

# Data stewardship

## Report by the CES Task Force

## Contents

1	Introduction.....	3
2	National statistical offices and data stewardship.....	6
2.1	Context .....	6
2.2	What NSOs have to offer.....	7
2.3	Enabling environment .....	11
3	Defining data stewardship and data governance.....	13
3.1	Data governance and stewardship in business and academic literature.....	13
3.2	Definitions of data stewardship and governance, and other related concepts.....	15
3.3	Data steward .....	17
4	Data governance in the context of data stewardship.....	19
4.1	Data Governance Models.....	19
4.1.1	Centralised model .....	19
4.1.2	Distributed model.....	20
4.1.3	Federated Hybrid model .....	20
4.2	Supporting principles of data stewardship.....	22
4.2.1	The UN Fundamental principles of Official Statistics .....	22
4.2.2	Standards.....	23
4.2.3	The FAIR guiding principles of data .....	23
4.2.4	The 4Gs Model of Data Lifecycle Management .....	24
4.2.5	The ‘once only’ principle .....	25
4.2.6	The European Interoperability framework .....	26
5	Responsibilities and skills for data stewardship.....	27
5.1	Involvement in Public Sector data strategies and data management .....	27
5.1.1	Data stewardship in different institutional settings.....	27
5.1.2	NSO role in Data Stewardship in the public sector .....	29
5.1.3	Setting a longer term direction for the system as a whole.....	31
5.1.4	Responsibility for trust and confidence in the system as a whole.....	31
5.1.5	Analytical support for public sector data providers.....	31
5.1.6	Data and information services .....	32
5.1.7	Promoting data quality and standards.....	33
5.2	Possible internal ‘core’ roles/tasks associated with data stewardship.....	34
5.2.1	Generic roles required to implement stewardship and data management in NSOs .....	34
5.2.2	The role of Data Steward.....	35

5.2.3	Other related roles .....	36
5.3	Skills required for Data Stewardship .....	38
5.3.1	Human resources requirements for the data steward role .....	38
5.3.2	Secondments .....	39
5.3.3	Partnerships .....	40
5.4	Services outside of the NSO mandate that could potentially be facilitated by NSO stewardship 40	
5.4.1	Business Intelligence .....	41
5.4.2	External 'data engineering' role .....	41
5.4.3	Operational analytics .....	41
5.4.4	Programme evaluation and related econometrics .....	41
6	How to communicate data stewardship? .....	42
6.1	A complex concept – how to communicate it? .....	42
6.2	Possible messages for communicating data stewardship .....	43
6.3	New products, services and partnerships as elements of data stewardship – seen from communication perspective .....	45
6.4	Conclusion .....	48
	Annex 1. Glossary of data and information stewardship .....	49
	Annex 2 Examples of data governance and data stewardship models in countries .....	59
	Argentina .....	59
	Canada .....	59
	Finland .....	60
	Mexico .....	62
	New Zealand .....	63
	New South Wales (Australia) .....	66
	Switzerland .....	68
	European Union: .....	70
	OECD .....	72
	References .....	74

# 1 Introduction

1. The world where statistical offices operate is increasingly interconnected, dynamic and digitized, bringing along new demands and new opportunities. Data has become a strategic asset and many organizations outside official statistics create statistical output using new data sources, tools and techniques. This has also changed stakeholders' expectations towards the demand for more open and more detailed data that is easily accessible and available faster. The COVID pandemic has further tested our ability to harness quality and timely data in a responsible way.

2. The abundance of data and the possibilities to derive new information by linking these data raise the questions of ethics, privacy protection, data security, accessibility, harmonisation of the concepts, formats and technologies used, knowing what is available, and a long-term view. To respond to that need, the term used is 'data stewardship' denoting the functions and competencies to enable access to and re-use of data for public benefit in a systematic, sustainable, and responsible way<sup>1</sup>.

3. The role of national statistical offices in the new data ecosystem has been widely discussed during the recent years. However, the discussions consider very different issues and there is no common understanding, as yet, what data stewardship is and what the roles of national statistical offices (NSO)<sup>2</sup> in this can or should be. The questions that come up are, for example:

- What is data stewardship? How is it linked with data governance?
- What role NSOs could or should play in data stewardship in the public sector, or more widely in a country's data ecosystem?
- What helps NSOs decide how far to go with their role in data stewardship for the public sector?

4. To move towards a common understanding, the Conference of European Statisticians (CES)<sup>3</sup> set up in 2021 a Task Force to help bring clarity into this area and facilitate the work of NSOs in adapting their role to the changing situation. The UNECE Task Force was mandated to:

- Define 'data stewardship' and related terms;
- Develop a set of possible responsibilities of a data steward and considerations to be taken into account in shaping the NSO role in this area. On this basis, a minimum core set of responsibilities and a maturity model of data stewardship can be developed;
- Develop a toolkit of good practices and tools for implementing the role of data steward;
- Clarify what NSOs can and what they cannot do in this respect, consistent with the Fundamental Principles of Official Statistics.

5. The following countries and organizations are members of the Task Force: Canada, Estonia (Chair), Finland, Hungary, Ireland, Mexico, Netherlands, New Zealand, Poland, UK, Eurostat and OECD. The Task Force members represent statistical offices and other government institutions dealing with data governance. UNECE is providing the Secretariat for the Task Force.

6. The work is a follow-up to the CES and its Bureau discussions in 2019 and 2020 that asked for proposals for practical ways forward to enhance the role of NSOs in the new data ecosystems and in

---

<sup>1</sup> <https://medium.com/data-stewards-network/data-stewardship-re-imagined-capacities-and-competencies-d37a0ebaf0ee>

<sup>2</sup> National Statistics Office (NSO) is a term used by the UN that is equivalent to the term National Statistics Institute (NSI) used by the EU.

<sup>3</sup> The Conference of European Statisticians is a permanent UN body. Its members are the heads of statistical offices of UNECE, OECD and some other countries, altogether about 65 countries.

public data governance. An important background document for the discussion was a paper by a group of countries led by Estonia ([UNECE 2020](#)) that laid out a number of considerations for a way forward for NSOs in data stewardship. Based on their extensive experience with ensuring data quality, protection of privacy, handling data and metadata, statistical offices would be naturally suited to take (at least some) role in the data stewardship in the data ecosystems. The Conference noted that NSOs can position themselves in the new data ecosystem in different ways, with a range of possibilities from minimal change to full data stewardship.

7. The work is planned to be carried out in two stages: first, defining the concepts and looking at possible tasks/responsibilities, and in a later stage adding recommendations. The current version of the report represents a draft outcome of the first stage of the work, explaining the context and enabling environment, defining data stewardship and the associated tasks and responsibilities, and describing the related models and principles. A maturity model and any guidelines or tools are expected to be developed in the second stage.

8. One of the incentives for the work was to develop a common definition of data stewardship that can be used in the context of official statistics and more widely, the public sector. The report is looking at the issue of data stewardship through the lens of statistical offices, identifying what role they could and should play, taking into account their mandates, experience and competencies. At the same time, it looks at data stewardship also from a government-wide perspective as there is no fixed borderline between what can be done by statistical offices or by other agencies. Coordination of official statistics has many aspects common with what governments aim at when establishing data stewardship structures for public data. The report makes use of the ample literature on data stewardship and data governance from business sector and academia but puts this into the context of public sector and official statistics.

9. The report aims to serve as a guide to statistical offices helping to understand what data stewardship is, what are its enabling factors and the associated opportunities and risks. It should help the offices who are just starting out to decide whether they should extend their role as government data stewards, as well as provide inspiration and ideas to those who are already undertaking this role. In addition to informing NSO heads about strategic direction and ways to advocating the role of NSOs, it also aims to function as a reference material to both experts and stakeholders in data stewardship. Public data holders will gain a deeper understanding of the value and role of NSO in the future data ecosystem from this report.

10. The report aims to:

- Describe the wider setting of the data stewardship role and why NSOs should be taking a stronger role in stewardship of public data (Chapter 2)
- Define 'data stewardship' and the related terms (Chapter 3 and Annex 1)
- Describe the current state of data governance models and principles (Chapter 4)
- Describe the responsibilities and skills of data stewards (Chapter 5)
- Discuss ways of communicating the role of NSOs in data stewardship (Chapter 6)
- Demonstrate the different pathways NSOs in different countries have taken to progress their role as data stewards (Annex 2).

11. **The draft report is distributed for consultation to gather as wide feedback as possible to improve the final product and ensure that it would be useful to all statistical offices, and for the public sector more widely.** The Task Force will continue work taking into account the comments. The final report is planned to be prepared by mid-2023.

## COVID-19 Impact

The COVID 19 pandemic shook the world. The health risk and the increase in mortality provoked widely varied responses from governments and societies. There was a serious dislocation of productive activities derived from massive business closures, unemployment increased and public capacities to face this emergency were put to the test.

An urgent need for more and better information, in more timely ways, to make better decisions put pressure on governments and companies. Important government-to-government, government-private, and government-civil society collaboration efforts started to obtain fresh data that would help understand the impact of the pandemic in multiple dimensions. This response was based on scientific and technical capacity, as well as the will to make things happen.

However, regulatory restrictions and the perception of risks have restricted the use of information to very limited purposes. It is necessary to offer stable legal, technical and administrative conditions to ensure proper use of data to allow its reuse so that more information for government and business decision-making can be generated. The capacity and reliability enjoyed by the NSOs opened the possibility of developing collaboration models that facilitate the reuse of information in a safe, confidential and constant way, respecting rigorous ethical principles in the treatment of information.

The COVID pandemic required many NSOs to respond rapidly and accelerate modernisation. There are many examples globally of NSOs providing support in the form of data collection, data standardisation, data hosting and data output to ensure the health, infrastructure and economic needs of citizens and businesses were met during the pandemic. Many NSOs have created data portals and data hubs to ensure that data, information and insights are made available to citizens on the subjects which are most important to them and delivered in as close to real time as possible.

These data must be analysed and treated in such a way that they can be harmonised and/or standardised, if necessary anonymised, so that they can be reused. In this work, NSOs have gained competencies and increased their technical strength to become a reference that gives confidence and supports the generation of statistical information from these sources.

The shared (but varied) experience with the COVID-19 pandemic has demonstrated the importance of well-designed, evidence-based policies as well as the key role of data and statistical information and the national statistical organisations that produce them. The fact is, the activities of NSOs are crucial to the provision of appropriate responses during crises. Good data governance and data stewardship are central prerequisites to effectively meet the extraordinary demands that societies face in a crisis, such as the one presented by COVID.

## 2 National statistical offices and data stewardship

### 2.1 Context

12. With the exponential growth in data generated worldwide (e.g., through Internet of Things (IoT) devices and individuals' interactions with service providers and public agencies) and the arrival of new technologies such as artificial intelligence (AI) and edge computing, governments worldwide are increasingly realising the importance of data. In this context, it becomes increasingly important to set the data stewardship and governance foundations that can enable data access and sharing while protecting the privacy of individuals and ensuring that the data are used ethically.

13. The use of electronic devices through communication networks has allowed companies to build up rich data pools, for example in online commerce and entertainment platforms. These data, combined with other - public or private - data offer enormous potential to discover patterns, infer indicators, develop business models, or make decisions for the common good. Access to these data pools is restricted due to reasons related to competitiveness, regulations on privacy and competition, as well as the implications associated with the costs of obtaining and sharing information. The development of public-private initiatives has made it possible to explore the benefits and challenges of sharing information, but the attention on a case-by-case basis limits the potential public utility to produce regular statistical information for public use.

14. The digitalisation of society and economy has placed data access and sharing at the core of innovation and public trust. The use of administrative registers and experimental techniques allow to obtain data from other sources and for different purposes. These advances now offer alternatives to the statistics that NSOs have traditionally been producing.

15. These developments offer both opportunities and challenges. There are huge possibilities for new types of data services, more timely and granular data, new insights by linking data from different sources and topics. But there are also huge risks: data could be used unethically, the 'digital divide' could become an 'information divide', invasion of privacy, etc.

16. The way public administrations have developed still originate from traditional paper-based information handling. Often government agencies or institutions have their own information collection for their own purposes. This has led to hundreds of data holdings in government institutions that are using different data concepts, formats, and technologies. There is no overview of what kind of information is available, high degree of redundancy, and high cost (or impossibility) of linking these data.

17. To overcome the risks requires governmental policies with a longer-term focus to deal with strategy, culture, ethics, roles and settings, and the capabilities of people that can support an effective data ecosystem. The new digital data governance requires common rules and standards to make interoperability possible. For this to work at an all-of-government level there must be coordination and accepted norms.

18. Data governance establishes and makes accessible the set of norms associated with data. Agreed data governance norms will provide clarity and assurance in the way data is overseen and managed, and support data availability, quality, security, usability, and integrity. While data governance has a strong emphasis on technical competencies, data stewardship adds an additional layer to support the growing maturity of data policy.

19. Evidence from the OECD (OECD, 2019) shows how national governments are taking steps to reduce policy siloes and make data strategies, projects and initiatives in the public sector more

coherent. Countries such as the United States<sup>4</sup>, the United Kingdom<sup>5</sup>, the Netherlands<sup>6</sup>, and Germany<sup>7</sup> have issued National Data Strategies to bring together, under a single policy instrument, different aspects ranging from data access and sharing within the public sector, open data, data for AI, and data ethics in the public sector.

20. In other cases, data-related policy aspects are part of broader digitalisation, digital government and AI strategies. Indeed, data from the 2019 OECD Digital Government Index show that “only 12% of countries have a single dedicated data policy (or strategy), while 82% embed data as part of broader related policies (e.g., digital government or open data)” (OECD, 2020).

21. In terms of institutional set-ups and leadership roles, some countries opted for defining specific Chief Data Officer positions for the whole-of-government (e.g., the 2015 UK CDO appointment, no longer in place<sup>8</sup>), while others have opted for creating bodies such as data councils where data leadership roles at the institutional level discuss priorities and agree on their coordinated implementation (e.g., the U.S. Federal CDO Council<sup>9</sup>).

