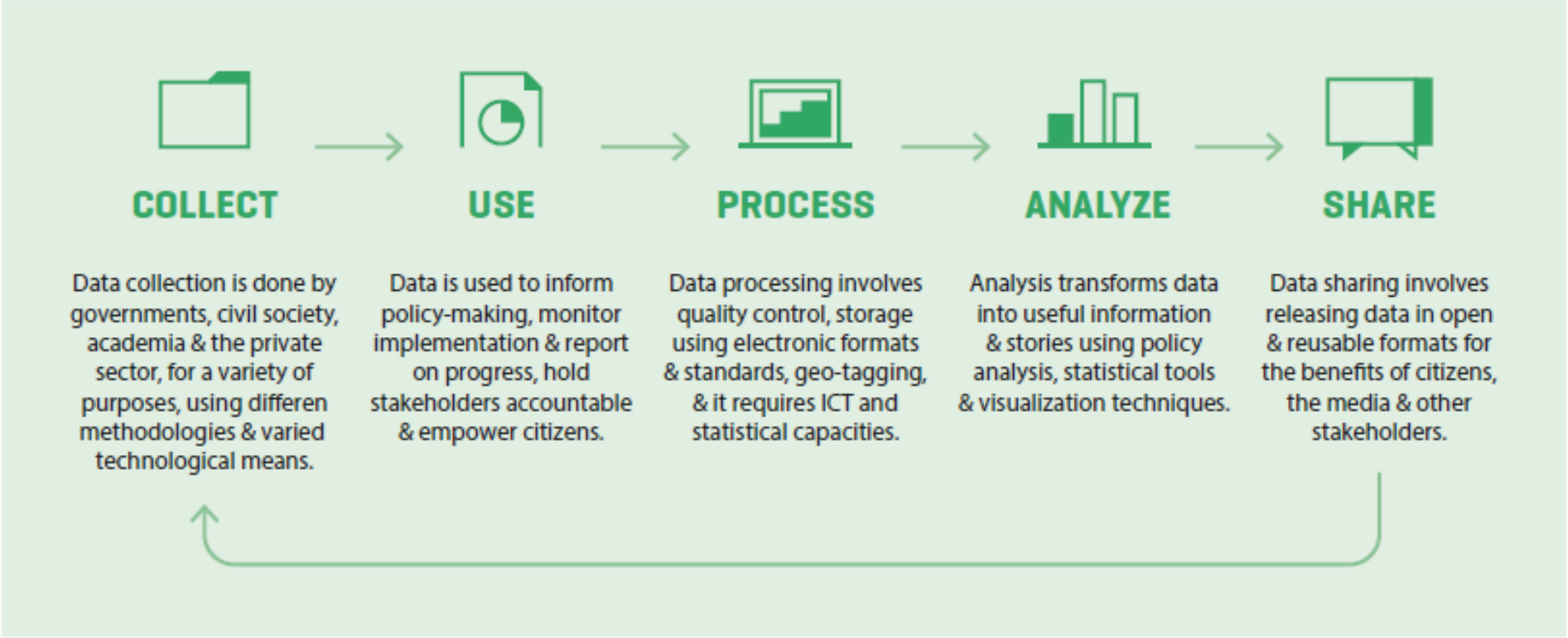
Open Government Data:

Data produced or commissioned by public bodies or government-controlled entities, which is then made accessible, and can be used feely, reused and redistributed by anyone.

Big Data: Volume, Velocity and Variety.

Data Value Chain

Figure 1 The Data Value Chain



(UNDP: Data Ecosystems for Sustainable Development

Important characteristics of Open Data

01. Availability and Access

The data must be available as a whole, preferably by downloading over the Internet. The data must also be available in a convenient and modifiable form.

