



Module #5 - Data: uses, opportunities and risks

Digital Transformation Learning Modules

Time	Slide #	Script (text and actions)
Data: uses, opportunities and risks (240:00:00)		
Introduction (02:30)		
00:15	1 	Share screen. Hello and welcome everyone. Thank you for coming along to the fifth session of a series of seven modules on digital government.
01:00	2 	This is meant to be an interactive session. We'll pause regularly for activities and discussions. But do not hesitate to interrupt us anytime for questions or comments. You can do this either by raising your virtual hand, or by using the chat. Unless you have connectivity issues, I'll ask you to keep your video on. But please stay on



		mute unless you're speaking.
00:05	3 	Today we will be talking about data: its uses, the opportunities it creates, and the risks associated with this.
00:30	4 	<p>At the end of this training session, you should be able to:</p> <ul style="list-style-type: none"> • Understand how organisations can create value from data, from building better services to policy making and evaluation • Understand the common barriers and enablers of data use in government • Explain the 'once-only' principle • Understand the importance of using data responsibly
00:40	5 	<p>The structure of today's session reflects the learning objectives. We'll cover these five main questions:</p> <ol style="list-style-type: none"> 1. How can we create value with data? 2. What are the benefits of open data? 3. Where to start to improve the use of data in government? 4. What is the once-only principle? 5. Why is it important to use data responsibly? <p>Any questions before we start?</p>



1. Creating public value with data (42:30)		
a. Introduction to data and its importance (27:30)		
00:05	6 	Let's start then!
05:00	7 	Group discussion How do you define data? Note: This could be a sli.do activity
01:00	8 	There are ways to define data. In simple terms, data is a collection of facts. It can be quantitative or qualitative. Quantitative data is data that can be counted or measured and given a value in numbers. Qualitative data is descriptive and usually expressed in language. But both are actually data and are important. Examples of quantitative data are age, salary, weight, height. Examples of qualitative data are doctor's notes, legal cases judgements or academic research reports.
02:00	9	In the early 2000s, tech giants such as Facebook realised how digital platforms and the 24-7 connected citizen provided the ideal context to collect and reuse data for business purposes. This opened a window of



		<p>opportunity to start selling data-driven products and services to any company and individual with an interest in designing ad hoc marketing and communication strategies – from businesses to politicians. Data collected through multiple sources (from mobile phones to smart home devices) are now analysed to better understand users and target potential customers, or service users. These insights are used to drive citizens’ choices, increase business revenues, and even influence public vote.</p> <p>It’s in this context that The Economist published in 2017 an article called “The world’s most valuable resource is no longer oil, but data” which soon became a new buzz phrase. With this data-oil analogy they wanted to increase public awareness in response to raising data monopolies and controlled data flows, but also stress how data could help organisations to make better decisions. We’ll see later in this session why this oil analogy is not as relevant as we may think.</p>
00:30	<p>10</p> 	<p>There is nothing new about governments and businesses collecting and using data. What has changed in the internet era is the volume of data that is now available. 90% of the world’s data has been created in the last 2 years. The amount of data available brings new perspectives.</p>
01:00	<p>11</p> 	<p>Why has there been such an increase in the amount of data generated? More data is generated through the increased use of connected devices and online services.</p> <p>As at January 2021, 53.6% of the world’s population were active social media users. The Internet of Things, or IoT, which refers to the billions of physical devices around the world that are now connected to the internet, is expected to generate 73 zettabytes of data by 2025. One zettabyte is approximately equal to a thousand exabytes, a billion terabytes, or a trillion gigabytes. To place that amount of volume in more practical terms, an</p>



		<p>exabyte alone has the capacity to hold over 36,000 years worth of HD quality video.</p> <p>Sources: Digital 2021 Global Overview Report, https://datareportal.com/reports/digital-2021-global-overview-report https://www.idc.com/getdoc.jsp?containerId=prAP46737220#:~:text=IDC%20predicts%20that%20by%202025,from%2018.3%20ZB%20in%202019)</p>
01:00	<p>12</p> 	<p>Why does data matter? Data enables better decision-making. Making decisions without having access to data is like trying to guess what a puzzle represents with lots of missing pieces, like in this picture. It's really hard! You can guess from the available pieces that the puzzle involves a child but can you tell what's the whole story. If decision-makers only see part of the problem or access part of the data, they might make the wrong decision or resolve only part of the problem.</p>
10:00	<p>13</p> 	<p>Group discussion</p> <ul style="list-style-type: none"> • What makes data initiatives fail or gain traction? • What are the biggest barriers to using data in government?
02:30	<p>14</p> 	<p>Maybe you mentioned:</p> <ol style="list-style-type: none"> 1. Poor data quality. Data enables decision-making. Poor data quality leads to poor decision-making. And when the same data is used in different systems, by different people, poor data quality can become very problematic.