## 2.2 What NSOs have to offer

22. National Statistical Offices are mandated to produce and disseminate relevant information to understand the evolution of society, the environment, and the economy. This mandate comes from the Fundamental Principles of Official Statistics and is enshrined in Statistical Laws of countries. For decades, NSOs have developed methods and techniques to produce information to meet the mandate. Alongside this, they have promoted a culture of using data and information in society and government for making decisions.

23. The United Nations’ *Fundamental Principles of Official Statistics* provide internationally agreed principles how NSOs can best provide their services to citizens (United Nations 2014). To summarize, the Fundamental Principles state that official statistics must be:

- practical and impartial;
- accurate, trustworthy and ethical;
- scientific and standardized;
- quality, timely, and with as low a response burden as possible;
- confidential and appropriately used;
- transparent and made public.

24. The Fundamental Principles also state that NSOs and other statistical agencies can draw data from a variety of sources, coordinate and cooperate with other organisations (nationally and internationally), and contribute to the improvement of statistical systems (United Nations, 2014).

25. Developed and adopted by the Conference of European Statisticians, the United Nations Fundamental Principles of Official Statistics (FPOS) were adopted in 1992 at the ministerial level by the Economic Commission of Europe (UNECE). These Fundamental Principles were then reaffirmed and

---

<sup>4</sup> See <https://strategy.data.gov/>

<sup>5</sup> See <https://www.gov.uk/guidance/national-data-strategy>

<sup>6</sup> <https://www.nldigitalgovernment.nl/overview/new-technologies-data-and-ethics/data-agenda-government/>

<sup>7</sup> <https://www.bundesregierung.de/breg-en/service/information-material-issued-by-the-federal-government/data-strategy-of-the-federal-german-government-1950612>

<sup>8</sup> See <https://www.techrepublic.com/article/uks-first-chief-data-officer-to-focus-on-making-data-a-public-asset/> and <https://techmonitor.ai/leadership/digital-transformation/uk-government-reverts-to-hire-cto-chief-data-officer-tech-leaders>

<sup>9</sup> See <https://www.cdo.gov/about-us/>



endorsed by the UN Statistical Commission at the global level (1994) and the UN General Assembly (1994). The preamble of FPOS highlights the importance of official statistics for national and global development, emphasizing the vital role that NSOs play in providing high-quality, official statistical data, and the positive effect that this has on policy decision-making and the overall development of countries.

26. The prestige of the NSOs is associated with the technical strength, reliability, and punctuality, as well as the careful handling of the information they collect. The track record of NSOs is a testimony to their experience and competence in working with data, including data and confidentiality protection, data ethics, ensuring comparability, and use of standards and classifications. While NSOs are often seen mainly as producers of statistics, their know-how in coordination, standardisation, confidentiality, handling vast data collections, communication, data literacy, etc. is a big asset for the whole public sector and beyond.

27. National Statistical Offices face the challenge – and the opportunity – to place themselves as key players in the data ecosystem given their well-grounded expertise in data management, access and sharing practices in the public sector, and their key role as producers of statistics and indicators. Figure 1 provides a summary of capabilities that NSOs have which give them a good basis to fulfil the data stewardship role.



*Figure 1. The new role of NSOs in the data stewardship landscape*

28. The following frameworks and expertise of NSOs can be leveraged to provide data, information, insights and visualizations to respond to any emerging need of citizens, businesses and other institutions:

- legislative frameworks and legal authority to collect, process and disseminate data and information (based on the Statistical Law);
- well established data quality frameworks;
- long history of data protection and leading-edge methods to protect privacy and confidentiality;
- modern statistical methods for data analysis, data integration and linkage and data categorization via metadata, and data visualization;
- growing data science expertise grounded in statistical and mathematical theory;
- expertise in data ethics and growing experience in the ethical use of artificial intelligence and machine learning to ensure methods are without bias;
- trust frameworks and development of new frameworks such as proportionality and necessity;
- story-telling and data visualization expertise from data to insights to policy to services for better outcomes;
- access to resources and networks within the national statistical system and a long history of data partnerships and collaboration both domestically and internationally;
- collaborative work spaces (protected cloud, data collaboratives, data trusts);



- high data maturity and high data literacy;
- statistical and data standards and classification systems to give data meaning and drive interoperability;
- subject matter expertise on socio-economic and socio-demographic phenomena;
- data collection expertise (censuses, different types of surveys, use of administrative registers, web-scraping, etc.);
- well established data dissemination vehicles including web presence, data hubs, data hosting services;
- expertise in modelling and forecasting;
- statistical registers – reference data and master data expertise;
- emerging entrepreneurial spirit and culture of continuous learning and development of innovation ecosystems.

29. In summary, the digitalisation of the society offers opportunities for NSOs' renewed role as governmental data stewards and as promoters of institutional data stewardship for example in terms of:

- **User-centricity.** Streamlining data access and sharing within the public sector supports the implementation of the *once-only* principle (the responsibility of public bodies not to ask the same information to citizens and businesses twice). NSOs can play a key role in facilitating and support data access and sharing given their experience in the area.
- **Performance:** Whereas data can be used as a key asset for public service design and delivery, NSOs can collect valuable data on issues such as citizens' satisfaction with those services to measure and evaluate performance and inform action towards continuous improvement.
- **Upskilling:** NSOs can act as hubs of knowledge-sharing and capacity building in areas such as data science and the application of data-intensive technologies.
- **Openness:** NSOs long-standing knowledge of information and data management offer a valuable opportunity to promote and advance the implementation of open data initiatives which could be used to promote data-driven innovation and research.
- **Public communication:** In the era of the fight against dis- and mis-information, it is fundamental to channel users of information and data to trusted data sources. NSOs can play a key role as sources of trusted data to fact-check the information provided by other sources. NSOs core know-how on the production on statistics can also help to build knowledge across public bodies for preparing effective data visualisations and avoiding misleading data representations.

30. NSOs are in a position to provide leadership in data governance and stewardship, as partners in a data ecosystem in charge of co-creating trusted smart statistics. The relevant visions, missions, mandates, and legal responsibilities of NSOs contribute to this unique position. This highlights societal needs, identifies opportunities for growth and development, and considers how NSOs may contribute to this growth and development.

31. Stewardship goes beyond sound data management and standards onto the domains of public acceptability and trust. Social license and public engagement are necessary for NSOs to be able to perform their duties. This is achieved through data stewardship and the adherence to principles of open data, transparent processes, ensuring confidentiality and security, and by communicating the value of citizen's sharing their data.

32. The data stewardship is a continuation of the role that NSOs have played in coordinating the national statistical system. But it has a potential of going deeper within the statistical system to ensure interoperability, and going wider, to the whole data ecosystem.

33. However, how the data stewardship role can be implemented in practice is highly context-dependent, varying based on national factors, such as legal frameworks and NSOs' mandates, maturity and use of digital technology, data ecosystems, and public sentiment. Thus, the extent to which NSOs provide this leadership in data governance and stewardship will vary by countries.

34. NSOs can position themselves in the new data ecosystem in different ways, with a range of possibilities from minimal change to full data stewardship. NSOs are at different starting points, some already have cross-government roles, and there is no "one-size-fits-all" solution. In moving towards data stewardship, it is important to take into account the legal basis, and recognize that NSOs can move forward at their own pace and take on a stewardship role to a different degree.

35. There are **risks** both ways in making and not making the change. If NSOs fail to take the lead in the process of data ecosystem transformation, they risk losing relevance. The outcome could be further fragmentation of public data holdings with little or no compatibility. While extending their role, NSO should maintain their independence and support it with promoting data literacy and data ethics.

36. On the other hand, there is also a risk of taking on a too ambitious data stewardship role that NSOs can not fulfil. This may impact adversely the image and reputation of the organisation. The aspects of cyber security can also not be downplayed - as data stewards, NSOs are more likely to become potential targets for cyber-attacks, and need to be prepared for this.

37. NSOs should consider these risks carefully: whether the damage to reputation may affect the core operations of an NSO, or it would still be better to be seen to be trying and failing, than doing nothing and watching from the side-lines.

38. **When successfully carried out**, the data steward role will increase the influence of NSOs across the public sector. They can be considered as trustworthy partners with data not only statistics, and allow NSOs to fulfil their core responsibility. NSOs will benefit by securing an important role for the future that will provide them with:

- access to more data sources allowing to produce more timely, relevant and disaggregated data
- increased possibilities for integrating data from different sources (especially if there are common identifiers).

39. Indirectly, this will increase efficiency in the whole public data ecosystem. Sound data stewardship has societal benefit and contributes to the public good because it enables ethical operation, which begets the trust, social license, and public support necessary for statistical work (Green, 2021). It minimises data misuse and enables reuse, allowing statisticians to access data already in the ecosystem. It also facilitates data sharing and the use of new, complementary data sources. All of this either directly or indirectly improves public trust and increases engagement, by saving time and money, decreasing response burden, increasing data value, and better enabling the communication of that value to citizens.

40. This may help to shift from a "vicious cycle" of statistical underdevelopment to a "virtuous data cycle". In their Development Co-operation Report, the OECD discussed the role of the national statistical systems in the data revolution, stating that increasing efforts in planning and production, strong data dissemination, and communicating value of data to citizens and partners alike, results in a more productive and virtuous approach (OECD, 2017).

41. In several countries NSOs have already been playing a key role in furthering data stewardship (see also country case studies Annex 2). For instance, in the United Kingdom, the Office for National Statistics' Data Science Campus acts as a space for skill development and data experimentation, and collaboration between the private and public sector (OECD, 2018). In Mexico, the INEGI played a key role in the implementation of open data policy and data skills in the public sector (OECD, 2016). In Canada, Statistics Canada is working closely with two central governmental agencies on evergreening the Data Strategy Roadmap for the Federal Public Service (2018), which supports a whole-of-government approach to digital and data issues.

## 2.3 Enabling environment

42. A number of enabling factors and trends are already in place in most countries that facilitate NSOs taking a proactive data stewardship role. Three of such aspects are described below: 1) legal and regulatory frameworks; 2) organisational interoperability; and 3) technological developments.

43. **Legal aspects.** The production of official statistics in countries and the role of NSOs is based on Statistical Law. The Law may or may not give sufficient basis for NSO's data stewardship role. Having a solid legal basis is important, but the absence of legal mandate should not stop from moving in this direction.

44. Regulatory environment is different in every country. What it allows to do and what should be better clarified. The legislation may be also outdated in regard to the technological development of society.

45. The (new) legal arrangement must address the implications of reuse and sharing throughout the data life cycle, from collection to dissemination. It will be necessary to establish the responsibilities of those who guard and manage the information (data stewards) as well as the legal and administrative basis (agreements, etc.) and mechanisms (federated, centralized, platforms, trusts, etc.).

46. The adaptation of the legal framework must be consistent with national data strategies and contemplate a flexible system that allows operational adjustments without the need to frequently modify the laws. That is, it must have a comprehensive conception that recognizes the continuous expansion of the border of the data ecosystem derived from technological development and the evolution of economic systems and public service models; and it must establish the obligations and rights for all actors in the ecosystem.

47. The rules associated with the operationalization of the information reuse mechanisms must be administrative in nature, ideally designed by consensus of the actors (standards, models, etc.). In this way, if it is necessary to adjust them in the presence of new risks or technological changes, their modification can be organized without the need for legislative approval.

48. In general, the legal frameworks of the NSOs allow agreements for research and projects to create statistical indicators using data held by third parties. The technical and legal challenge then focuses on the possibility of having this information regularly available, guaranteeing that its reuse does not legally compromise any of the parties that participate in the generation of the data.

49. It is necessary to maintain and strengthen communication so that the participants in the data ecosystem take ownership and use the Fundamental Principles of Official Statistics as a common framework, as well as the established Codes of Ethics. In this way it will be possible to strengthen trust between the actors under a transparent frame where the incentives to participate and the rules to operate are crystal clear for all.

50. In the European Union, a number of legislative acts have been adopted in recent years that have a bearing on data stewardship. ([To Add] related legislation: the European Data Act; GDPR).

51. **Organisational aspects.** The extent to which data stewardship role can be implemented depends on the maturity of the data ecosystem in the country, including the existence of public sector/service data strategies, national data strategies, etc. When such strategies are developed, NSOs should be actively involved in the discussions. However, this is often not the case.

52. **Technological aspects** have to do with the level of technological development:

- digital services: national and cross-national based on data usage, also cross-border use;
- digitalization of public sector processes (from paper-based forms and documents towards digital forms and documents). This is a very important opportunity for NSO to be involved from the start to ensure that the data that will come out is useful for producing statistics;
- digitalization of economy: production, logistics, trade, selling;
- impact of web technology (semantic interoperability, Internet of Things, data linkage);
- privacy enhancing technologies.

53. Many enablers are specific to countries, related to the legislative basis, institutional setup, status of the statistical office, existence and quality of registers, etc.

### 3 Defining data stewardship and data governance

54. To advance data stewardship and governance, *unified* understanding and definitions of concepts are needed. Currently, grassroots collaboration and international work are yielding definitions and conceptual relationships, as is the work of the academic sector. However, a consistent and commonly used definition is an important step in ensuring quality and consistency in the fields of data and statistics.

55. This chapter will describe and explain important concepts and principles related to data stewardship and propose definitions for endorsement and use. The first section will present an overview of the concepts of data governance and data stewardship used in business and academic literature, followed by a discussion of their relationship. The second section will discuss the scope of data stewardship and data governance in the public sector and will provide specific definitions for data governance and data stewardship along with a few other relevant definitions, sourced from experts from public and private sectors and academia. The third section defines Data Steward as a role, that can be carried out by different persons, not necessarily a position in an organization.