02. Reuse and Redistribution

The data must be provided under terms that permit reuse and redistribution including the intermixing with other datasets

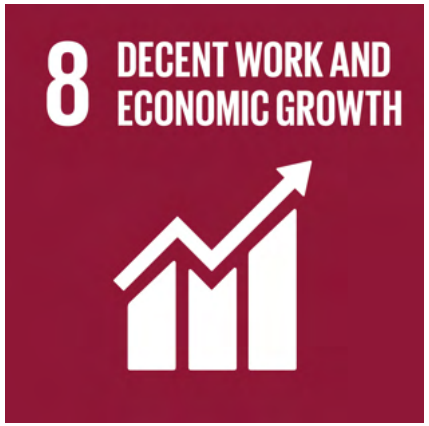
03. Universal Participation

Everyone must be able to use, reuse and redistribute - there should be no discrimination against fields of endeavor or against persons or groups.

Data-Driven decision-making in the public sector

Example:
Detecting epidemics in early stages

Example:
professional social networks analysis for a more prospective vision of the employment market



(UNDESA, 2018)

Data-Driven decision-making in the public sector

Example: Global Fishing Watch prototype

Identifying potential fishing vessels by collecting behavioral patterns of vessels.

Example:
Data mining techniques to support the structuring of evidence in court cases.

(UNDESA, 2018)

Why data matters

Transparency and accountability

Participation and citizen engagement

Improved efficiency and effectiveness of government services

Empowerment of citizens and communities

Impact measurement of policies

Improved or new public services through benchmarking

Innovations and economic growth

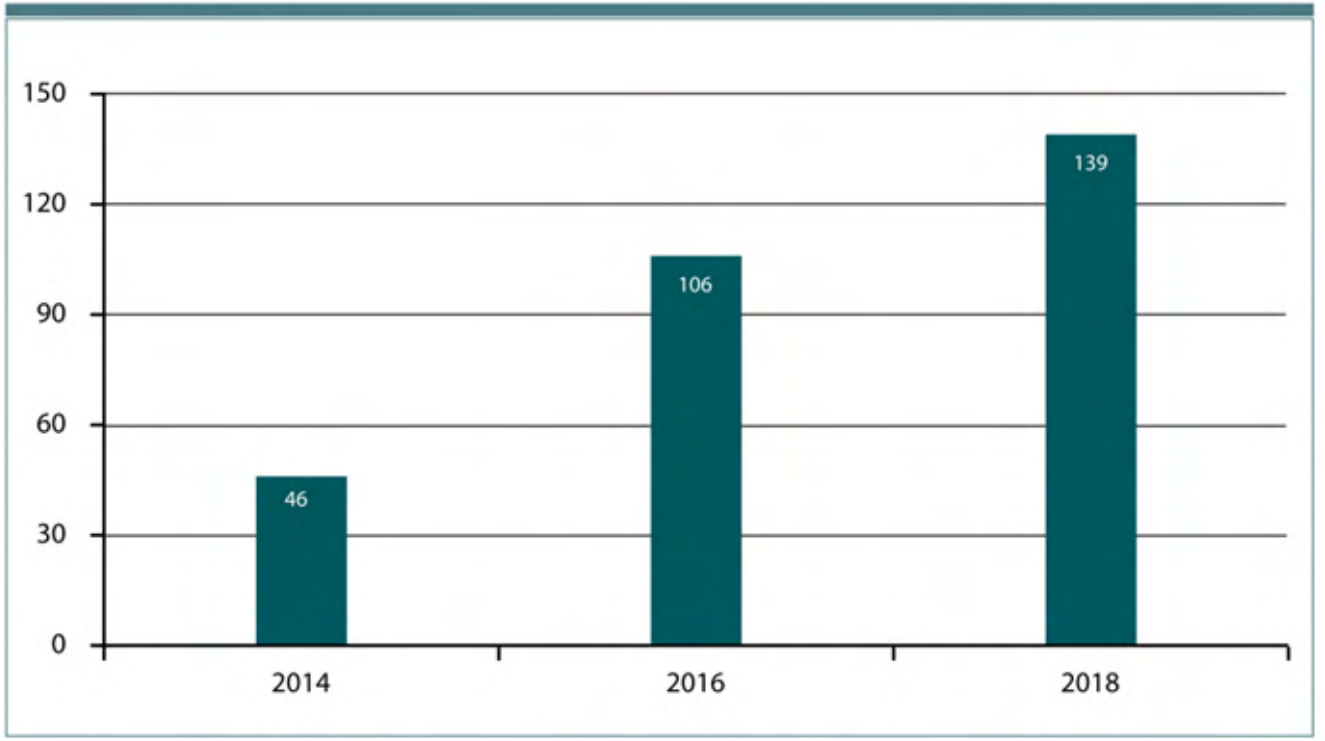
New knowledge from combined data sources or analysis of big data