		<p>2. Lack of interoperability. Interoperability is the property that allows data to be easily reused and processed in different applications, allowing different information systems to work together. There are three types of data interoperability: (1) foundational interoperability, which enables different systems to communicate (2) structural interoperability, which refers to the format in which data is shared and (3) semantic interoperability which refers to data having a shared and unambiguous meaning. A lack of interoperability can make it hard to gather data from different sources together. We'll talk about interoperability in more detail in future slides.</p> <p>3. Data protection is another aspect which is often overlooked. When we talk about data protection, we talk about security but also data privacy and the responsible use of data. Governments have a duty not only to ensure that users' data is safe, but also that's is used in accordance with human rights laws, and in an ethical way.</p> <p>4. Finally, some people think that working with data is the job of data experts, like data scientists. But data is everyone's concern. Everyone plays a part in building a data-driven organisation, from data collection to data security.</p>
01:45	15 	<p>This is an example of how things can go bad when data protection fails. In 2020, Public Health England, the UK government agency in charge of protecting and improving health in the country, lost nearly 16,000 records of covid-19 cases due to a poor use of Microsoft Excel. What happened?</p> <p>In the UK, private firms were involved in analysing covid-19 swab tests. They filed their results in CSV files.</p>



		<p>Public Health England had an automatic process that pulled all these different files together in a single Excel file, and uploaded them into a central system so that the National Health Service ‘Test and Trace’ team could access it. The issue was that the Excel file could only support 65,000 rows of data. So data coming in after the 65,000 limit was not captured. The impact of this simple problem was that thousands of people were unaware that they were positive to covid-19, and potentially kept spreading the virus.</p> <p>https://www.bbc.com/news/technology-54423988</p>
00:05	<p>16</p> 	<p>Now let’s see what good looks like. We have seen that data can enable better decision-making. When used properly, data can have a positive impact and even help to save lives.</p>
01:00	<p>17</p> 	<p>Here’s an example of how the government of Sierra Leone used data to fight the spread of Ebola. Their approach was later replicated in other West African countries.</p> <p>In 2016, the Infectious Diseases Data Observatory (IDDO) put together an online platform dedicated to Ebola, in close collaboration with countries impacted by the virus, including Sierra Leone. The platform was the first global data repository for clinical, epidemiological, and laboratory data on Ebola. The IDDO was able to collect and upload to the platform data from various countries including Guinea and Liberia. The amount of data that was gathered enabled better investigation on treatment outcomes, allowing the analysis of patients’ groups who would have been underrepresented otherwise, like pregnant women or children. The outcomes of the research have helped to inform health responses in countries impacted by Ebola.</p>

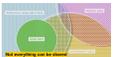


		Sources: https://tdr.who.int/newsroom/news/item/13-07-2020-sharing-research-data-to-better-understand-and-tackle-ebola https://www.iddo.org/research-themes/ebola
b. Who uses data? (15:00)		
03:00	18 	<p>We've talked about how the use of data can create value. Let's now look at different categories of data users:</p> <p>Governments, because they offer public services, collect loads of data. If analysed and used properly, this data can support evidence-based policy making. Evidence-based policy making consists in using evidence (data) to design policy interventions that respond to existing or anticipated challenges.</p> <p>This can apply to a variety of areas, like:</p> <ul style="list-style-type: none"> ● looking at past flood data to inform real estate development ● changing the assessments for children to promote coursework rather than exams ● creating marine protected areas (no-take zones) based on fish spawning data etc. <p>Lots of public organisations have now published guidelines to help civil servants adopt evidence-based thinking when developing new policies.</p> <p>Civil society can use data to innovate, and conduct research (eg. students, researchers, journalists). Access to data also encourages greater citizen participation. People can better hold governments accountable, for</p>



		<p>example by monitoring public spending and its associated impact.</p> <p>Businesses can use data to better understand their market, and innovate. Here's an example.</p>
02:00	<p>19</p> 	<p>With people working remotely during the pandemic, having a reliable power supply was extremely important. Malaysia's single electricity supplier, Tenaga Nasional Berhad (TNB), used data to identify and monitor power issues. Thanks to the analysis of past electricity consumption data, TNB managed to forecast demand and plan for electricity to be dispatched accordingly. TNB readjusted the amount of electricity dispatched to match real-time demand. This helped to eliminate energy waste and reduce costs.</p> <p>The utility provider also collects data from calls reporting power disruptions and shares it with its maintenance crews. This way the causes of power disruption are identified and fixed more quickly.</p> <p>In the future, TNB plans to install smart metres across the country for predictive maintenance purposes.</p> <p>More reading: https://govinsider.asia/digital-gov/how-data-is-revolutionising-malaysias-electricity-sector-tenaga-nasional-berhad/</p>
10:00	<p>20</p> 	<p>Group discussion</p> <p>Give an example of how data that can be useful for (1) government organisations, (2) civil society and (3) private companies.</p>



Break (05:00)		
2. The value of open data (39:00)		
a. What is open data? (08:00)		
00:05	21 	Let's move to the second item on our agenda: how open data can help create value, for public sector organisations but also civil society and private companies.
04:00	22 	Open government data refers to the information collected, produced or acquired by public bodies and made freely available for re-use for any purpose. Let's watch a short video about this. Click on <u>https://vimeo.com/266308637</u>
01:00	23 	Open data can come from governments - here government data - or other sources - here publically available data. But not all government data can be released publicly as open data, as we have seen in our previous module on 'building trust in digital government'. Some data needs to be kept private. Open data is a subset of what can be made publicly available both from government and other sources as well.