#### 3.1 Data governance and stewardship in business and academic literature

56. **Data stewardship** is concerned with the science and practice of data collection for the purposes of analysis, reflecting the values of fair information practices. In practice data stewardship is a collection of methods and mechanisms of data management encompassing acquisition, storage, protection, aggregation, deidentification, and procedures for data release, use, and re-use, to ensure that the data assets are of high quality, easily accessible, and used appropriately. An important aspect of data stewardship is collecting and documenting metadata which, along with standards, is critical to data quality. The concept of a **data steward** is intended to convey a fiduciary (legal or ethical trust) relationship with data, where a data stewards' loyalty is to the interests of individuals and entities whose data are stored in, and managed by, the system (New Zealand Government, 2020; Plotkin, 2021; Rosenbaum, 2010; StatCan, 2020; 2021b ).

57. **Data governance** concerns decision-making and authority for data related matters, whether within or between enterprises and public agencies. According to a seminal text on data stewardship, data governance is a system of decision rights and accountabilities for information related processes, executed according to agreed-upon models governing the kind of data stored, the authority to access data, and the methods of data access (Plotkin, 2021). Consistently, the Data Management Body of Knowledge (DMBOK) defines data governance as “the exercise of authority, control, and shared decision-making (planning, monitoring, enforcement) over the management of data assets” (DAMA, 2017).

58. Data governance is more about establishing roles and responsibilities – the actual practice of governing. In effect, it is the management of the availability, integrity, interoperability, and security of the data stored. It establishes and enforces the policies for access, management, security, sharing and uses of data; identifies the methods and procedures necessary to the stewardship process; and establishes the qualifications of those who would use the data and the conditions under which data access can be granted (Rosenbaum, 2010). This encompasses the people, processes, and technologies required to manage and protect data assets, and can be in reference to strategic data asset management at the level of the private enterprise or public agency, interdepartmentally, or even at the national level.”

59. Data stewardship is the necessary, operational aspect of data governance. It is an approach to data governance that formalizes accountability for managing information resources on behalf of others and, in the case of government, for the best interests of the public (Plotkin, 2021). Data stewardship

implements the policies, standards, and principles outlined by data governance. While there are similarities in the scope of data governance and stewardship, they are not synonymous. Rather, in practice they seem to occupy two sides of the same data management coin. The European Union's *Data Governance Act* demonstrates this with its focus on interoperability and sharing, on ensuring public sector data is available for reuse, and on leveraging General Data Protection Regulation (GDPR) to balance the use of personal data with individuals' rights (Economic Commission, 2020a). We can also observe this interconnection in the fact that many "best practices" principles of data governance are, in fact, data stewardship principles, which likely contributes to this confusion.

60. The "best practices" and principles for data governance and data stewardship include the following understandings:

- Data are a strategic asset
- Data requires stewardship and accountability
- Data quality preserves and enhances the value of data
- Data must be secure and follow privacy regulations
- Metadata must be easy to find and of high quality



Figure 2. Governance 'V' Model (source: Ladley, 2020)

61. The governance of data can be conceptualized using the "Governance 'V' Model" (Figure 2), which is useful to facilitate the understanding of the subtle differences between governance and stewardship. On the left is data governance – the authorities and policies previously discussed. The right side shows the "hands-on" management of data and information – the function and goals of sound data stewardship. This model emphasizes that data governance is NOT just a function performed by those who manage data (Ladley, 2017). They have a somewhat cyclical, mutually enabling relationship, but the distinction between them is important, both conceptually and practically.

62. Data stewardship is not a new focus, and the data steward is not an entirely new role. Rather, these constitute an extension and re-definition of existing organizational positions that govern, manage, and use data. Traditionally, the focus of a data steward, or of data stewardship activities, was data integrity in the context of internal data governance and management with an emphasis on technical competencies. However, public, private, and academic data experts agree that this narrow conception is no longer sufficient (Verhuulst, 2021). With the rapidly accelerating proliferation of data and the increasing demand for, and potential of, data sharing and collaboration, NSOs and public governance organizations alike need to re-imagine data stewardship to a function and role encompassing a wider range of purposes and responsibilities (Henning et al., 2001; Verhuulst, 2021). The result will be a more efficient and effective use of data assets to address socio-cultural issues, economic challenges, and to improve people's lives.

63. In this way, those performing a data stewardship function (including NSOs) should be empowered to create public value by re-using data, by better leveraging data expertise through strategic partnerships and collaboration, and through the facilitation of functional access to expertise, data, and evidence-based insights (Verhuulst, 2021). This contributes to the promotion of trust and is vital to data collaboratives and public governance, providing functional access to unlock the potential of siloed data sets.

64. Data stewards – at the organizational or interorganizational level – form a new and essential link in the data value chain. Most broadly, data stewards and those NSOs performing data stewardship functions are responsible for the collection, storage, handling, and usage of data. This responsibility can be conceived as three main tasks related to collaboration, ethics, and sharing (Verhuulst, 2021). Responsible collaboration is necessary to unlock data when there is a public interest case, data must be managed ethically to prevent harm and misuse, and action must be taken to ensure that insight is shared with those who need so that it may be translated into meaningful impact.

### 3.2 Definitions of data stewardship and governance, and other related concepts

65. **Data** is the representation of information, in a manner suitable for storage, communication, interpretation, or processing by human beings or by automatic means, and from which knowledge can be drawn, including structured or unstructured forms. While there are many different types of data (digital or hard-copy data, qualitative, quantitative, etc.), statistical data refers to data used to produce statistics (it often comes from a census, survey or administrative source but may also come from other sources) (OECD, 2006; StatCan, 2019, 2021b; UNECE CES, 2000).

66. The data (including statistical data), along with the data subjects, collection of stakeholders and data users, capacities, processes, policies and infrastructure used to capture and analyse data are referred to as the **data ecosystem** (European Commission, 2017; StatCan, 2020b). The data governance and data stewardship activities take place in data ecosystem. They manage the interactions of four main categories of actors in the data ecosystem, which include (European Commission, 2017):

- data generators - the primary sources generating data, whether actively or passively, such as consumers, customers, or citizens engaging with the statistical system;
- data services - the actors participating in data value creation chain; those collecting, organizing, storing, processing, retrieving, sharing, using, reusing, restricting or destructing, such as NSOs;
- data business users - those parties, like companies and public administrations who use the insights derived from data analytics to improve performance or quality of life;
- end customers - the data consumers, customers, or citizens.

67. The information that can be shared in a data ecosystem can be very diverse in nature (statistical-spatial/geographic, structured-unstructured, images-data, text-data, etc.) and should thus not be limited to any specific type of format. Standards should be open to those interested in establishing the formats and type of information they want to share.

68. **Data governance** can be defined as

**a system of decision rights and accountabilities for the management of the availability, usability, integrity and security of the data and information to enable coherent implementation and co-ordination of data stewardship activities as well as increase the capacity (technical or otherwise) to better control the data value chain, and the resulting regulations, policies and frameworks that provide enforcement.**

This includes the systems within an enterprise, organisation or government that define who has authority and control over data assets and how those data assets may be used, as well as the people, processes, and technologies required to manage and protect data assets.

(Data Governance Institute, n.d.; OECD, 2001, 2019; Plotkin, 2021; StatCan, 2020, 2021b).

69. Following the explanation in para 61, data stewardship brings in the aspect of ‘achieving goals’, that is, the purpose **why** the data are collected, processed, used, re-used, etc. According to the first



Fundamental Principle of Official Statistics, 'Official statistics provide an indispensable element in the information system of a democratic society [...] to honour citizens' entitlement to public information.' The Principle also speaks about serving the government and the public with data about the economic, demographic, social and environmental situation, meeting the test of practical utility, and being compiled and made available on an impartial basis.

70. Impartial official statistics do not only help to build trust in the NSOs themselves, they contribute to a wider social goal of fostering transparency and accountability, building an open relationship between society and states. Because of this purpose, data stewardship in official statistics and in the public sector may differ from data stewardship in business or academic environment, bringing in issues of trust, public good, data ethics, serving different types of users, and a long-term view.

71. **Data stewardship** can be defined as

**ensuring the ethical and responsible creation, collection, management, use, and reuse of data. It is expressed through long-term, inter-generational curation of data assets so that they benefit the full community of data users, and are used for public good.**

Data stewardship

- is enabled through good data governance and data management, which provide oversight of data assets throughout their lifecycle to ensure their proper care;
- influences proactive and responsible data practices to help deliver data strategies, maintain trust, and promote accountability;
- works to support the growing maturity of data policy;
- is made visible through a range of internal and external functions associated with stewardship roles - including data access, security, data quality and standards;
- is applicable at all scales, from the national or data system level, to the organisation or enterprise level, to the individual or dataset.

(New Zealand Government, 2020; OECD, 2018; Plotkin, 2021; StatCan, 2020, 2021b). See also the **open data** definition in the glossary available in the appendix.

72. Both data stewardship and data governance professionals are highly focused on data ethics. **Data ethics** is a newer branch of ethics that focuses on moral problems related to data. Data ethics provides guidelines for the handling of data (from collection to reuse), algorithms and corresponding practices to formulate and support morally positive solutions to data ethics problems. This occurs in full compliance with human rights and works against the risks of misuse and non-use of data for public good. Data ethics refers to the knowledge that allows a person to acquire, use, interpret and share data in an ethical manner including recognizing legal and ethical issues (e.g., security, biases, privacy, confidentiality, and public support or social acceptability) (Economic Commission, 2020b; Floridi and Taddeo, 2016; StatCan, 2020a).

73. The data stewardship and governance professionals working in this sphere are subject to data accountability. **Data accountability** is the ongoing liability and responsibility an individual, enterprise, organisation or government is subject to regarding the management of data. It is the principle that ensures that appropriate technical and organizational measures and records are in place to demonstrate compliance to data related principles and policy instruments. Examples may include data protection policies, impact assessments, maintaining necessary documentation, data security measures, roles and authorities related to data protection, privacy management frameworks, and appropriate recording and reporting (Information Commissioner's Office, 2019; StatCan, 1998).

74. These accountable professionals are concerned with **data management**, discipline that directs and supports effective and efficient management of information and data in an enterprise,

organization, or government administration, from planning and systems development to disposal or long-term preservation. Data management involves the development, execution, and supervision of plans, policies, practices, concepts, programs, and the accompanying range of systems that contribute to the organization and maintenance of data processes to meet ongoing information lifecycle needs. Data management enables the delivery, control, protection, and enhancement of the value of data and information assets through integrated, user-based approaches. Key components of data lifecycle management include a searchable data inventory, reference and master data management, and a quality assessment framework (DMBOK, 2017; Government of Canada, 2019; StatCan, 2020a, 2020b, 2021b).

75. The focus on management of data *throughout* the data and information lifecycle, is **data lifecycle management**. An aspect of data management, it is characterized by four main phases outlined in the “Four G” model of data lifecycle management (DMBOK, 2017; Government of Canada, 2019; Rancourt, 2019; StatCan, 2020b). See Section 4 for a discussion of models of data governance and data lifecycle management.

### 3.3 Data steward

76. The role implementing data stewardship is the **data steward**. A data steward is accountable for data assets and resources from a strategic perspective. Data stewards are responsible for ensuring that the data acquisition, entry, quality, interoperability, and overall management supports the needs of consumers, citizens, enterprises, organizations, or governments, while also ensuring adherence to social license, legislative, and regulatory requirements. They work with stakeholders and other deliberative or advisory bodies to develop definitions, standards, and data controls, and perform key functions in the ideation and implementation of data policies that are scalable, sustainable, and significant (OECD, 2001, 2018; Plotkin, 2021; StatCan, 2021b).

77. There is a distinction between ‘data steward’ as a specific person or agency carrying out the task, and ‘data stewardship’ that reflects a collective effort. These responsibilities can be linked with an organisation like a National Statistical Office, as a data steward for the National Statistics System, or for whole or only part of public sector data.

78. For the purposes of this report, the **data steward** (DS) should be viewed primarily as a role, not necessarily a position in an organization or public administration. This means that the DS’s role concerns expected behaviours and responsibilities within a particular position, not a career placement or a title. To illustrate, in a role-based system an individual may hold multiple roles and each role may be held by multiple people. In a position-based system, everyone has only one job position.

79. The data steward role must have a clear mandate identifying their responsibilities throughout the data life cycle, from collection to dissemination and preservation of the data. Depending on the complexity of the organisation or public administration and the functions associated with the production and processing of data, a specific organizational unit may be designated that includes a position of a Chief Data Steward. It is also vital that this steward be part of a network of data stewards outside the organization to promote the use of data as public good for public interests.

80. At the country or federal level, the setup of the government system must be taken into account. That is, a person should be designated to assume the coordinating role in each sphere of government: one in the executive branch, one in the legislative, one in the judicial and one in each autonomous body. This scheme would be replicated at the state and municipal level according to the applicable models of data governance (such as the centralised, federated, and hybrid models that will be discussed in chapter 4). It will be useful to define the role of a data steward also in private organizations and companies that own information assets that can be used by others.

81. In the case of the National Statistical Offices, many positions within their structure are de facto data stewards, fulfilling different tasks and responsibilities necessary to implement data stewardship. When looking at this in the government and wider public administration, the Data Steward role has to be clearly designated as this person (or persons) have to coordinate with other institutions. Special attention needs to be paid to the relations with Privacy Protection or Personal Data Protection Agencies, to avoid duplication of effort and clarify the division of responsibilities. Chapter 4 explains in more detail the different data governance models and where a Data Steward may be situated.

82. When using new/alternative public and private data sources from outside the statistical system, the Data Steward will need to work closely with other relevant staff in NSO, such as persons responsible for data acquisition and data engineering. This is explained in more detail in Section 5.2.3.

DRAFT

## 4 Data governance in the context of data stewardship

83. As described in Chapter 3, data stewardship is enabled through good data governance and data management which provide an oversight of data assets throughout their lifecycle to ensure proper care.