Open Government Data for Sustainable Development

Open government data (OGD) contributes to :

- Generating better data for monitoring SDGs
- Supporting Goal 16: to build effective, accountable and inclusive institutions at all levels!

Figure 5.18 Countries with Open Government Data Portal and/or Catalogues in 2014, 2016 and 2018



OGD and Services

UNDPIDG has been conducting research on Open Government Data , services and guidelines for building OGD Portals are provided in following areas:

- **OGD for Citizen Engagement Guidelines**
- **OGD for Sustainable Development**
- **OGD Readiness Assessment**

Available at : https://publicadministration.un.org/en/ogd

Let’s take a break!

Are you up for a challenge?

MCQ_1

Which of the following can be considered ‘Open Government Data’ according to what we have learned?

A) A government published the annual government planning online in PDF format.

B) A government released the national real-time air quality data to all the environmental institutes nationwide.

C) A government disclosed the geo-spatial data of all the available kindergartens in its capital area, and prohibited all commercial use of that dataset.

D) A government released its data of oil import (1970-2017) online in CSV format and can be downloaded and reused for free.

Data Governance

What is Data Governance?

Data Governance is

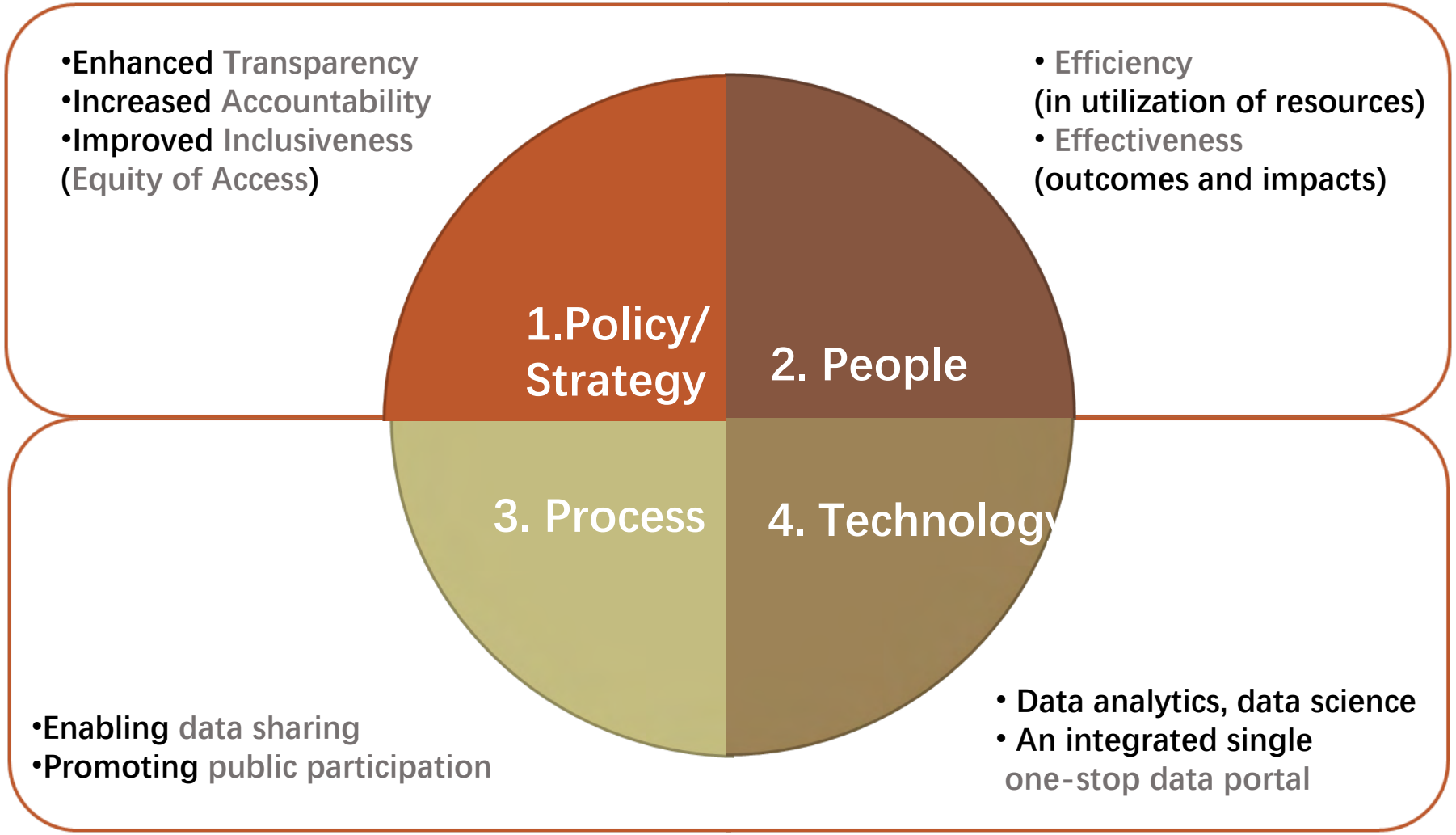
"the formal orchestration of people, processes and technology to enable an organization to leverage data as an enterprise asset."

Ref: John Ladley, *Data Governance* (Morgan Kaufmann, 2012).

"the organization and implementation of policies, procedures, structure, roles, and responsibilities which outline and enforce rules of engagement, decision rights, and accountabilities for the effective management of information assets"