02:00	<p>24</p> 	<p>Data is considered open if anyone can access, use and share it. More specifically:</p> <ul style="list-style-type: none"> • For data to be open, it should have no limitations that prevent it from being accessed, used and shared by anyone - whoever they are and whatever their occupation is. • Open data is in the large majority of cases free to use, but this does not mean that it's free to publish open data. There is often a cost to creating, maintaining and publishing usable data. In some instances, this cost can be translated into an access fee for users, but it should never be more than the reasonable reproduction cost of the unit of data that is requested. • Open data can be re-used for any purposes, even commercial ones.
01:00	<p>25</p> 	<p>One frequent question that comes up within governments is why do we need to release open data when we are already publishing statistics for example? Or why do we need to change the way in which we are publishing our data?</p> <ul style="list-style-type: none"> • Open data must be shared in a machine-readable format so that it's easily re-usable. If data is available in a machine-readable format, it can easily flow from one system to another through what we call APIs: application programming interfaces. • Open data must be licensed for reuse.



b. How can open data create value? (31:00)		
00:30	26 	By sharing data in the open, governments make a resource available to others to create value, either in terms of economic, governance or social benefits.
00:30	27 	Economic: organisations can use open data to develop innovative services, and explore new business models.
01:00	28 	Europe did an exercise to estimate the benefits of open data. They estimated that open data in Europe had enabled the creation of 25,000 jobs by 2020, and more than EUR 30m savings in public administration in 11 countries. Source: https://data.europa.eu/sites/default/files/analytical_report_n9_economic_benefits_of_open_data.pdf
01:00	29 	China launched the Shanghai Data Exchange in November 2021 to help businesses in China access data from other organisations easily and transparently, so they can use it to get loans, gain insights or build new tools. One of the key goals of the data exchange is to encourage the opening up of China's digital economy. The exchange will eventually become a hub which can allocate data resources in the market beyond buying and selling. China believes that data is able to improve the efficiency of other resources, such as land and labour, to boost economic growth.



		<p>Source: https://govinsider.asia/digital-gov/yong-lu-exclusive-how-chinas-open-data-trading-could-power-growth/</p>
03:00	<p>30</p> 	<p>Governance: open data can help increase government transparency and accountability by shedding light on their activities, decisions, and expenditures, allowing citizens and government to better monitor the flow and use of public money within and across borders. Open data can facilitate this by:</p> <ul style="list-style-type: none"> • showing how and where public money is spent, which provides strong incentives for governments to demonstrate that they are using public money effectively; • making existing information easier to analyse, process and combine, allowing for a new level of public scrutiny; <p>For example in procurement. It's one of the largest government spending activities, often considered by societies in many countries around the world as an area prone to corruption. Signs of collusive behaviour can be detected by analysing price- related variables like bid distribution characteristics; specific bidding patterns like bid rotation or bid suppression; or market structure-related variables such as market concentration. Constructing co-bidding networks of public procurement bidders allows for differentiating healthy competition from potentially collusive bidding. Co-bidding clusters where most firms bid with all the others and many different firms win contracts suggests healthy competition on the face of it. Instead, a firm winning many contracts while it bids with companies which always lose and only bid with this firm suggests a cartel formation.</p>



		<p>Open Contracting has actually become a new global norm and is endorsed by organisations like the OECD, the World Bank and the European Commission. The Open Contracting Partnership is a non-profit public charity that advocates for Open Public Contracting and supports the implementation of contracting reforms to adopt open contracting principles.</p> <p>Source: https://www.open-contracting.org/about/</p>
01:00	<p>31</p> 	<p>In 2020 in Thailand, the cost of corruption was estimated to be USD 3.31 billion. By publishing data on large infrastructure projects in the open, Thailand was able to increase efficiency in infrastructure projects. The Ministry of Finance reported USD 460 million savings in September 2020.</p> <p>Source: https://govinsider.asia/smart-gov/comptroller-generals-department-pattaraporn-vorasaph-how-thailand-is-cutting-infrastructure-costs-with-transparency/</p>
01:00	<p>32</p> 	<p>Social: there are many ways in which open data can create value for people. Open data can be used by students, local communities, NGOs, technologists, etc.</p>
01:30	<p>33</p> 	<p>In Uruguay, open data is helping people choose their health care service provider. In Uruguay, patients are allowed to change their healthcare service provider every three years, except under exceptional circumstances. Without access to information about healthcare providers, people were unable to make informed choices. In 2015, the Ministry of Health partnered with an open data civil society organisation (DATA)</p>