84. This chapter looks at data governance models in the context of where the Data Steward role is situated within the structure of the government or organization. Three types of data governance models (Centralised, Distributed and Federated Hybrid) are used here as a framework of data governance models deployed at a government level. The chapter considers advantages and disadvantages of the different models and the situation where they may work best. The chapter also looks at data governance principles and how they fit with data governance frameworks.

85. Contextual factors such as the levels of centralisation/federation, legal barriers to data access and sharing, institutional arrangements, cultural context, and the type of data (personal data vs. administrative data) play a role when deciding on the data governance arrangements. The arrangements should enable to deliver value from data as an asset. In practice, the complexity of data governance can differ depending on the level of its application (e.g., within a single organisation, at a sectoral level, across government, internationally).

### 4.1 Data Governance Models

86. Many data governance models are available. This chapter focuses on three categories of models depending on where the Data Steward role is situated within the government structure, as follows:

- Centralised
- Distributed
- Federated hybrid

#### 4.1.1 Centralised model

87. In a centralized data governance model a single individual or institution/agency makes decisions and provides direction for data governance. It is a top-down approach where usually a single person is accountable for data governance for the whole government. Data governance is centrally managed, funded and resourced.

88. The centralised model's **advantages** are that it is:

- Easy to manage.
- Dedicated Data Governance Lead.
- Permits strong coordination through the data steward's team.
- More efficient decision making.
- Easier to focus on policy, guidelines.
- Easier to control costs.

89. However, there are a number of **disadvantages**:

- Potentially increased bureaucracy and lower redundancy due to the linear structure.
- Can lead to operational rigidity.
- More time required to accomplish data governance operations.
- Tension can arise between data governance from a centralised location and the business reality of individual agencies having to run their operational systems, including funding to meet centrally mandated data requirements/standards.

- Mostly concerned with all-of-government data priorities, which will tend to narrow the focus of the data under governance to that which is commonly shared.
- May hinder experimentation and creativity.

90. The centralised model can work well in small organizations with low data maturity and a small capacity for data management. It is less suitable for an all-of government approach where multiple autonomous agencies are involved, as the negotiations that would be required to make this work would effectively turn it into a Federated Hybrid model.

#### 4.1.2 Distributed model

91. In a distributed data governance model individual agencies will have independent governance processes, and are likely to have different governance models at individual agency level. There will be no central authority responsible for co-ordinating data governance or decision making.

92. This model is characterised by a bottom-up approach. Government agencies own and operate independent governance processes. It is likely to lead to differing models, maturities and difficult coordination between agencies.

93. This model's **advantages** are that it:

- Empowers agency chief executives to make decisions that meet their agency's needs
- Is relatively easy to establish
- There are clear rules for responsibilities and processes (data governance models fit well for organization's own needs)
- Lowers bureaucracy

94. But there are a number of **disadvantages**:

- Individual agencies tend to be focussed on their own internal data governance –there is no external agent to coordinate inter-agency work. Some collaboration is possible, but difficult to sustain.
- It is challenging to address government-wide initiatives or ensure interoperability and data sharing because of lack of inter-agency data standards. Agencies tend to become more siloed which may reduce the capacity for agencies to integrate data.
- Reaching consensus tends to take longer.
- The model can make it difficult to coordinate and commit the needed resources from agencies particularly as funding tends to be focussed on individual agency needs rather than interagency needs.
- The model lacks a co-ordinated approach to data governance across agencies – potential for incompatible models.
- Potentially increases redundancy.

95. This model works best in small or medium size enterprises/systems with multiple locations/organizations involved. The model was widely used in traditional paper-based data governance, but the operating environment has changed, and the new digital data governance requires common rules and standards to make inter-operability possible. For this to work at an all-of-government level there must be some degree of coordination or at least accepted norms.

#### 4.1.3 Federated Hybrid model

96. In this model there is still a centralized structure which oversees the government sector data environment, but there is also bottom-up input enabling participation from individual government agencies. The centralized structure provides a framework, tools, and best practices for the agencies to

follow but leaves the agencies enough autonomy to manage agency-specific data and offers channels of influence to gather input for data sets impacting all-of-government data or the other way around.

97. This model is characterised by centralized control & distributed management. The government-wide strategy is centralized but its execution and implementation are decentralized. Data governance is centrally lead with representation from all individual agencies. Collaboration and good organizations is key.

98. The decision-making regarding standards and shared data is centralised. The program, processes, standards, guidelines, and systems are usually set up by a core team, task force or board, and individual agencies are responsible for applying these locally.

99. This model may involve the adoption of common data model by different organizations (Replicated model). It may also adopt a common approach across ecosystem (cross-agency) to serve as response to a domain specific data issue (e.g., COVID-19).

100. The **advantages** of the Federated Hybrid model are:

- Individual government agencies are responsible for their own data and metadata
- Allows for broad membership for working groups (but these should include those responsible for data and decision making).
- Provides the ability to focus on specific data sets at the individual government agency level and how these datasets relate to all-of-government data.
- Provides the autonomy to develop standards, policies, procedures at the individual agency level (but requires that these align with an all-of-government model for data sharing to ensure interoperability).
- Issue resolution at a bottom-up approach (must have clear accountability and process for issue resolution – taken to new or existing committee).
- Relatively easy to establish.

101. The **disadvantages** are:

- A highly skilled Data Governance and Stewardship lead position (or team) is required full-time.
- Requires clear roles & responsibilities for collaboration & sharing.
- May require strong levers (including legislation) to drive adoption of all-of-government standards, meta-data & models.
- May get very political at the working group level.
- Decisions made at the group level may tend to be pushed up to the upper levels for approval.
- It can be difficult to find a balance between all-of-government priorities and those of the individual government agencies.
- Oversight over the autonomy of individual government agencies can be challenging and relies a lot on self-reporting.
- Individual agency efficiency depends on localized skills – data maturity.
- Metadata management is not simple to address as it can differ widely from one agency to another.

102. The Federated Hybrid model tends to work well in large organizations with varying levels of data maturity. It is the most practical model for an all-of-government environment.

103. The principle of the hybrid model requires establishing norms (laws, administrative provisions, etc.) that are necessary for data reuse. That underlines the need for a Data Steward. The adoption of this model may require legislative changes to avoid the multiplication of negotiations to obtain

resources on a case-by-case basis. It involves establishing the institutional arrangements and clarifying the necessary roles to make this model work.

104. The centralised model can work in small organizations. In this case the data stewardship role would be established within this organization (explicitly or implicitly), linked with data governance. However, it would be very difficult to make the centralised model work for all-of-government involving multiple autonomous agencies.

105. The distributed (decentralised) model is the one that happens naturally if there is no coordination and there are no attempts to establish common data strategies and standards. It leads to redundancy and lack of interoperability. There may be Data Stewards in individual agencies but their possibility to implement an ethical and responsible approach to data is limited with their own agency.

106. The federated hybrid model looks like the best one for an all-of-government approach but is the most complex one to implement effectively. In reality, many different setups and combinations of this model are possible depending on the institutional structure, the level of centralisation of the decision making on standards, formats, metadata, etc. In this case, there would be a need for a data steward at the whole-of-government level and data stewards in individual agencies.

107. Annex 2 provides examples of data governance models and frameworks used by different countries.

## 4.2 Supporting principles of data stewardship

108. This Section discusses some of the supporting principles of data stewardship and other enabling pieces of the data governance puzzle. The concepts help facilitate governance activities and drive sound data stewardship. The concepts include Fundamental Principles of Official Statistics, standardization, the FAIR Guiding Principles of data, the Four G Model of data lifecycle management, the 'once only' principle and the European Interoperability framework.

### 4.2.1 The UN Fundamental principles of Official Statistics

109. Developed and adopted by the Conference of European Statisticians, the United Nations Fundamental Principles of Official Statistics were adopted in 1992 at the ministerial level by the Economic Commission of Europe, and at the global level by the UN Statistical Commission in 1994. The Fundamental Principles were then reaffirmed and endorsed by the United Nations General Assembly in 2014. The preamble of the Principles highlights the importance of official statistics for national and global development, emphasizing the vital role that NSOs play in providing high-quality, official statistical data, and the positive effect that this has on policy decision-making and the overall development of countries.

110. The Fundamental Principles<sup>10</sup> provide an international, intergovernmental guide for how NSOs can best provide their services to citizens. The first Principle states that '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 summarize, the Fundamental Principles state that official statistics must be: practical and impartial; accurate, trustworthy and ethical; scientific and standardized; high quality, timely, and with as low a response burden as possible; confidential and appropriately used; transparent and made public. The document also states that NSOs and other statistical agencies can draw data from a variety

---

<sup>10</sup> 2022 marks the 30<sup>th</sup> anniversary of the Fundamental Principles. The anniversary campaign includes useful materials explaining and promoting the Principles. For more information, see: <https://unece.org/FPOS30>



of sources - that should be assessed at the discretion of the office or agency. And they coordinate and cooperate with other organizations (nationally and internationally), to contribute to the improvement of statistical systems (United Nations, 2014).

#### 4.2.2 Standards

111. As professionals of the international data community, we understand the importance of **data standards** and **metadata** (see Annex 1. Glossary for definitions). The guidelines by which data are collected, described and recorded, as well as the accepted practices, technical requirements, and terminologies for the field, are critical for data quality. They also provide information about data collected to help further understand and interpret data. To share, exchange, combine and understand data fully, the format and the meaning must be standardized. Statistical and data standards enable organizations and governments to collect, process, and produce quality statistics using consistent methods and procedures. This includes standard names and definitions (or ontologies), statistical units, concepts, variables, and classifications. Standards in statistical programs ensure that the data are collected, represented and interpreted responsibly and accurately, in a way that is *consistent* (Green, 2021).

112. If standards are adhered to, data over time and from different sources can be better integrated, thereby maximizing the data's productive value. Standards provide the basis for consolidating statistical information, increasing the capacity for interoperability by eliminating the need to conform data or metadata to new specifications, and reducing time spent cleaning and translating data – a common barrier to data analysis that accounts for much of data users' on-the-job time (Plotkin, 2021). Considering traditional statistical methods, having standards in place – and the infrastructure to monitor, assess, and improve those standards – also reduces the resources required to develop and maintain surveys. This is vital, as national statistics offices (NSOs) around the world are looking to harness the power of digitization and of “Big Data”, not only to increase the efficacy and efficiency of the data for public use, but also to reduce the response and financial burdens of traditional statistical systems on citizens.

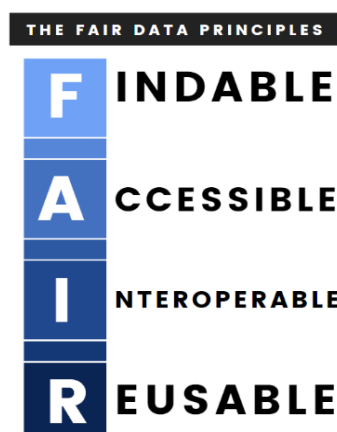
113. It is important to establish standard methods and formats for storing data, so that information collected by one department or agency is comprehensible and valuable to another. Standards are the common languages professionals use to communicate ideas between people, departments, agencies, nations, and even throughout time; they are necessary for productive collaboration (Green, 2021). This is why, globally, many NSOs are participating in multinational task forces, on international manuals, in classification working groups, or in industry conferences and general discourse about data and metadata standards, to elevate and advance these important standardization principles.

114. Through this work, the international statistical community not only ensures that the quality of data is consistent across history and geography, but also helps to equip public, private and academic sectors that produce and manage data to better integrate data from various sources, as well as enable international partners in complying with those transnational reporting obligations that ultimately encourage data comparability between countries. NSOs are also incorporating data standards directly into their data strategy and data stewardship frameworks. Given this clear importance of standards in data stewardship, a definition and explanation of the FAIR Guiding Principles is useful.

#### 4.2.3 The FAIR guiding principles of data

115. Data standards and metadata make data **Findable** and easily searchable, make data **Accessible** and easy to use, make data **Interoperable** and more easily interpretable, and they make data easy to use, share and **re-use** data.

116. The Principles define characteristics that modern data resources, tools, vocabularies and infrastructures should demonstrate in order to assist in the discovery and reuse of data by other parties (Wilkinson et al., 2016). The FAIR Guiding Principles were developed by a consortium of scientists and organizations and then defined in a 2016 paper in the journal *Scientific Data*. These high-level FAIR Guiding Principles are not, themselves, a standard or a specification. They provide guidance to data producers, publishers, and stewards to support them in evaluating “whether their particular implementation choices are rendering their digital research artefacts Findable, Accessible, Interoperable, and Reusable” (Wilkinson et al., 2017).



117. If consulted regularly and employed diligently, these principles will enable a more integrative and exploratory approach to data discovery, data storage, data sharing, and data use and reuse. When implemented, the management and stewardship of these valuable digital resources is maximized, to the benefit of the entire statistical community, as well as the citizens that NSOs serve. As data stewardship is concerned with the management of data and information assets throughout their lifecycle, adherence to the FAIR Principles (and even incorporation of them explicitly into data stewardship messaging) will contribute to a focused yet encompassing approach to ensuring data quality.