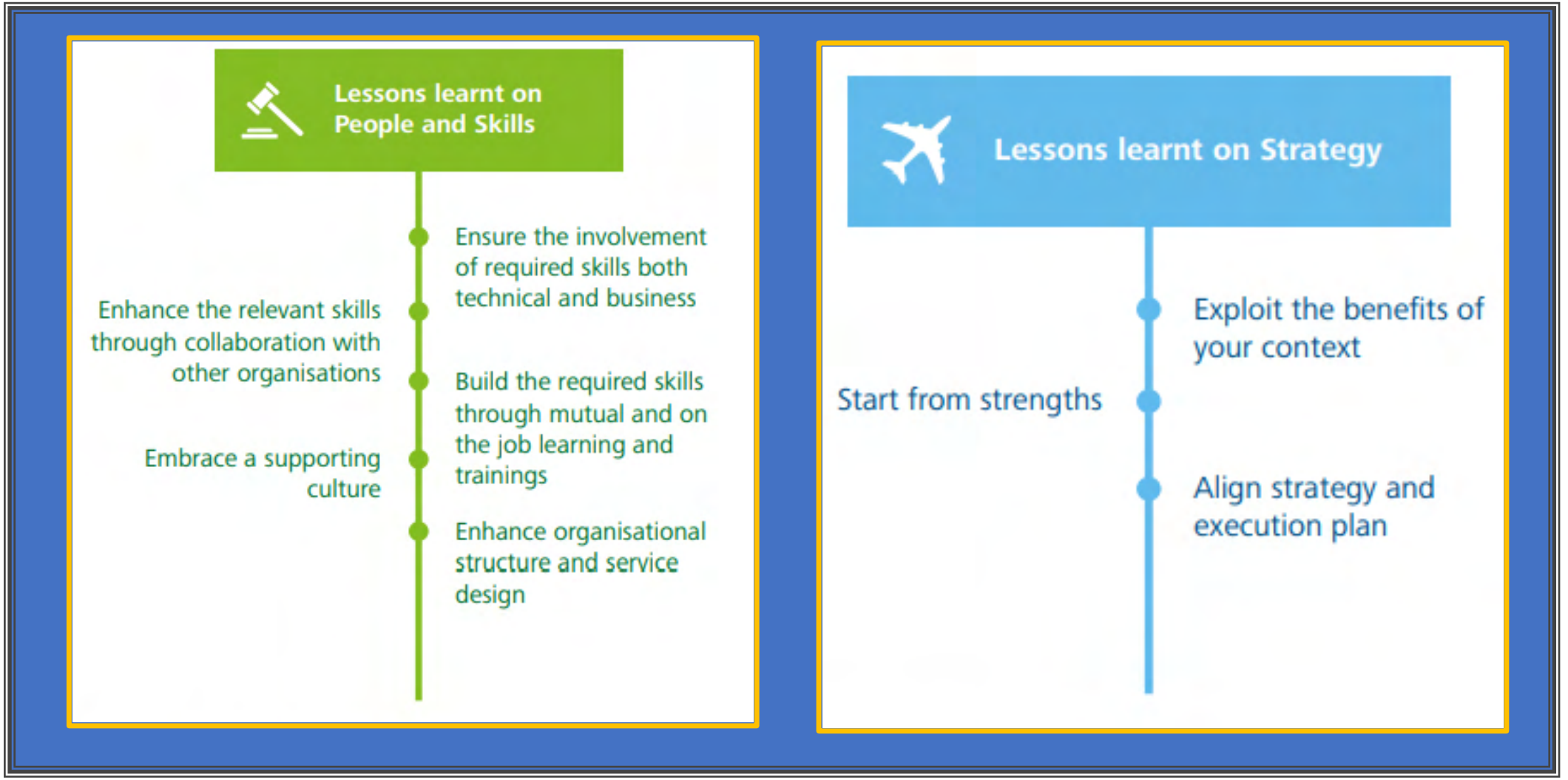
Ref: Berson and Dubov, *Master Data Management and Data Governance* (New York: McGraw-Hill, 2011)

Data Governance: A framework Approach



(Dizon, et al. 2017)