		<p>to create a website providing easily digestible, searchable and visualised infographics based on open government health data. This web site has since helped people feel more secure and confident in choosing their healthcare providers for them and their families.</p> <p>Source: https://www.ogpstories.org/open-data-empowers-public-to-opt-for-better-healthcare/</p>
01:00	<p>34</p> 	<p>In Taiwan, open data helped assess the availability of masks in the context of the covid-19 pandemic. Personal protective equipment was key in preventing the spread of the virus. In 2020, many countries faced a shortage of face masks. Taiwan National Health Insurance Administration published online data with real-time updates on face mask availability in all contracted pharmacies and selected local health centres. Based on this data, numerous software applications quickly emerged to assist the public in finding sales locations efficiently.</p> <p>Source: https://www.oecd.org/gov/digital-government/open-data-in-action-initiatives-during-the-initial-stage-of-the-covid-19-pandemic.pdf</p>
04:00	<p>35</p> 	<p>By sharing data in the open, governments enable value creation for civil society, businesses, and for public organisations themselves. However, open data is still a relatively new concept, which is not easy to make happen. In this slide we explore some of the main barriers to achieving impact from open data, as well as ways to overcome them.</p> <p>Users don't know about your data: you may have excellent data that could be used to achieve a positive impact, but that is only possible if potential users know that the data exists.</p>



		<p>Governments can promote their open data websites, for example by organising hackathons, where actors from the civil society take up a specific challenge, like ‘how can we use open government data to better protect the environment?’. Hackathons are a nice way to bring people together and think about how data can be used. By offering a cash prize and setting out an open innovation process, you can increase open data’s visibility and encourage people to re-use it for purpose-led services.</p> <p>Showcasing stories of how government data has been used in the past is also a way to inspire others to innovate.</p> <p>Fear of breaching data protection rules: people may not use open data because they’re unsure about what the different licences mean, and if they could be held accountable for unwillingly using personal data without people’s consent.</p> <p>Open data publishers should be clear about the legal underpinning of the data they publish (especially if it includes people’s data), by providing information on the privacy control process and how they’ve anonymised the data, to help to reassure users.</p> <p>Data quality and usability concerns: Another barrier to the reuse of open data is the fear that it may be of poor quality.</p> <p>Some community-based standards exist to help organisations make sure that they publish high-quality open data, and to help users assess the quality of the data they’re interested in.</p>
--	--	--



		Making sure that data is usable, from a legal and practical perspective will make it easier for people to re-use it.
01:30	36 	<p>Government organisations are not the only actors that can release open data. Private organisations - among others - also collect large amounts of data that could be reused. Creating public-private partnerships and involving the private sector in open data initiatives can encourage companies to publish their data in the open.</p> <p>For example, Syngenta is a global agriculture business that publishes open data on crops, from its research on sustainability. Through their open data initiative, Syngenta has helped farmers understand the performances of their crops, and make more sustainable decisions. And by doing this, Syngenta, gets consideration from the agricultural ecosystem.</p> <p>Source: https://www.syngenta.com/en/innovation-agriculture/our-stories/open-datas-potential-agriculture</p>
15:00	37 	<p>Group discussion</p> <ol style="list-style-type: none"> 1. What data does your department organisation collect? 2. Identify at least one data set that you could release in open data. 3. How could they use this data to create public value?

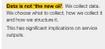


Break (05:00)		
3. Collecting, storing and maintaining data (53:30)		
a. Collecting data (38:00)		
00:05	38 	Our third agenda item for today is collecting, storing and maintaining data.
02:00	39 	<p>The right data is not always available at the right place and at the right time to enable decision making in government. This may be due to various reasons:</p> <ul style="list-style-type: none"> • Data is not always seen as a priority for governments, resulting in a lack, or poor strategic thinking and data governance. In this case, it can be unclear who is accountable for data collection, analysis or maintenance. • It can be hard to secure funding for data initiatives. Investments in data bear fruit in the long run. In the short term, significant investments in time, money and people are necessary to build and maintain interoperable databases. Yet this is the foundation for successful digital transformation and the development of AI-based solutions.



05:00	40 	<p>Data goes through a lifecycle from raw, isolated and unstructured datasets to information and knowledge on which governments can make decisions. Understanding the data lifecycle can help digital teams and governments to maximise the value of data. The diagram is presented as a cycle rather than in a linear fashion, because data can inform and affect the nature of decision-making processes, which in turn can lead to the production and collection of different or more data.</p> <ol style="list-style-type: none">1. Collecting and generating: The data accessed by public servants can come from multiple sources. Internal data sets may include service usage data, or data from enquiries, for example data requested as part of the design of a service. While much of that data is generated by government activity, it is also possible that this first stage in the cycle involves non-governmental sources. For example, data held by private sector actors working in conjunction with the public sector to deliver goods and services.2. Storing, securing, processing: Once governments have access to data, they need to store it. We've seen in the previous session how important it is to store data securely. It's also important to assess the quality of data, and cleanse it. Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. Most government data need to be aggregated or manipulated in order to support decision making. Many users would in fact not be able to understand and make use of the data 'as it is'.3. Sharing, curating, and publishing: data needs to be distributed to the potential users within and outside the public sector to enable access and re-use. If government data is made public, then any actor can generate value from it, as we'll see later in this session.
-------	---	--