#### 4.2.4 The 4Gs Model of Data Lifecycle Management

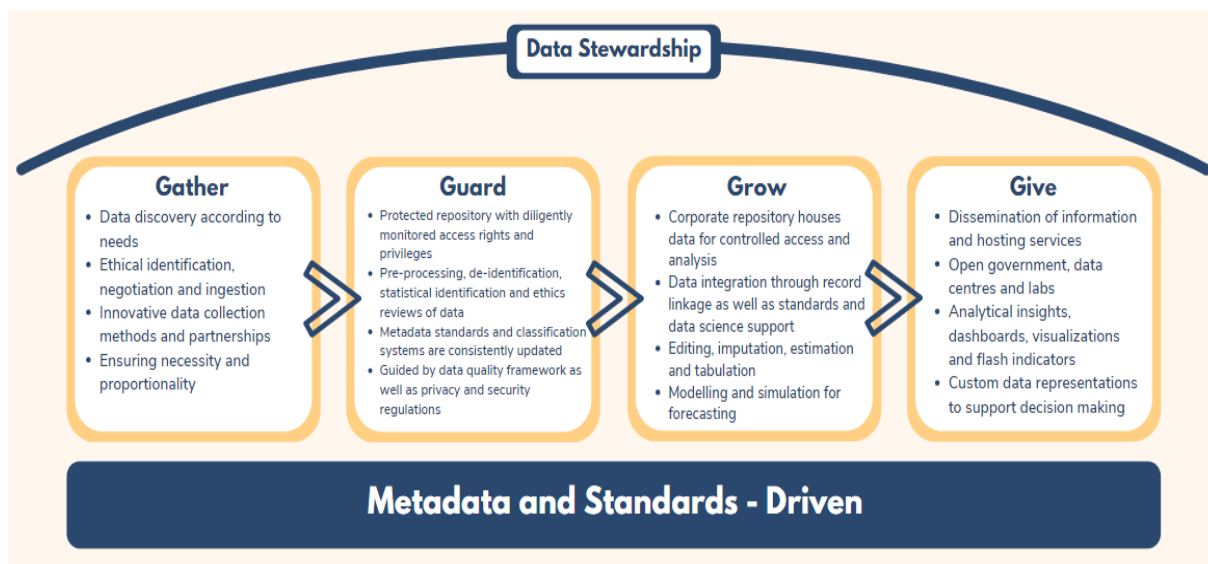
118. To advance National Statistical Systems (NSS) globally, a commitment to sound stewardship is required throughout the data lifecycle. To achieve this, Eric Rancourt of Statistics Canada developed the “Four Gs” model of data life-cycle management – *gather*, *guard*, *grow*, and *give* (2019).

119. In this metadata-driven system, “**gather**” refers to all data ingestion, including the collecting and integrating of data assets through various systems of acquisition, as well as the policy instruments and ethics-based legislative frameworks through which the agency gains access to data and information (Rancourt, 2019). Sound data stewardship ensures that this data is acquired efficiently, ethically, and without duplication or redundancy.

120. To “**guard**” data, special attention is paid to access rights and privileges, data audit trails are performed, data monitoring and back-up protocols are systematized and ongoing, and metadata standards and classification systems are consistently updated (Rancourt, 2019). The goal in guarding data is to adhere to the “privacy by design” principles, ensuring that data is secure and encrypted, confidential and de-identified, and with all necessary privacy protocols in place in order to function ethically and according to our trust framework.

121. To “**grow**” data, the data is organized, processed, transformed, integrated, and extracted from for various uses (Rancourt, 2019). During this phase, data is cleaned and verified, quality assurance is performed, data is analyzed, explanations are developed, and hypotheses are tested. Efforts are made to grow data by ensuring its optimization and adhering to (and continually developing) data quality frameworks.

122. Finally, to “**give**”, data and statistics are shared and published. Here, data access and interoperability are ensured, dissemination occurs regularly and with quality and accessibility, and the appropriate metadata is made available based on strategic requirements (Rancourt, 2019). The goal for Statistics Canada, and the public service more broadly, is to increase data discoverability and be “open by design”, having sharable and open data, metadata, metainformation, and analysis.



123. Organizing data stewardship activities along these data lifecycle phases and aligning them to our overall role as an NSO – which indeed, is to gather, guard, grow, and give data – has been extremely valuable. It has allowed us to ensure that data is efficiently and optimally used and reused, that high-quality data is consistently discoverable and accessible, that expertise is appropriately leveraged, that standardization is maintained, and that public trust and engagement is preserved and encouraged, by operating ethically and transparently (Rancourt, 2019). For a visualization of the 4G Model, see the graphic above.

#### 4.2.5 The ‘once only’ principle

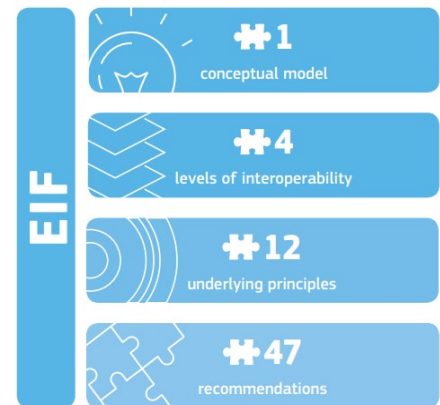
124. The once-only principle is an e-government concept that aims to ensure that citizens, institutions, and companies only have to provide certain standard information to the authorities and administrations once. By incorporating data protection regulations and the explicit consent of the users, the public administration is allowed to re-use and exchange the data with each other. The once-only principle is part of the [European Union's \(EU\)](#) plans to further develop the [Digital Single Market](#) by reducing the administrative burden on citizens and businesses (to be launched in 2023). (*"European eGovernment Action Plan 2016-2020". Digital Single Market*).

125. When data becomes increasingly digitized, the desire to integrate data sets from different organizations for evidence-based decision-making increases. The use of existing data should be managed efficiently, and no extra data collection should be used if there is a government organization that has already collected the required data. This Principle provides a strong incentive to establish common rules and standards to make inter-operability possible.

126. It is also closely linked with ethical and responsible handling of data and therefore data stewardship.

#### 4.2.6 The European Interoperability framework

127. The European Interoperability Framework (EIF) is a generic framework devised to be applicable to all public administrations in the EU. It outlines the conditions necessary for achieving interoperability at all levels, including European, national, regional and local, embracing public administrations, citizens and businesses. This guide to EIF is addressed to all those involved in defining, designing, developing and delivering evidence based and data-informed European public services. EIF begins with guiding principles for interoperability and an interoperability model appropriately layered to address the aspects of EU public service administration. A conceptual model for interoperability in public services follows, promoting the idea of ‘interoperability by design’ as a standard approach. The EIF guide concludes by providing 47 concrete recommendations to be reflected on and actioned by public administrations.



EU Interoperability Framework Structure, (European Commission, 2017)

128. The EIF aims to inspire European public administrations in their efforts to design and deliver the European public services in a manner that is open, digital and cross-border-by-default, to the degree that is possible. For example, providing services and data through digital channels to all citizens in the EU and enabling access and use, reuse and transparency of data. The EIF also aims to provide guidance to public administrations on the design and update of national frameworks, policies, and strategies around data interoperability, such as national interoperability frameworks (NIFs) and national policies, strategies and guidelines promoting data interoperability. The goal is to contribute to the establishment of the digital single market by fostering cross-border and cross-sectoral interoperability for the delivery of European public services.

129. The EIF's scope covers three types of interactions:

- 1) A2A (administration to administration), referring to interactions between public administrations, such as a Member State or EU Institution;
- 2) A2B (administration to business), which refers to interactions between public administrations (in a Member State or an EU Institution) and businesses;
- 3) A2C (administration to citizen), which refers to interactions between public administrations (in a Member State or an EU institution) and citizens

130. EU and national policies (e.g. NIFs) are expected to build upon the EIF, by adding new or fine-tuning existing elements and ensuring its contextual relevance. In a similar way, domain-specific interoperability frameworks (DIFs) should remain compatible with, and where necessary or possible, expand the scope of the EIF to capture the specific and positioned interoperability requirements of the domain.

## 5 Responsibilities and skills for data stewardship

131. Data stewardship deals with strategy, culture, ethics, roles and settings, and the capabilities of people that support an ethical and responsible creation, collection, management, use and reuse of data. Data stewardship has a long-term focus, to provide oversight of data assets throughout their lifecycle. It can be exercised at an individual agency and at system level. There is a distinction between 'data steward' as a specific person or agency carrying out the task, and 'data stewardship' that reflects a collective effort.

132. The responsibilities of a Data Steward can be linked with a person or an organisation. For example, the National Statistical Office can act as a data steward for the National Statistical System (NSS), or for only part of public sector data. Within an organisation there must always be a person or persons responsible for data stewardship. There can be data stewards (or somebody to whom the stewardship tasks are assigned) in each unit of the organisation dealing with data.

133. Data Stewardship responsibilities can be externally or internally oriented (see figure 3):

- Externally at different levels:
  - within the wide data ecosystem (including towards private data holders);
  - concerning public data; and
  - concerning other producers of official statistics, parts of national statistical system.
- Internally within an organisation, involving a 'core' set of responsibilities for all organisations dealing with sensitive data and roles that are appropriate for the mandate of a National Statistical Office.

134. This chapter outlines the roles for consideration for both internal and external orientation, and gives some examples of the skills required where such information is available from NSOs. For internal roles, the overall NSO context based on the GSBPM is outlined to show the relationship of data stewardship to the business areas.

### 5.1 Involvement in Public Sector data strategies and data management

#### 5.1.1 Data stewardship in different institutional settings

135. National Statistical Offices have extensive experience with dealing with data throughout its lifecycle. The digitalisation of society and economy has placed data access and sharing at the core of innovation and public trust, with the use of data from other sources and for different purposes growing in importance. Therefore, the questions of data and information management, data access, privacy protection, and data security are high on the agenda of governments and the public sector. Several countries have developed or are developing public sector data strategies. Statistical offices have a lot to contribute on these topics and should be involved in this discussion.

136. A Data Steward is a strategic and tactical advisor within all Government organisations, who may also have overarching responsibilities across organisations, and therefore a liaison with data owners about the use of standards and classifications, description of metadata and data quality requirements. While each NSO is clearly Data Steward (DS) for its own data, including acquired public sector data, the extent of the outward facing role depends on the institutional setting for each country.

137. The role that the National Statistical Organisation (NSO) takes in government sector data stewardship can vary from being limited to just within the NSO to providing guidance and direction on data management for the whole Government sector. In Figure 3, these different responsibilities and coverage are shown. The outer circle shows the greatest coverage – the whole government sector; and the inner circle shows the coverage of only one institution - the NSO, with each successive circle including the ‘smaller’ roles.

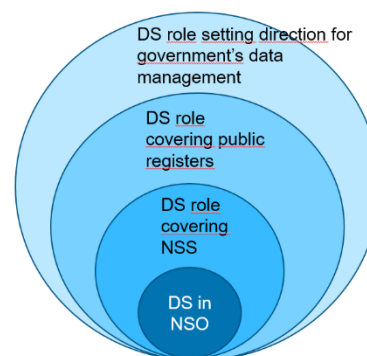


Figure 3, Possible coverage of the Data Stewardship role

138. It may be also the case that the data stewardship role for the whole government sector lies with another institution and the NSO has this role only for official statistics. In this case, the NSO should extensively cooperate with the public sector Data Steward, and the NSO should have an integral part in elaborating standards, quality frameworks, etc. as they have an impact on official statistics. As some NSOs might not want to take on the widest role of data stewards, or some governments might decide to implement different organisational solutions, it is vital to have some minimum safeguards in place for NSOs as regards their rights and responsibilities, and their relations to data stewardship should be outlined in legislation.

139. The broadest role is also relevant to NSO's in countries where public registers do not exist, since the NSO may still set the direction for Government data (e.g. New Zealand). Section 4.1.3 sets out examples of hybrid approaches within this framework with shared responsibilities.

140. The position and responsibilities of the NSO in Government data stewardship must be connected with the Government's Public Sector Data Strategies and recommendations on data management. The current state of data management in different countries varies widely, so a single solution that would meet all NSOs needs cannot be given. If the Government has an Information Policy which directs the development of Government activities in data management, the role of the DS situated in the NSO can be negotiated. But if the Government or Ministries do not have a co-ordinated view on government sector data management, then the role of the NSO is limited, as all institutions and organizations in the government sector will be making their own decisions on how to handle the data according to good principles.

141. When data becomes increasingly digitized, the desire to integrate data sets from different organizations for evidence-based decision-making increases. The use of existing data should be managed efficiently, and no extra data collection should be used if there is a government organization that has already collected the required data ('once-only principle'). Thus the path to better use of data is through increasing the interoperability of data and the use of common technical interfaces.

142. To have good interoperability, it is necessary to have common guidelines, structures, metadata definitions, quality frameworks, common identifiers etc. in place. The statistical society has solved this problem decades ago by using international concepts, classifications and methods to promote the consistency and efficiency of statistical systems (Fundamental Principles of Official Statistics 9).

143. This pool of knowledge of interoperability and data quality inevitably developed in NSOs can be used to promote the interoperability and the use of standard classifications, metadata structures and quality frameworks in the NSS, among holders of administrative data or/and among other institutions and organization within government. There are some key relationships with data holders who have compatible data and skillsets to NSOs (e.g., Mapping and Environmental Agencies) that can be organized through partnerships, which is described in Section 5.3.3.

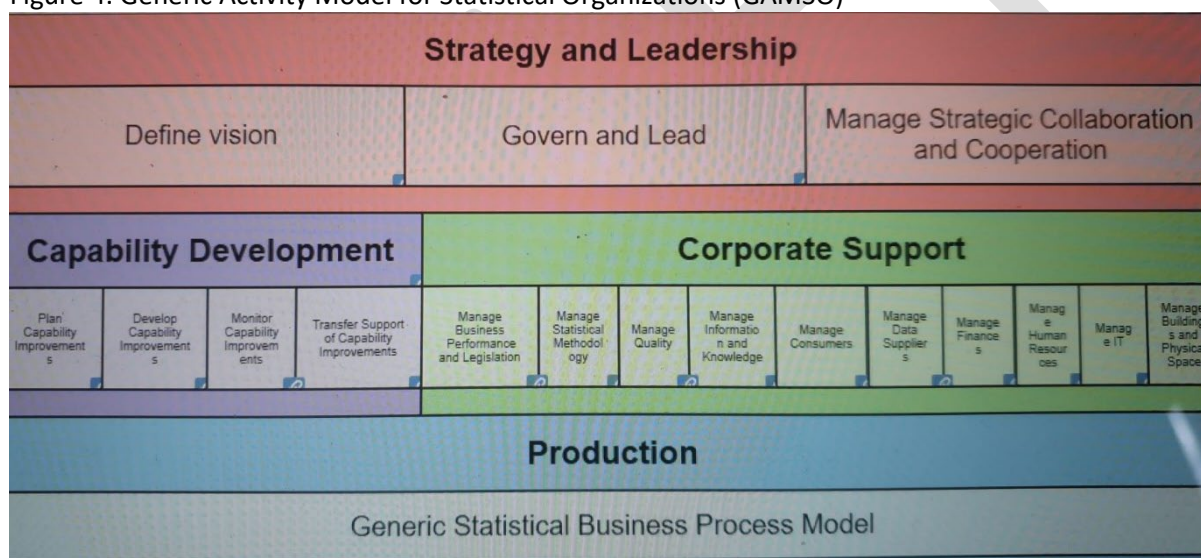


144. Increasingly, NSOs are seeking to access privately held data to fulfil their public service mandate for faster and more disaggregated information for Government and society. While NSOs are unlikely to be consulted on the interoperability or quality dimensions of such data, the concepts of stewardship described in this paper are even more relevant to reuse of such data, since the networking, management and legal skills of the Data Steward will be essential to achieve sustainable access. This role is described in Section 5.2.2.