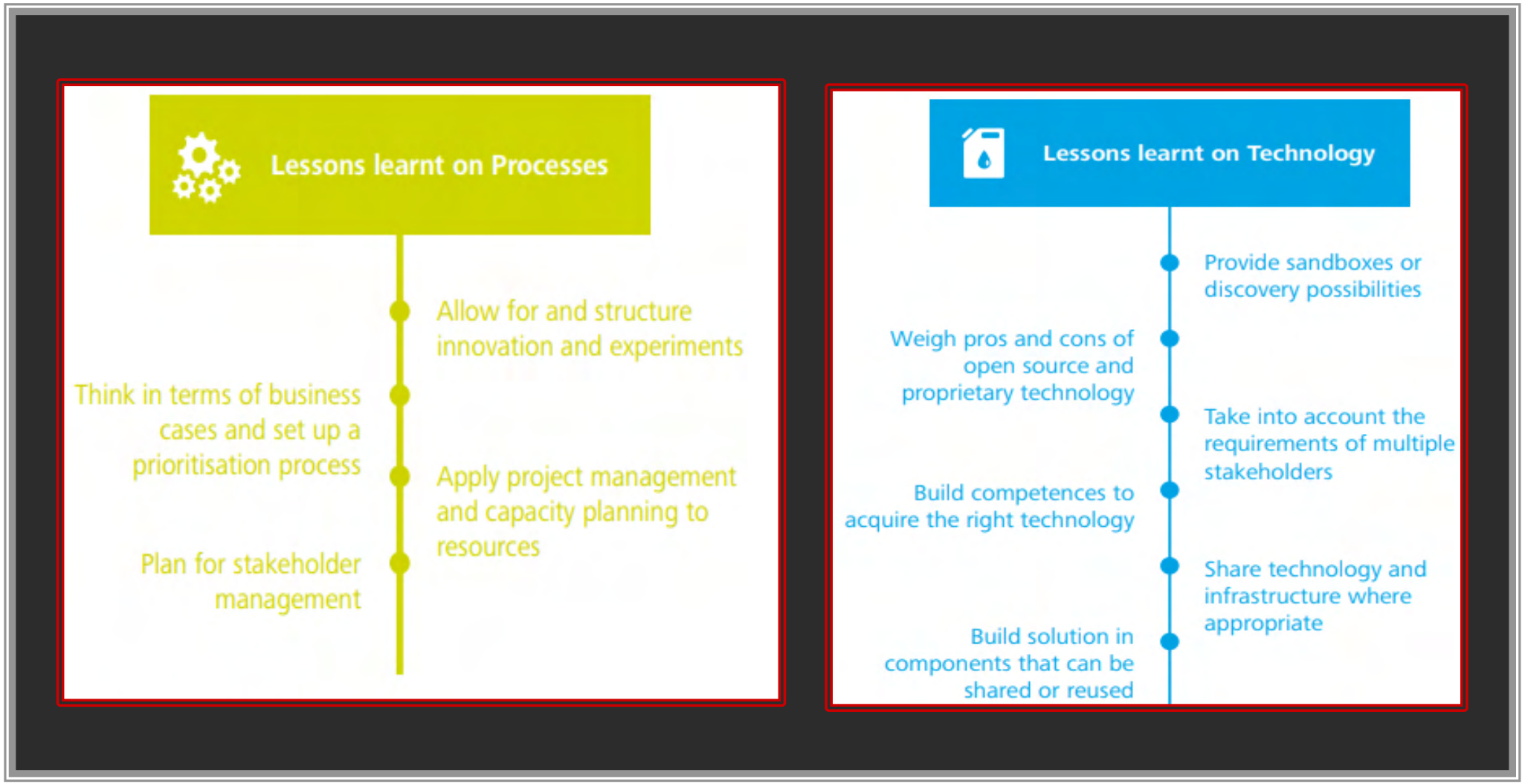
Data Governance: A framework Approach



Source: 'Big data and analytics for policy making' (EU, 2016: p53-59)

(Dizon, et al. 2017)

Data Governance: A framework Approach



Source: 'Big data and analytics for policy making' (EU, 2016: p53-59)

(Dizon, et al. 2017)

Data Governance pitfalls

Fail to ……

- Take **an organizational approach**
- Develop **a clear chain of accountability**
- Consider **interoperability**
- Make **use of legal and regulatory tools**

Big Data & Applications

- Big data and SDGs
- Challenge and opportunities
- Policy Recommendations
- Data-driven technology

1 NO POVERTY
Spending patterns on mobile phone services can provide proxy indicators of income levels

2 ZERO HUNGER
Crowdsourcing or tracking of food prices listed online can help monitor food security in near real-time

3 GOOD HEALTH AND WELL-BEING
Mapping the movement of mobile phone users can help predict the spread of infectious diseases

4 QUALITY EDUCATION
Citizen reporting can reveal reasons for student drop-out rates

5 GENDER EQUALITY
Analysis of financial transactions can reveal the spending patterns and different impacts of economic shocks on men and women