		<p>4. Use and reuse: A data-driven government can generate public value by using data to plan, to deliver better services to citizens as well as to get insights for evaluating policy and monitoring performance of public institutions. Public value can also be generated by sharing data in the open as we have discussed earlier in this session.</p>
03:00	<p>41</p> 	<p>In modules 2 and 3, we talked about the importance of building human-centred services that anyone can use. The same applies to data services. Data is a means to an end, it's here to inform decision making or build services that answer user needs.</p> <p>Teams that use data to build services need to take a human-centred approach from the start, and that includes the data collection stage. Before collecting data, digital teams need to ask themselves what problem they're trying to solve. Because the way data is collected influences the outputs of data services.</p> <p>Do you remember our earlier analogy: 'data is the new oil'? It's not exactly true. Oil is not something we - humans - create. It's something that's naturally formed from organic elements across millions of years. We find it, we extract it, we burn it. It's not the same with data, we - humans - create data. We either do this manually, like when we fill out forms, or we ask machines to collect data for us, like traffic sensors. But even then, we - humans - decide what data we want to collect (and not collect), and we decide how to structure and categorise it. This has significant implications on service outputs.</p>
01:00	42	<p>A few years ago, there was lots of debate about how airport scanners were discriminating against coloured minorities. These scanners used facial recognition algorithms to verify people crossing borders were actually</p>



		<p>who their passport said they were. But the algorithms had been trained on photos of mostly white people. So the algorithms were not as precise in recognising coloured people, and some of them got stuck at the airport. From the start, the service team should have thought, who will use the airport scanners? And from there, they should have fed computers with pictures of both white and coloured people.</p>
02:00	<p>43</p> 	<p>Data collection is key. Let's say your government wants to help reduce school dropout rates. What's the best way to do this? Do you go to schools and ask dropout rates? No. To reduce school dropout rates, you first need to understand why students decide to stop going to school. You can think of some reasons: maybe some students live too far from schools, maybe some students have to help their parents make a living, maybe it's about teenage pregnancy. So you don't need - just - dropout rates. You need drop out rates per region, per school, per school year. You need to know who's dropping out, by looking at sex, gender, parents' occupation, earnings.</p> <p>It is difficult for people without data literacy to think about all these dimensions. That's why data literacy is important. People in government need to be able, willing and confident to engage with data, use it in their day to day work and communicate data.</p>
00:30	<p>44</p> 	<p>Group discussion</p> <p>Let us work through an example. With the covid-19 pandemic, people's lives changed drastically, between lockdowns and other mobility restrictions. Let's say that your government wanted to promote e-commerce as a way to ensure safer shopping for people. How could data have been helpful?</p>



Note: For this activity, it is suggested that the trainer moves to a Miro board if the training is done virtually.		
00:30	45 	<p>We should first start with the problem we are trying to solve. What are we trying to achieve?</p> <ul style="list-style-type: none">● Allow citizens to shop for what they need● Ensure safety
02:00	46 	<p>Let's put ourselves in the shoes of our users. What goods or services are essential for them to buy?</p> <ul style="list-style-type: none">● Groceries● Pharmaceutical products● ... participants to list
03:00	47 	<p>What data may we want to collect to assess whether e-commerce is a viable option?</p> <ul style="list-style-type: none">● Does the shop already have a delivery option?● Are there any special requirements for delivery, e.g. frozen goods?● Does the shop have a digital way of marketing their goods, e.g. web site, social media, etc? Small entrepreneurs may not have an online presence.● Does the shop allow for online payment?● ...participants to list
02:00	48	<p>Now let's try to see which of these data are already available and what data we need to collect.</p>



02:00	49 	<p>Now let's think about users. What data do we need about users to assess whether e-commerce answers user needs, and is an inclusive option? Who may be left out? Senior people with little digital skills, people in rural areas without internet connection, etc.</p> <ul style="list-style-type: none"> • Do people have access to the internet? • Are people digitally literate? • Can people use online payment? • ...participants to list
01:00	50 	<p>When implementing digital services, governments must aim to leave no one behind. While there has been a lot of progress towards gender parity in internet use according to the ITU 2021 survey, recent research shows that women are less likely than men to have reliable access to the internet and are less likely to have the necessary digital skills to use the internet. To use digital to achieve more gender equality, governments must ensure that policies and solutions address these challenges.</p> <p>Sources: https://www.itu.int/en/ITU-D/Statistics/Documents/facts/FactsFigures2021.pdf https://femlab.co/2020/11/14/beyond-access-towards-meaningful-connectivity/</p>
10:00	51	Group discussion