### 5.1.2 NSO role in Data Stewardship in the public sector

145. The national context is crucial to the role that the NSO adopts in the data ecosystem. There may already be national bodies with responsibility for aspects of data management. The NSO will be required to collaborate with these bodies to, at a minimum, protect the integrity of the NSO and NSS. The NSO may also seek to maximise its influence in the ecosystem by ‘marketing’ its skillsets and taking advantage of any roles that are appropriate to an NSO in the broader ecosystem. The NSO could of course assume the role of Data Steward for the system, but that may not be an option in all countries. This section sets out the list of potential tasks at public sector level, and some examples of different approaches for countries. Firstly, it should be noted that the GAMSO model, with its focus on Strategy and Leadership, is ideal for framing the NSO’s role in the national context.

Figure 4. Generic Activity Model for Statistical Organizations (GAMSO)



146. The ‘Strategy and Leadership’ sub-activities of ‘Define vision’ and ‘Manage Strategic Collaboration and Co-operation’ are specifically relevant to identification and management of the national data stewardship role. ‘Capability Development’ is necessary to identify the areas where the NSO can have most impact in the National context, while ‘Corporate Support’ activities also include many potential externally focused services described in this Section.

147. Examples of the new tasks in the context of Data Stewardship that NSO can take on are:

- Management of all government data in common **data lakes** or integrated data bases
- Development and maintenance of the **quality framework** for administrative registers
- Development of **research services** for scientific purposes and **information services** for evidence-based policy making by integrating statistical and administrative data
- **Holding of data space** (e.g., sector-specific data lakes in health, transportation, energy) and the production of information and research services from data spaces
- Development of **integrated data sets** and enhanced access to open data



- Holding of information on existing data /databases in data **catalogues**
- **Standardization of metadata structures** and classifications for the use in government organisations
- Development of **common data handling methods** with other producers of data
- Management of a **joint dissemination platform** with other producers of data
- **Analytical support** for public sector data providers
- Establishing **common identifiers** for data

148. A description of some national settings is given in boxes below.

#### Box 1. Examples of Data Stewardship

##### Finland

Statistics Finland operates in a decentralised system, which provides opportunities for the NSO-based Data Steward role to extend its focus outside the NSO. In this case the Data Steward role could extend to cover other producers of statistics (Other National Authorities - ONAs). According to EU law, the NSO can give guidance to other offices on data production, dissemination and quality etc. In Finland, this guidance is provided by the Advisory Board of Official statistics where the NSO and ONAs all participate. This guidance can/could be understood as a Data Stewardship role. In this external role, we need to work more co-operatively with a focus on coordination than we do in the Data Stewardship role inside the NSO. Also an understanding of data availability, quality and interoperability is crucial.

Another possibility for the Data Stewardship role within the NSO is to cover other institutions and organisations in the government sector. In Finland, the discussion is moving to that direction; the whole picture is not yet clear but is likely to focus on quality. The Digital and Population data services agency has a Data Stewardship role concerning metadata, standardisation and interoperability. Also, here the skills required concentrate on cooperation and coordination as well as having good understanding of working environment and roles of different players.

##### Ireland

The Government CIO also has responsibility for the Public Sector Data Strategy (PSDA) and the implementation of the Data Sharing and Governance Act (DSGA). These initiatives facilitate data sharing and use outside of pre-existing legislation such as the Statistics Act. Both will facilitate 'levelling up' of data access and use across the public sector, but are relatively immature at the moment. The Central Statistics Office (CSO) were heavily involved in the drafting and implementation of the PSDA and share responsibility for data issues with the Government CIO in Public Service Reform initiatives. The focus of CSO in Public Service Reform so far has been on promoting the use of common identifiers for people, businesses and addresses. We now see ourselves moving more into classifications and methodology, which is a role similar to CBS. This role will be facilitated by the committee structure of the Data Sharing and Governance Board, which oversees the implementation of the DSGA.

The four sub-committees of the Board are:

- Data Architecture and Technical (which includes analytics) – CSO leads on this group
- Data Protection and Ethics
- Advocacy and Communications
- Data Sharing - to oversee processes/frameworks/templates as well as recommending approval of data sharing agreements to the main Board; CSO has no involvement in this activity.

In addition to leadership of the first sub-group, the CSO DG sits as an observer on the main Board.

##### Netherlands

Statistics Netherlands (CBS) sees a role for itself as Bureau of Standards with regard to statistical data. This mainly concerns classifications and methods relating to statistics. CBS also has a [potential?] role in helping departments with [implement?]federated models (meta) and techniques. CBS has knowledge and skills of data processing and in this role it can share this knowledge and skills. CBS is closely involved in the development of the Interdepartmental Data Strategy (IBDS). The implementation of this strategy may give substance to the CBS's data stewardship within the government sector.

### New Zealand

The Government Chief Data Steward (GCDS) role is held by the Chief Executive of Statistics New Zealand (the NSO), and this position also has the role of Chief Statistician. The GCDS role leads by facilitating and enabling a joined-up approach across government, rather than by directing. As well as developing policy and infrastructure, the GCDS provides support and guidance so agencies can use data effectively, while maintaining the trust and confidence of New Zealanders.

GCDS:

- Sets the strategic direction for government's data management
- Leads New Zealand's state sector's response to new and emerging data issues
- Is in the process of co-developing a Data Stewardship Framework to enable agencies to manage data as a strategic asset and benchmark their data maturity
- Leads the government's commitment to accelerating the release of open data.

### 5.1.3 Setting a longer term direction for the system as a whole

149. It is increasingly recognized that NSO's are in a unique position to evaluate the strengths and weaknesses of the national administrative data system, as we have the skillsets and data access to perform this evaluation. 'Once only' requires that such gaps are identified and the value of data collected in each silo is maximized through subsequent integration and analysis. The improvements we identify for statistics are generally also necessary for administrative purposes, so our requirements are similar. However, NSO's are often not aware of the full set of requirements in the public sector, which is an additional stewardship requirement to identify these needs. The NSO's role in setting the direction for the system can range from setting reference classifications to monitoring of uptake of classifications and standards affecting statistics (e.g. common identifiers). The most evolved role is where the NSO sets the standards for the 'Integrated Data Infrastructure' (NZ/Estonia?) which are subsequently implemented and monitored across the public data system.

### 5.1.4 Responsibility for trust and confidence in the system as a whole

150. Privacy preservation and social licence are second nature to NSOs as we have operated in this context for a long time, but these discussions are becoming much more relevant to the broader public sector now as public sector data quality improves and demands for analysis within individual ministries increase. These ministries operate in a multidimensional context, where the social impact of their policies have impacts or barriers in other domains, which increasingly requires data linking across domains. While NSOs can meet some of these demands, the NSO may also need to act to ensure that overall trust in the system is maintained.

### 5.1.5 Analytical support for public sector data providers

151. In many cases, data providers are also data users since Government ministries usually have policy functions as well as operations. At a minimum, information on relevant outputs should be viewed as part of the overall relationship with such Public Sector Bodies (PSBs). A more proactive approach would involve the analysis side of such PSBs in the liaison groups governing data flows, where they could have an input into the statistical work programme based on the PSB's data sources.

152. This requires more responsive official statistics. Openness to the outside world and some level of healthy competition between statistical areas (while maintaining coherence of outputs) to meet these needs has been very helpful in making some NSOs more responsive externally. Societal impact and related metrics should be measured to direct external focus. The NSO should demonstrate its leadership role in data analysis in these outputs, which derives from broad data access and the concentration of relevant skillsets and specialisations in the NSO.

153. Support for analysts in the public sector can also be organized through networks, which the NSO is ideally placed to facilitate. These networks can be used to facilitate external analysts in peer reviewing their work, but also allow for information sharing by NSO's in a broader forum. Networks can be focused on data quality, data analysis or other topics of common interest.

#### Box 2. Example

##### Ireland

Since 2017, the Statistical System Co-ordination Directorate in CSO has offered 'pathfinder projects' to Departments to demonstrate the power of linking data already held by CSO across social policy domains. This is an important service for the various PSBs and also acts as an encouragement for the Departments to collect identifiers and implement standard classifications in their data holdings. It has also been useful in acquiring new datasets from PSBs. There are no new skillsets required for the statisticians involved, and it has particularly suited new hires. CSO also operates two networks which are well attended; The Formal Statistician Liaison Group focusses on data analysis, while the National Data Infrastructure Champions group focusses on promotion of identifiers and classifications.

#### 5.1.6 Data and information services

154. This is the stewardship area many NSOs are considering as they decide whether to become 'data and statistics' offices. 'Differential privacy' is key to these services. NSOs provide open data, and can also facilitate other public bodies to meet their new obligations under the Open Data Directive. Researcher access is one of the key areas for development, including the new role in facilitating access to data for COVID researchers in some countries. It is also a focus in EU legislation, as the Data Governance Act envisages pseudonymised public sector data access for researchers, which may be considered in the stewardship activities of NSO's. Between open and pseudonymised data services, many NSO's also provide anonymised data services to encourage data skills among inexperienced researchers.

155. The main requirement here is to understand different types of data that can be used for research purposes and how that data has to be protected. Also, the NSO staff member giving these services must understand what information is needed for the researcher's research plan/program and whether the required data fulfils these needs.

156. The agreement process has to be documented and security rules defined before the data can be transmitted to the researcher.

#### Box 3. Examples

##### Ireland

There has been a heavy demand on our legal and data governance teams in CSO for COVID data access as the researcher access needs to be within our legal mandate and GDPR. We will become more active in technology services in the coming years. The main proposed CSO Government data services in addition to researcher access are surveys for Government, dissemination tools (the PXStat service), advice on statistics/analytics tools and technology (in partnership with the Government CIO) and consultation in new systems development, where we have an underused legal mandate at the moment. No new skills are required but it will require more external focus from CSO experts. This is all resource-intensive work and is likely to require additional staff. Activities related to citizen data science and machine learning will also be part of this role, which will be done on a partnership basis with geographical, environmental and transport agencies who are active in these fields.

#### Netherlands

With regard to data services, CBS provides remote access to microdata. In the coming years CBS intends to establish even more links with government and businesses with regard to AI applications. What and how this will be done is currently being worked out in a policy framework. CBS also works together with a number of municipalities in the form of an Urban Data Centre (UDC). The aim of this UDC is to stimulate data-driven work. CBS data sources are combined with local data sources. CBS works together with individual municipalities to fulfil the tasks of data stewardship and CBS employees work in the municipality's office. At the moment there are about 14 UDCs in the Netherlands. <https://www.cbs.nl/en-gb/onze-diensten/urban-data-centres>.

Experiences with the UDSs have led CBS to continue working with the current UDCs, but for new social issues at the municipal level we are looking for generic solutions / instruments for all municipalities in the

#### 5.1.7 Promoting data quality and standards

157. To lead and govern the quality of the data of public organisations, the data steward needs to have several kind of abilities and capabilities. First of all, the NSO needs to have a proper understanding of the public data space/ infrastructure and data flows in between different public organisations, even including private data (depending on the coverage of its role, see Figure x Circles). In other words, architectural understanding of the National Data Infrastructure is a necessity. In some countries the importance of the quality of public registers is especially high to NSOs.

158. Secondly, basic understanding of the different dimensions of data quality and different standards and scientific research on this area is a good basis for creating national solutions. Thirdly, data strategies and initiatives ongoing, for instance at EU level, need to be understood as well as the recognition of the National political level in this area.

159. Moreover, coordination and communication and networking skills are all needed when creating and maintaining national data quality solutions. This is long term work and needs to have specific resources as well. Different kind of resources, which are familiar with methods and also resources with deep knowledge on data in question.

160. It is also important to note that the level of data quality required depends on how the data is used, in other words the purpose of the data usage in question. There might be different kind of data usages for the same data leading to different kind of quality assumptions and needs concerning that data. These need to be discussed at an appropriate level. Following that, the data steward might need abilities and skills to negotiate and prioritise as well as coordinate the National data quality improvement actions within the public sector. Specific topics the NSO could consider as part of the stewardship role include:

- 1) **Metadata** is also a crucial dimension of quality and needs to be maintained within NSO's for administrative data as the data were not originally collected for statistical purposes and each source as its own dimensions of quality depending on mandatory/legislative fields, coverage, etc. The NSO may also consider making the metadata publicly available, at a minimum to researchers. As a further step, the NSO may define common metadata standards, or collaborate with the body developing the standards
- 2) **Development and promotion of other data standards** in the community of data owners, such as input and output **classifications**. Again, as these standards are needed for internal purposes, they may be useful to other bodies collecting similar data to help with use and linkage.
- 3) Promotion of **unique identifiers** as a special category of standards, which is particularly relevant to countries without registers. If coverage of such identifiers is poor, the NSO can promote standards for identifiable data to improve the subsequent level of probabilistic matching.
- 4) While the primary concern in relation to **privately held data** so far in NSO's has been access, we also need to think about the **quality frameworks** that we will apply to such data. The quality

model that applies to public sector data may not be relevant to private sector data. ‘Volume, velocity and variety’ will result in different, but very informative, statistical products than traditional administrative sources with high and predictable population coverage levels.