6 CLEAN WATER AND SANITATION
Sensors connected to water pumps can track access to clean water

7 AFFORDABLE AND CLEAN ENERGY
Smart metering allows utility companies to increase or restrict the flow of electricity, gas or water to reduce waste and ensure adequate supply at peak periods

8 DECENT WORK AND ECONOMIC GROWTH
Patterns in global postal traffic can provide indicators such as economic growth, remittances, trade and GDP

9 INDUSTRY, INNOVATION AND INFRASTRUCTURE
Data from GPS devices can be used for traffic control and to improve public transport

10 REDUCED INEQUALITY
Speech-to-text analytics on local radio content can reveal discrimination concerns and support policy response

11 SUSTAINABLE CITIES AND COMMUNITIES
Satellite remote sensing can track encroachment on public land or spaces such as parks and forests

12 RESPONSIBLE CONSUMPTION AND PRODUCTION
Online search patterns or e-commerce transactions can reveal the pace of transition to energy efficient products

13 CLIMATE ACTION
Combining satellite imagery, crowd-sourced witness accounts and open data can help track deforestation

14 LIFE BELOW WATER
Maritime vessel tracking data can reveal illegal, unregulated and unreported fishing activities

15 LIFE ON LAND
Social media monitoring can support disaster management with real-time information on victim location, effects and strength of forest fires or haze

16 PEACE, JUSTICE AND STRONG INSTITUTIONS
Sentiment analysis of social media can reveal public opinion on effective governance, public service delivery or human rights

17 PARTNERSHIPS FOR THE GOALS
Partnerships to enable the combining of statistics, mobile and internet data can provide a better and real-time understanding of today's hyper-connected world

[source: survey 2018 P186 Box 8.3.]

Let’s take a break!

Are you up for a challenge?

Brainstorming

Think of some examples that big data may help in following SDGs

Goal 4: Quality Education

Goal 1: Climate Change

BIG DATA: Challenges and Opportunities

Opportunities

- Rich international cooperation opportunities – exchange knowledges, information, experiences

('International conferences on Big data for health')

- Strong commitments to big data application (especially in health sector) from the government and stakeholders

Challenges

External

- Increasing data volume, velocity and variety..(data' s complicated nature)
- Expensive facilities and hardware equipment
- Cybersecurity issues

Internal

- Asking for high capacity in big data analytics, management and governance to ensure data quality, quality of analysis and security

[bangladesh-big data 2.18]

Case Study on Big Data Innovation

The purpose of CBS

Statistics Netherlands (CBS) is responsible for collecting and processing data in order to publish statistics to be used in practice, by policymakers and for scientific research. In addition to its responsibility for (official) national statistics, Statistics Netherlands also has the task of producing European (community) statistics[42]. CBS also participates in other supranational collaborative experiences such as the UNECE Sandbox.

The mission of Statistics Netherlands is to publish reliable and coherent statistical information that meets the needs of society. In view of this mission, the quality of the statistical information must be guaranteed[43].

The information Statistics Netherlands publishes incorporates a multitude of societal aspects, from macro-economic indicators such as economic growth and consumer prices, to the incomes of individual people and households[44]. Since 2004, Statistics Netherlands is an autonomous agency with legal personality.

Statistics Netherlands' approach to innovation and big data

In January 2012, Statistics Netherlands formally started its Innovation programme in order to accelerate innovation and thus facilitate dealing with trends such as, the high volatility of information and the increasing need for rapid, to-the-point and easily accessible information, the shift to mobile devices and the increasing importance of internet[45].

Their innovation program is loosely organised in five priority areas: Data Collection Innovation, Efficient Processes, Output Innovation, Big Data, and IT Innovation.

In the last years CBS has been experimenting with the following Big data sources:

- Mobile phone data for mobility;
- Scanner data for consumer price index;
- Social Media (twitter) for sentiment analysis;
- Traffic loops data for traffic intensity statistics;

In order to prioritize well and focus on value, they have developed a phased approach. The innovation program foresees three phases: idea generation, proof of concept to test the value and implementation if previous processes have given enough proof of value.