		<ol style="list-style-type: none"> 1. From your experience during the pandemic, can you think of digital services that have led to some people, (e.g. women) to be excluded? 2. What could be some policy considerations to ensure greater access to digital services?
b. Data infrastructure and architecture (12:00)		
01:30	52 	<p>We'll now talk about data infrastructure and architecture. Data infrastructure describes the set of components that make up a system, while data architecture describes the design of the components and their relationships. In a nutshell, a system is built on an infrastructure that has a particular architecture. We won't go into deeper technical details than this.</p> <p>Data infrastructure and architecture are key because they support the publication, sharing and reuse of data. They can allow or prevent smooth data-sharing across institutions. The lack of an overarching data governance model can lead to the proliferation or duplication of data standards and technical solutions for data sharing, thus hindering data interoperability.</p>
00:30	53 	<p>Data is said to be interoperable when it can be easily reused and processed in different applications, allowing different information systems to work together. Data interoperability helps to make the lives of civil servants easier but also the lives of citizens.</p>
05:00	54, 55	Group discussion



		<p>Suppose we want to keep track of who attended this training programme, and that we captured participants' details in a different format at each session. At the end of session 1, we received hand-written information and then scanned it as PDF. In session 2, we gathered details in a spreadsheet. In session 3 in a word document. How easy would it be to gather all this data in one single place?</p> <p>Now suppose that on the participant registration form, there is a field 'status'. Some of the participants understood the status to be marital status whereas some of them understood it as employment status. What consequences does this lead to?</p>
02:00	56 	<p>The aim of this example was to introduce the 4 types of interoperability:</p> <ul style="list-style-type: none"> ● Foundational: different systems and platforms are able to talk to each other. ● Structural: the format in which you store data is important. When you start sharing data with others, you need to be able to share it in a format that others can work with. ● Semantic: it's important to ensure that different people interpret the same data in the same way. ● Human: even if data is interoperable, there won't be much happening if humans don't collaborate. This is probably the most important point. <p>Interoperability enables data to be shared between different systems. For data to be interoperable, it must be used consistently. This is why standardisation matters.</p>



00:30	57 	Data standards provide a common way for various departments to collect, store, manage and share data in a consistent way. Data standards make it easier for people to access, use and share data across different organisations or departments.
01:00	58 	<p>What can be standardised?</p> <ul style="list-style-type: none"> • Vocabulary: standardised vocabulary helps to achieve semantic interoperability. One example for standardising vocabulary is registers. Registers are authoritative lists that can be trusted. For example, a list of countries or postal codes. • Data exchange: this refers to file formats, data transfers and rules for sharing data. Simple example is CSV. • Guidance: this is more about processes and best practices. <p>There are also open standards for data that everyone can access, use and share.</p> <p>More reading on: https://standards.theodi.org/introduction/what-are-open-standards-for-data/</p>
00:30	59 	<p>The UK Office for National Statistics has developed data standards that define rules for using and managing data consistently. It includes guidelines of how to classify data, data formats as well as data sharing standards.</p> <p>More reading on: https://www.ons.gov.uk/aboutus/transparencyandgovernance/datastrategy/datastandards</p>



c. Data maintenance (03:30)		
01:30	60 	<p>Data quality is important. This means that data must be kept up-to-date. Government organisations need to think early on about how they can do this - and ideally automate it - and who should be responsible for it.</p> <p>It's worth highlighting here that the EU General Data Protection Regulation makes it mandatory for organisations that store personal data to allow data users to access and update their data.</p>
01:00	61 	<p>Data retention refers to how long organisations need to maintain their database. This has an impact on organisations' data infrastructure needs. Bear in mind that organisations may be required by law to keep records for a specific period of time. Having a data retention policy helps to have a clear and common understanding of all these considerations. As part of that, organisations need to think about how they archive data when it is no longer required.</p>
01:00	62 	<p>The use of data for decision-making purposes increased during the covid-19 pandemic. It highlighted gaps in governments' data governance. The United Nations Department of Economic and Social Affairs (UNDESA) published a report in 2020 called 'Strengthening Data Governance for Effective Use of Open Data and Big Data Analytics for Combating covid-19'. It highlights the importance of taking a whole-of-government approach to data governance. To be successful, data initiatives require a coherent implementation. Having a centralised data agency that steers data-driven policies, strategies and initiatives across the government can help to ensure a coherent implementation and a whole-of-government approach.</p>



		<p>New Zealand has a central agency under Stats NZ that leads data initiatives across government as well as a Government Chief Data Steward since 2017.</p> <p>Sources: https://www.un.org/development/desa/dpad/wp-content/uploads/sites/45/publication/PB_89.pdf https://www.data.govt.nz/about/about-data-govt-nz/ Short video from former Government Chief Data Steward: https://www.youtube.com/watch?v=rGCR51tFnTY</p>
Break (05:00)		
4. The Once-only Principle (28:30)		
00:05	63 	Let's move on to the fourth item of today's session: the once-only principle.
01:00	64 	Data interoperability makes it possible to share data between different systems and organisations. This means that people should not have to waste time submitting the same data to different organisations, but that government organisations should instead speak to each other and share the data they hold. This is what we call the once only principle. It relies on the fact that governments' organisations' databases communicate with each other.
02:00	65	Let's look at the once-only principle from the perspective of a service user. Meet Jane. She just turned 18 and