#### Box 4. Examples

##### Finland

The Government of Finland has a development programme on data usage and data dissemination. The main goals are to improve the usability of public data resources, improve the interoperability of the data and to open public data as much as possible. The programme is divided in four subprojects: data strategy, data quality, opening data and technical & semantic interoperability. Statistics Finland is in charge of the data quality framework project.

The aim of *data quality framework project* is to increase the usability and uniformity of data and extend the use of data for decision-making in society and by enterprises. The project has studied the current state of data quality management and on this basis, preliminary quality criteria were formed and published in early 2021. The piloting was undergoing in two separate phases during 2021. In autumn 2021, a public official consultation round gave 71 responses including good feedback as well as opinions and suggestions for the implementation of these common criteria.

The quality criteria with metrics for public sector will be launched in March 2022. During 2022, also requirements and a couple of core models supporting the implementation at national as well as organisational level will be planned and concrete data quality evaluation -examples utilising the metrics will be published.

The purpose is to describe the quality of data in a uniform manner. With the help of the model developed in the project, government agencies will produce together, in the longer run, more easily utilised and higher-quality data for public data resources. The quality criteria and metrics will be published as a public administration recommendation.

A large group of maintainers of central government registers and users of register data have been participating in the project. In addition to Statistics Finland, the main participants include the Digital and Population Data Services Agency, the Social Insurance Institution, the National Land Survey of Finland, the Finnish National Agency for Education, the Finnish Patent and Registration Office, Finnish Customs, the Finnish Institute of Occupational Health, the State Treasury, and the Tax Administration. Other actors, for instance municipalities, research open data as well as data security experts, have been consulted during the project.

##### Ireland

The CSO of Ireland has been actively promoting the ‘National Data Infrastructure’ (identifiers for people, addresses and businesses) since 2016. The scope of this work will expand to include classifications. While quality has been a key focus for CSO in the past decade, rolling it out externally seems like a different challenge. In fact, external Departments often respond more positively to CSO quality support than internal business areas. Our rollout of process mapping and other quality support for the NSS was considered to be an ‘innovation’ in the recent ESS peer review. We have offered methodological support externally in recent years which may eventually require more resources in our methodology team and will require them to become more externally focussed.

## 5.2 Possible internal ‘core’ roles/tasks associated with data stewardship

### 5.2.1 Generic roles required to implement stewardship and data management in NSOs

161. Six key roles can be identified to fulfil all the requirements of good data governance within an organization, in addition to the traditional statistician/data analyst role.

- 1) Data Steward

- 2) Data Acquisition Lead
- 3) Data Engineer/Research Software Engineer
- 4) Statistical Product Owner
- 5) Domain Data Manager
- 6) Corporate services supporting Data Stewardship

162. These roles are described in more detail below. The Data Steward has to work closely with the staff in the other roles. Individuals can also hold multiple roles. The concepts and associated actions must be clear to everyone involved in working with data, as the Data Steward will be involved in all of the other roles described here in a 'spider's web' – hence the reason why the roles are described. In different NSOs, those roles can be assigned and distributed in different ways, here we are aiming to give a generic description.

163. Knowledge and expertise on data stewardship are often fragmented at NSOs. Ideally it should be a centrally organised role, with a mandate (responsibilities and skills) as steward of the data (including meta-data). The Data Steward would not be guarding the content, because that is done by the statistical product owner, but guarding the processes surrounding the data and meta-data. Data stewardship is comparable to the role of a group controller with division controllers. So a central data steward is required, but also decentralised stewards who are close to the line employees. It may even be the case that the statistical product owner or domain data manager is also a data steward. The data steward makes policy, disseminates it and gives requested and unrequested advice.

164. A formal approach to data management is especially important when data is collected outside of the NSO since metadata need to be much more comprehensive to indicate issues likely to affect quality in the collecting organisations. These include mandatory/statutory variables, collection channel (online, face-to-face, etc., is the data used in the organisation, default values, etc.). This role is critical as NSOs seek to move away from primary data collection towards using more secondary data, since statistical product owners need additional safeguards to use data sources that are not directly under their own control. As NSOs seek to reuse privately held data in future, the specialized skills of the data steward will become essential to the dialogue with private sector organisations.

### 5.2.2 The role of Data Steward

165. Data Stewardship is an evolving role within NSOs, and will depend to some extent on the level of involvement of the NSO in the National Data Ecosystem. The 'internal' Data Steward within the NSO may also be a significant player in the Department's external role, given the knowledge and networks the Steward will develop over time. For the moment, the role of the Data Steward is set out here in terms of minimum activities and responsibilities in the organization, which are common to most organisations who acquire and process sensitive data. The NSO context is further explained later by relating the Data Steward to other roles in the organization based on GSBPM and RACSI matrices. The 'maximum list' of activities and responsibilities will depend on the degree of external engagement in the NSO.

166. The list of minimum activities of a data steward within the NSO oversees includes being responsible for:

- Data description (data catalogue, data dictionary)
- Quality of the metadata (the responsibility of the data lies with the data owner)
- Data life cycle management
- Data ethics (the Data Steward is the connection between the data management officer/Data Protection Officer and the Ethics Committee)
- Overseeing data security, protection and confidentiality
- Data audit (monitoring the use of data)



167. The responsibilities of a data steward within an NSO are related to:

External/internal networking	Data management/technical	Ethics and legal
Supporting acquisition of new data and using new data sources	Maintaining meta/classifications codes	Link to legal services
Liaising with external providers for metadata and data life cycle	Data life cycle management	Management/compliance of Data Governance Model
Liaising internally for quality	Monitoring 'once only' principle	Developing rules for acquisition of data
Making internal and external data available based on FAIR principles	Overseeing quality and security processes	

Box 5. Examples:

**CBS Netherlands:**

The roles listed under minimum activities for a data steward are distributed throughout the organisation. The purpose of the data steward at CBS is to make the connection between everyone in the production lines and ensure that agreements (policy) are observed.

**CSO Ireland:**

The activities are more centralised in CSO, as much of the administrative data it holds is managed in the Administrative Data, Governance and Analysis Division, though some administrative and other secondary data remains outside the Division. CSO also have a Data Office which maintains processing registers and advises data owners on legal issues. Our internal data strategy in development at the moment with a strong focus on skills and roles, and the 'data engineering' function envisaged is very similar to the RSE role identified by CBS (described next).

### 5.2.3 Other related roles

168. **Data Acquisition** as a process is the exploration and acquisition of new (sometimes big) data sources (from public or private sources) for the purpose of usage in the statistical process, both for improving existing statistics and developing new statistics.

169. The person(s) responsible for Data Acquisition coordinate, organise and facilitate the data acquisition activity for a specific (group) of new data sources. It may not be one centralised function but carried out by many Data Acquisition Leads directly where concrete needs for a specific new data source trigger the process. Typically, Data Acquisition Leads may come from statistical subject domain departments or from departments tasked with research and innovation. Responsibility for acquisition for a specific (group of) new sources lies thus with those teams. In the context of Privately Held Data there is an increased focus on 'reuse' rather than Data Acquisition or ownership. In this context, the Data Steward becomes much more central to acquisition, 'holding' and usage of such data as the networking, management and legal skills described earlier for that role are required to negotiate ongoing access.

170. Responsibilities of a Data Acquisition Lead include organising and coordinating:

- the definition a specific data need, together with both the internal and/or external users
- leading the team for the acquisition activities for a specific data source
- the search for, and evaluation of potential new data source for a specific data need, including an assessment of necessity and proportionality



- liaising between the various stakeholders (such as business, legal, technical and domain experts), both internally at CBS and external (such as data owners and collaboration partners) to the end of acquiring the targeted data source

Box 6. Examples:

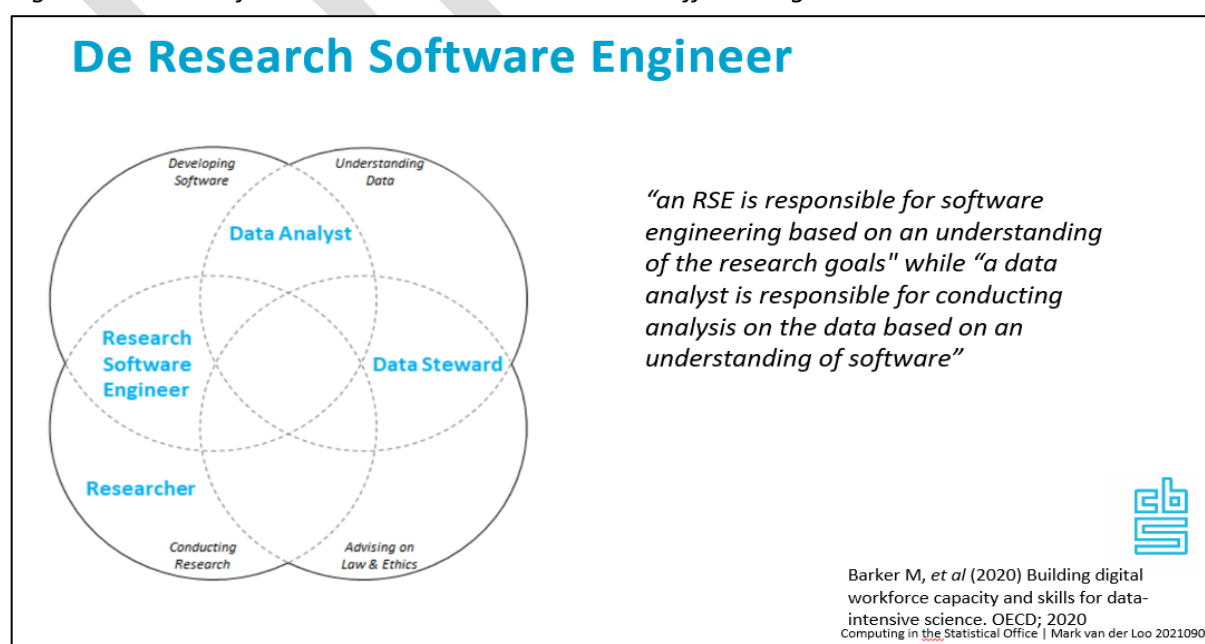
In the Netherlands, strategic and policy questions that concern acquisition issues that go beyond individual activities (for a specific source) are coordinated by the CBS Data Acquisition Community. In Ireland, no such group exists in CSO at the moment; more direction for the Data Acquisition Lead role and relationships with other key roles such as legal and data stewardship/governance will be provided in the upcoming Data Strategy.

171. **Data engineering** (or Research Software Engineer in the CBS example), similar to data stewardship and acquisition, is a new speciality in NSOs which brings together aspects of technology and statistical skillsets, with a stronger emphasis on technology or analytics training. Technology experts in NSOs need to have a strong awareness of business requirements to support it effectively in an agile approach. The main activities of the data engineer include operational aspects of the data stewardship role as well as supporting business areas to produce statistics from secondary sources;

- Maintenance of source level data description (data catalogue, data dictionary, data life-cycle)
- Provision of data extracts to data analysts in line with research goals

172. Data engineers can be located with the data steward or in business areas, and ideally are separate from the software engineering and other technology functions in the NSO. The figure below<sup>11</sup>, from a CBS presentation on the Research Software Engineer role, shows where and with whom the Data Steward works. The locations of the statistical product owner and the domain data manager (again described later, in the context of the GSBPM) are missing from the image, but are logically in the middle. A data steward is responsible for data and needs to have a thorough understanding of the legal and ethical ramifications of its use, an RSE is responsible for software engineering based on an understanding of the research goals, and a data analyst is responsible for conducting analysis of the data based on an understanding of software.

Figure 5. Position of data steward in relation to other staff working with data



<sup>11</sup> [Building digital workforce capacity and skills for data-intensive science | en | OECD](#)

173. The **statistical product owner** (SPO) has responsibility for a set of outputs and products related to a statistical domain. This role corresponds to the 'Analyse' and 'Disseminate' GSBPM overarching processes. 'Evaluation' is also a key function for these Divisions. Data analysts work in the product owner teams.

174. The **domain data manager** (DDM) is responsible for data collection and integration activities within a statistical domain. These activities correspond to the 'Collect' and 'Process' GSBPM processes. There are a variety of 1-1 and 1-many relationships between DDMs and SPOs in the organisation. DDM managers have responsibility under the RACI model (responsible, accountable, consulted, and informed) for all transformed external data and primary data related to a statistical domain. A combination of data analysts and data engineers work in the DDM teams, while data engineers may also be centralized within the data steward's team.

175. These two overarching roles are related to data stewardship and the processes of GSBPM below, based on the Irish example;

Figure 6. Responsibilities according to the GSBPM processes

<= Shared, involving Data Steward =>	<=DDM=>	<=SPO=>
--------------------------------------	---------	---------

176. Alternatively, the role of the Data Steward can be presented as part of a RASCI (Responsible, Accountable, Supported, Consulted, Informed) matrix based on the GSBPM processes;

GSBPM process;	DS	DDM	SPO
Specify needs	S	R	R
Design	S	R	R
Build	S	R	R
Collect	I	R	C
Process	I	R	C
Analyse	I	C	R
Evaluate	I	C	R

177. New forms of data stewardship, based on data reuse rather than ownership through the processes, may result in more responsibilities being vested in the Data Steward over time.

178. **Corporate services** supporting data stewardship include the data protection officer, legal department, internal and external communications, ethics committee, human resources, etc.