Through this innovation program CBS gained a lot of experience in experimenting with alternative (big data) data sources and in assessing the value of these for statistical production purposes.

Source: 'Big data and analytics for policy making' (EU, 2016: p53-59)

Case Study on big data in tax administration

The purpose of the Estonian tax and customs board

EMTA (Eesti Maksu- ja tolliametile) is the Estonian Tax and Customs Board. The area of activity of the Estonian Tax and Customs Board includes administration of state revenues, implementation of national taxation and customs policies and protection of the society and legal economic activities[54]. The journey towards data analytics started back in 2004 on Management requests. Starting with a pilot, now they optimised their work through data analytics and achieved significant results in terms of costs reduction. The use of data analytics within EMTA is now also driven by the Estonian Government strategy on this matter which promotes further and further big data take up across administrations.

At European level, EMTA is collaborating on customs subjects with the European Commission within the framework of customs working groups, one of them being the one created by the Lithuanian Customs Authority on data analytics.

The Estonian customs analytics project

EMTA uses big data and data analytics technology for fraud detection and evaluation purposes. Through data analytics they redefined their strategy towards identification of cases to verify. They moved from an unstructured approach to this case selection towards a data based methods driven by an algorithm identifying risk coefficient for each case with the overall objective of increasing tax compliance and prevent fraud. For this purpose EMTA analyses a large amount of structured data coming from government sources mainly such as business register and tax declarations.

EMTA management is nowadays also taking decisions on the organisation's activities based on big data mainly.

Case Study on Integrated Approach to AI

Box 8.5. Europe rolls out an integrated approach to Artificial Intelligence

In April 2018, the European Union chose to pool its resources to foster innovation through the use of artificial intelligence. The Declaration[51] signed by European countries aims to ensure a sustainable vision for AI to thrive, by collectively addressing ethical and societal challenges linked to its growing and pervasive use. This states "where needed [to] review and modernise national policies to ensure that the opportunities arising from AI are seized and the emerging challenges are addressed." The European approach is based on three pillars.[52] The first foresees an increase in financial support, to reach 20 billion Euros by 2020, thereby promoting the uptake of AI in both the public and the private sector. The second pillar is based on ensuring framework conditions for socio-economic success. Actions here aim at accompanying the transition of the labour market by modernizing education and training. The third pillar addresses the development of an adequate ethical and legal framework. The first series of draft guidelines is expected by the end of 2018 and will build upon the Union's Charter of Fundamental Rights[53].

Source:http://ec.europa.eu

[Source: UN E-Government Survey 2018]

Conclusion

Contextualisation is a key

Contextualize using data for public goods

- ✓specify subjects
- ✓specify goals

for better policymaking (e.g. formulating a more equal, inclusive, preventative healthcare strategies by using big data to monitor population health trends)

Congratulations!

You have reached the end of submodule 2.4 on Data.
Thank you for joining us in this exciting journey.

Under this submodule, you:

- ✓ Understand the definition and characteristics of open data
- ✓ Understand the framework of data governance
- ✓ Understand data-driven technologies and applications

You may proceed to the next submodule 2.5 on E-Participation

Sources & Recommended Reading
Please check the PDF in the folder

Contact us for inquiries or questions

DPIDG@un.org

Or post your questions/comments in the forum!

Please note that this is a beta version. We appreciate your feedback so we can further improve our toolkit

Acknowledgement

The toolkit DiGIT4SD (beta version) was developed under the general guidance of Juwang Zhu and Vincenzo Aquaro. The conceptual framework and overall content development of the toolkit was guided and facilitated by Wai Min Kwok and Olivia Lin. This submodule was developed by Chenzhi Han and reviewed by Wai Min Kwok. Substantive contributions were made by Weiyu Wang.

United Nations Department of Economic and Social Affairs
Division for Public Institutions and Digital Government