		<p>goes to the Civil Status Office to request her identity card. To obtain her identity card, Jane will need to submit her personal information such as her name and date of birth which is verified against her birth certificate. Her request is processed and Jane is issued with her national identity card with an ID number. Her records are updated in the Civil Status database.</p> <p>A few months later, Jane wants to apply for her driving licence. She goes to the Police Traffic branch. The Police first needs to verify her identity and she again has to enter her name, date of birth as well as her ID number on her application form. The Police officer will then need to verify that these details are correct by checking them against her ID card.</p> <p>With the once-only principle, information is shared securely between different departments within Government so that information needs not be submitted more than once. This is usually done through a Data Exchange Layer. We will look at an example of Data Exchange Layer in more detail in the next slides.</p>
03:00	66 	<p>Let's see how this scenario changes when applying the once-only principle. When Jane turns 18, she makes a request to the Civil Status Office to get an identity card. As this is the first time she will be issued her ID card, she will need to call in at the Civil Status Office for her identity to be verified.</p> <p>When Jane now wants to apply for her driving licence, the Police also need her personal details and need to verify her identity. With the once-only principle, the Police should be able to access this information already available in the Civil Status Office database. This is typically done through a Data Exchange Layer. Jane submits her ID number to the Police department. Using this ID number, the Police submits a request to the Civil Status Office through the Data Exchange Layer to access Jane's records. The Civil Status Office gets the</p>



		<p>request and publishes the data to the Police department. The Police department verifies Jane’s identity and is able to proceed with her application.</p> <p>Another scenario applying the once-only principle could be that when someone dies, the Civil Status Office automatically notifies the institution responsible for paying pensions benefits to stop the pensions payment for that person. If this had to be done manually there are great chances that it would take time until the payment is stopped.</p> <p>This scenario is a good example of how using the once-only principle can not only improve citizens’ experience but also ensure efficient spending of public funds.</p>
01:00	<p>67</p> <p><small>Benefits of the once-only principle</small></p> <ul style="list-style-type: none"> • Less error-prone • Reduced inefficiencies • Better user experience 	<p>There are several benefits to applying the once-only principle:</p> <ul style="list-style-type: none"> ● Less error-prone: As the same data is entered only once, it’s less error-prone. There’s a single source of truth. If changes are made, all organisations will access the same updated data. ● Reduced inefficiencies: people can access data without asking for csv files from their colleagues. ● Better user experience: users spend less time entering their data
03:00	<p>68</p> <p><small>Common barriers to the once-only principle</small></p> <ul style="list-style-type: none"> • Governance • Interoperability • Data quality • Data ownership and control 	<p>But there are barriers to implementing the once-only principle:</p> <ul style="list-style-type: none"> ● Governance: when introducing the once-only principle, it may not be clear who should control the data. Institutions may be reluctant to collaborate with each other or rely on another organisation or department to collect, store and maintain data they need. Implementing the once-only principle requires buy-in at a strategic level, policy direction and governance to be well-defined to ensure coordination



		<p>between the different parties.</p> <ul style="list-style-type: none"> ● Technology interoperability: there might be legacy systems in place that don't readily support data sharing in an appropriate format. This may pose problems to implementing the once-only principle and upgrading or replacing those legacy systems would require effort and funding. ● Semantic interoperability: the use of data may not be standardised across different departments or systems. In some cases, this can be fixed by implementing standards but not always. For example, the list of countries for the Passport and Immigration department may be different from that of the Civil Status department for valid reasons. <p>What this means is that a big-bang approach to implementing the once-only principle is usually not recommended. It's usually best to take an incremental approach to it. For example, it could be easier to start implementing the once-only principle when planning for new projects rather than legacy systems. Over time, the legacy systems will be replaced at which point the once-only principle can be applied to them. It's also ok to decide to have an exception to the once-only principle if it makes sense (like in our previous example for the 'Countries').</p>
15:00	<p>69</p> 	<p>Group discussion:</p> <ol style="list-style-type: none"> 1. As a user of public services, do you have to submit the same data multiple times to different organisations? 2. What data does your organisation collect? How other departments may use this data?



03:00	<p>70</p> 	<p>Estonia is well-known for its once-only policy. Since 2007, the Public Information Act prohibits the creation of separate databases for the collection of the same data.</p> <p>Estonia uses an open source data exchange platform called X-Road which enables departments to exchange data securely between themselves. Put it in another way, X-Road provides a secure data exchange layer that connects the different organisations of the Government of Estonia. Each organisation is a member of the X-Road ecosystem and can both produce or consume data from other members.</p> <p>Other governments are using X-Road, like Iceland.</p> <p>Source: https://x-road.global/x-road-organizational-model Additional reading: https://x-road.global/iceland-joins-the-nordic-interoperability-league-with-straumurinn</p>
5. Responsible use of data (36:00)		
00:30	<p>71</p> 	<p>To finish this module, let's talk about the responsible use of data.</p>
01:00	<p>72</p> 	<p>Data ethics refers to using data without harming anyone directly or indirectly, even if the distribution of data is lawful.</p> <p>We have discussed in the last module how important it is to build trust in digital services. One important</p>



		component to that is to ensure that data is used in an ethical way.
01:00	73 	<p>Data ethics apply to all stages of the data lifecycle, from the collection of data to the deployment of policies or services informed by data analysis.</p> <ul style="list-style-type: none"> ● Stewarding data – collecting it, and cleaning ● Creating information from that data – data modelling training and testing ● Deciding what to do – governments need to ensure that the policies or services they’re creating are aligned with their organisation’s values and culture
02:00	74 	<p>Stewarding data – collecting it, and cleaning: You may wonder what kind of ethical issues can emerge at this early stage of the data cycle. It’s actually a very sensitive stage, because data sets can be biased. The Amazon team had been building computer programs since 2014 to review job applicants’ resumes with the aim of mechanising the search for top talents. But by 2015, the company realised its new system was not rating candidates for software developer jobs and other technical posts in a gender-neutral way. That is because Amazon’s computer models were trained to vet applicants by observing patterns in resumes submitted to the company over a 10-year period. Most came from men, a reflection of male dominance across the tech industry. In effect, Amazon’s system taught itself that male candidates were preferable. It penalised resumes that included the word “women’s,” as in “women’s chess club captain.” And it downgraded graduates of two all-women’s colleges, according to people familiar with the matter. They did not specify the names of the schools. Eventually, Amazon had to scrap its AI recruiting tool.</p>
01:00	75	Creating information from that data – data modelling training and testing. Remember our previous example



		<p>of airport scanners discriminating against colored people. This is an example of bias from data modelling. Another example: Google Translate tends to be biased towards women when translating from languages with gender-neutral pronouns. The algorithm behind it is more likely to produce translations like “he invests” or “she takes care of the kids” even if the sentence in the original language was gender neutral.</p> <p>Source: https://itrexgroup.com/blog/ai-bias-definition-types-examples-debiasing-strategies/#</p>
01:30	<p>76</p> 	<p>Deciding what to do – governments need to ensure that the policies or services they’re creating are aligned with their organisation’s values and culture, and understand the unintended consequences that could happen as a result of their actions. For example, in the UK, the police have started exploring the use of facial recognition algorithms to identify people wanted by the police. There has been lots of debate in the public scene about whether this is a good thing or not. Studies so far show that citizens are quite supportive of the use of live facial recognition(LFR), but only in some specific cases, for example where and when necessary (e.g. football game in a stadium, when under terrorist threat). Among others, they want to take into consideration the potential impact of the use of LFR by the police on the normalisation of surveillance.</p>
01:00	<p>77</p> 	<p>How can governments make sure they leverage the potential of data in an ethical way?</p> <p>Legislation is one route to ensuring ethical management and use of personal information in both the public and private sectors. But, the increasing focus is on establishing ethical frameworks or guidelines. These provide civil servants with information, resources and approaches to help them achieve ethical practices and decision making.</p>



		<p>The ODI Data Ethics Canvas is a good tool to help anyone who collects, shares or uses data to identify and manage ethical issues.</p> <p>http://theodi.org/wp-content/uploads/2021/07/Data-Ethics-Canvas-English-Colour.pdf</p>
10:00	<p>78</p> 	<p>Group discussion</p> <p>Let's go back to the group discussion we had on the Once-only principle and the different data sets that you collect as civil servants. Select a few of these data sets and try to test them through the Data Ethics Canvas.</p> <p>Note: This will likely be a Miro board activity with the canva template created on Miro</p>
02:00	<p>79</p> 	<p>I'd like to end this session by sharing some good practices to encourage a good use of data:</p> <ul style="list-style-type: none"> • Develop a government-wide data policy that highlights the importance of publishing data in the open, and using data in an ethical way • Make sure that you collect and use data for a specific purpose, tha answer user needs • Be transparent about your data governance. and how data is used in your organisation. • Be aware of data bias. When developing data services, make sure you consider potential bias in data collection. Having a diverse multi-disciplinary team makes it easier to gather different perspectives, and anticipate and mitigate bias against specific groups of people.



		<ul style="list-style-type: none"> • Monitor and evaluate how data is used: does it create value? Is it ethical? Keep iterating on data policies accordingly. <p>Additional reading: Good Practices Principles for Data Ethics in the Public Sector</p>
15:00	80	<p>Group discussion</p> <p>What are the goals of your organisation and how can data help you better achieve these goals? In one or two sentences, articulate a data vision for your organisation.</p>
01:00	81, 82	<p>Takeaways</p> <ul style="list-style-type: none"> • Data enables better decision making. • Open data creates economic, governance and social value • Data can be used to improve policy-making and ensure no one is left behind. • Interoperability allows data sharing between different systems • The once-only principle is to ensure citizens do not need to provide governments with information they have already supplied. • Implementing the once-only principle improves the citizen journey and increases efficiency. • Data ethics apply to all stages of the data lifecycle, from the collection of data to the deployment of policies or services informed by data analysis.



ASIA AND THE PACIFIC
**Regional
Innovation Centre**

**public
digital**

	83 	Next session: Managing digital technology risks
--	---	---