## 5.3 Skills required for Data Stewardship

### 5.3.1 Human resources requirements for the data steward role

179. The potential scope of the external and internal data stewardship roles described above requires a new approach to human resource practices in NSOs. The statisticians and data engineers increasingly come from analytics backgrounds. It is often not well understood how complementary the analytics and statistician skillsets are in the public sector. However, there are also areas of expertise in

the public sector which are complementary skillsets and where we should engage in partnerships rather than becoming involved in recruitment, training and secondments. These issues are described in this Section.

#### Box 7. Example

##### Ireland

The CSO believes that the centralisation of the recruitment and training of all Statisticians/Data Analysts for the Civil and Public Service within the CSO for subsequent secondment in a managed and strategic manner would support the systematic implementation of common skills/standards on data related matters across the public service. This will need to be balanced against CSO's own staff requirements, geographical considerations and the higher level of 'churn' experienced by seconded statisticians in the public service. All staff would spend a period of between 3 and 6 months within the CSO receiving training and working within business areas before secondment across the system, from where they would continue their Continuous Professional Development through CSO's Statistical Training Framework.

#### *Managerial level skills*

180. The skills required to 'carve out' each NSO's public sector role depend on the level of engagement described in Section 5.1 and in the different national examples.

181. In a system where the NSO has a clear data stewardship role covering a specific area (as described in the Finnish and NZ examples in Section 5.1, the skills required are strategic, managerial and operational at the planning level. The NSO has to convince others on the long-term direction, strategy and value-added that joint efforts will provide. NSO leadership on data issues must have good communication skills and be able to work co-operatively with all participants involved. Also, it is important to clearly state problems and solutions for common decision making, to communicate information on the progress made in the task etc.

182. In a decentralised system where all organisations take care of their own data management according to best principles, the skills required of NSO data leadership are also managerial skills – strategic thinking, the ability to influence large external organisations, communications and networking (also strategy drafting). But operational level skills internally are also required, from data management to data descriptions and data bases. In Sections 5.22-5.33 these roles are described with national examples of roles and required skills in more detail. The NSO leadership will be trying to influence a range of larger, better funded organisations and also smaller, less organised data holders, without any formal defined role which requires very strong networking capability.

#### *Technical level skills*

183. A range of skills are needed to fulfil all the requirements of good data governance from specialists on legal/data protection to the statistician/data analyst level. For the large group of statisticians, the main implications from data stewardship will be the increased level of external focus described in Section 5.1. The skills involved are similar to managerial role, but at statistician/data analyst level rather than managerial.

### 5.3.2 Secondments

184. Depending on the national context, it may not be sufficient to pass down guidelines, classifications, standards etc. from "the centre" and expect that they will be adopted seamlessly without direct support from "the centre". Some Public Sector Bodies may not have the skill-sets available to them to engage with such standards or indeed to interrogate data holdings in a meaningful

way. It may be necessary in these circumstances for NSOs to second out staff with these skill-sets to ensure that standards are adopted and data can be exploited. There are downstream benefits for NSOs where data suppliers can see the benefits from data quality by using the data effectively.

#### Box 8. Examples

##### Ireland

CSO second out statistical services through professional statistical staff. In addition to working on the compilation of Official Statistics and the development of administrative data sources for statistical purposes our seconded staff will and can be asked to use their skills to support administrative activities. Many of the CSO seconded statisticians (35 in 12 Public Sector Bodies) have been new hires in recent years since the option has become unpopular among experienced statisticians, who were hired before the financial crisis and are well settled in their CSO roles. New hires with experience have often been seconded while new graduates up to PHD level have been retained in CSO. The seconded statisticians tend to do well in senior statistician promotion competitions as they bring very interesting experiences back to CSO. We have four senior statisticians on secondment, and most have been appointed from general senior statistician competitions.

##### Netherlands

CBS (Netherlands) currently has 2 staff seconded to other Departments for data-driven work to pass on expertise held in CBS.

### 5.3.3 Partnerships

185. Skills to cooperate, create partnerships and communicate with partners are crucial in the data stewardship role, especially if DS role extends outside of the NSO. This requires good negotiation and communication skills and also some legal expertise if the partnership is documented in Memorandum of Understanding or other agreement formats. New partnership arrangements for Privately Held Data access will need to be developed by NSOs, requiring deep understanding of privacy and commercial sensitivity of such data sources.

#### Box 9. Examples

##### Ireland

CSO has engaged in a number of strategic partnerships in recent times with agencies that have complementary skills. The most successful partnership recently has been with the national mapping authority in Ireland, Ordnance Survey Ireland (OSi). Our work until March 2020 was focused on collaborations involving our complementary skillsets, such as SDGs, Census mapping and near-Earth observation data. The partnership was subsequently the entry point for our work to support the COVID crisis in Ireland, as the OSi offered (with our support) to develop a dashboard to support decision makers in the health sector with real time data on infections, deaths and hospital activity. However, it quickly became apparent in midMarch as the crisis took hold that CSO's involvement would need to be much broader than this to be an effective support to the health system. Openness to partnerships is encouraged by the Irish Government, and also can open other doors based on this example.

## 5.4 Services outside of the NSO mandate that could potentially be facilitated by NSO stewardship

186. It should be recognised that other Public Sector Bodies have a different set of requirements within their mandates for administrative data than NSOs. Specifically, they usually collect their data for e.g., operational measurement or policy/programme development and evaluation, which is not compatible with NSO's core values confidentiality and independence. This should be taken into account in analysing how our data stewardship role could potentially interact with their non-statistical stewardship requirements if we are to influence effectively for data access and quality.

187. Non-statistical PSB requirements that fall outside of the NSO mandate are summarised here as considerations for ‘what do we do and what do we facilitate’ – a key question in data stewardship. While NSOs will not be directly involved in this analysis, if they do not provide support they may lose some influence in the public sector since it will either be ‘contracted out’ or done sub-optimally with fewer benefits for data quality.

#### 5.4.1 Business Intelligence

188. Classifications and data services (mainly visualisations) are crucial for effective Business Intelligence implementation in PSBs. However, there are some clear distinctions with normal statistical production. The audience for the reports is usually ‘in house’ and self-service, so there is less interpretation needed of the data. The metrics provided are based on combinations of events which are supposed to represent progress or otherwise in operational performance. Clearly processing times for e.g. social welfare offices vs. hospitals involve a different combination of events, so business knowledge is required to produce accurate metrics. Also, business processes can change over time so ongoing business interaction and understanding is needed to ensure that metrics do not become misleading. In summary, this is an activity that we can only do via ad-hoc ‘pathfinders’ but we can facilitate via the data services described earlier and our secondment programme. Also, any improvements to data for statistical purposes, such as identifiers and classifications, are also strongly supportive for BI analysis.

#### 5.4.2 External ‘data engineering’ role

189. Our seconded statisticians tell us that relationships with technology colleagues are crucial to effective analysis. While this relationship is second nature in CSO, it is not always obvious to technology experts in PSBs how they should support data analysis, since data integrity and security are paramount in their work. The best approach to this interface is a collaboration as described in the RSE/‘data engineering’ role, where data is made available to analysts via warehouses and data quality and potential are assessed by both sides in conjunction with business areas. While this is clearly something we do effectively ‘in-house’, there is also an important facilitation role that we can play in PSBs based on the data services described earlier and training for our seconded staff.

#### 5.4.3 Operational analytics

190. Operational analytics involve a range of activities in PSBs such as operations research, fraud measurement/detection, segmentation of clients requiring additional supports and financial management, often based on similar regression techniques to those used for programme evaluation. While these activities are clearly not compatible with the CSO’s mandate, there are some indirect supports that we can provide via the data services described earlier and our secondment programme. Also, any improvements to data for statistical purposes, such as identifiers and classifications, are also strongly supportive for operational analysis.

#### 5.4.4 Programme evaluation and related econometrics

191. Modern programme evaluation attempts to control for background factors to determine whether participation in a programme is beneficial or cost-effective for clients or not. While there is a strong scientific basis to this analysis it also requires some subjective decisions to decide what needs to be controlled or otherwise to obtain a clear signal about a programme’s value. As with operational analytics, this is not compatible with the CSO’s mandate, but in addition to the supportive services identified for analytics CSO can also provide ‘data hub’/researcher access services and partnership options for programme evaluation, since the results involved are generally aggregates.

## 6 How to communicate data stewardship?

192. This chapter presents how the complexity of data stewardship in terms of semantics, differences across countries' data ecosystems and different understanding of the roles of data stewards can be approached through communication. The chapter includes some examples and success stories as regards communication of a holistic approach to public data governance and data management across countries.

### 6.1 A complex concept – how to communicate it?

193. Over the last decades, the environment where official statistics has been operating is changing rapidly. It can be described as heterogeneous because of proliferation of new data sources and data stakeholders. Data ecosystems – expanding the notion of national statistical systems – represent an inclusive concept of interacting data producers and users whose common characteristic can be paradoxically described as “difference and diversity” in all data-related activities: production, exchange and use.

194. Users not only search for reliable information, they more and more often tend to search for a fast one. They frequently are not interested who is at the beginning of the information chain, i.e. who stands behind data. Instead, they want to have a quick access to an extensive and comprehensive information, in an easily approachable format, accompanied with hints how to interpret it, so called “data stories”. Users want to be “assisted”, treated with attention and care. They also want to be guided in the maze of data, standards, research and definitions. Data stewards can be expected to help with all those features.

195. When talking about the data stewardship, the first aspect that comes to mind is purely semantic: what does it mean? The current report aims to provide a common definition (see Chapter 3). However, how the stewardship role is implemented in countries in reality looks very different due to the national contexts. There is no ‘one size fits all’. An added difficulty is the need to translate the term into national languages. Sometimes the same term is used as for ‘data governance’ but this does not convey the important differences between ‘data stewardship’ and ‘data governance’ (see Section 3.2).

196. A practical solution to communicating data stewardship effectively can be to focus on what is the purpose of data stewardship, who performs the related activities and what they concretely are. The communication can also concern new products and services that data stewardship engenders and inspires.

197. Failing to communicate the role properly brings about considerable risks. For example, lack of specifying who is responsible for collecting and storing data may result in data gaps. The collection of the same data by different institutions can lead to excessive burden on respondents and budgets. Lack of understanding of data may lead to data misinterpretation or misuse. Failure to determine deadlines for data transmissions may result in an inability to make decisions in appropriate time. Those risks can be reduced by well-balanced and targeted communication.

198. The chapter will demonstrate the variety of ways to communicate the said aspects. Communication of data stewardship, bearing in mind the ambiguity of the concept, should be action-driven and focused on the usefulness of this approach for the benefit of different audiences, and for strengthening the position of NSOs who often are the ones assuming the stewardship responsibilities.



## 6.2 Possible messages for communicating data stewardship

199. The official statistics has rightfully recognised the **importance of communication of its products and roles in the society**. In the times of digital revolution, the increasing role of social media, disinformation and different actors on the information market, official statistics has to reposition itself leaving behind its long-time “dissemination-oriented” nature. Communication is more user-oriented and concerns the ability of getting through the message and being understood. The set of values<sup>12</sup> and principles<sup>13</sup> that official statistics is based on constitute a distinctive factor vis-à-vis other data producers. Although for official statisticians this distinction is a reason for valuing official statistics higher than information by other producers, a characteristic of today’s communication market is that “the value, as much as beauty, remains in the eye of the beholder”<sup>14</sup>. Therefore users should be the focal point in communicating and promoting the value, relevance and trustworthiness of official statistics.

200. Data stewardship-related communication may focus on the **roles which data steward can or should play in the data ecosystems**, related to data governance and coordination, acting as data curator<sup>15</sup>, and assisting and facilitating data use. The communication **can also focus on challenges** which data stewardship can help to address in national data systems. An example of such case is the joint European Statistical System (ESS) communication to private data holders (discussed further on), defined as one of the pivotal targets of a unified ESS strategic communication approach.

201. A **core message** could revolve around “officialising and enabling the roles assigned to the national statistical offices within a constantly developing data ecosystem, aiming at creating public value beyond official statistics by reusing data for the public interest in a systematic, responsible and sustainable manner”. Such message features a three-fold approach that is:

- a) systematic i.e. beyond experimentation and referring to methodological framework and quality standards;
- b) responsible i.e. ethical use, privacy preservation and protection of confidentiality; and
- c) sustainable i.e. ensuring continuity, consistency, partnerships, etc.

202. The setups of data ecosystems in countries are different and NSOs are not entirely free to decide about their role. The role of data stewards – irrespective of strategic vision and positioning of an NSO – can sometimes be reduced to data dissemination in a national data ecosystem. In those cases, users need proper communication and additional services, such as metainformation or user guidance. It may also happen that users may request data that would compromise statistical confidentiality, one of the Fundamental Principles of Official Statistics. In this case users must get a clear message that some information is public, and therefore not protected, and other data have a status of statistical information and therefore covered by statistical confidentiality. When data are already publicly available outside official statistics, the statistics obtained from those data shall not be considered confidential when disseminated<sup>16</sup>. The imposition of data confidentiality in such a case may have a reverse communication effect. Here a strong coordination role of NSOs is vital.

---

<sup>12</sup> See the works of the UNECE TF on Core Values of Official Statistics.

<sup>13</sup> Starting from the [UN FPOS](#), through all the ethical codes and guidelines stemming from them.

<sup>14</sup> See the report of the UNECE TF on Measuring Value of Official Statistics. The TF has been considering the issue of how to measure the value of official statistics and how to communicate this value to both users and the wider public, putting forward a thesis that the inherent value of official statistics is ultimately decided by users only. They will attach value to what is relevant, timely and meets their needs.

<sup>15</sup> See the roles of Data Steward described in Chapter 5. Responsibilities and skills for ‘Data Stewardship’.

<sup>16</sup> Art. 25 of the regulation 223/2009 on European statistics  
<https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02009R0223-20150608&from=EN>



203. NSOs are naturally **well positioned to be data stewards because of their coordination role** – having a general overview of standards, classifications (and being responsible for their implementation), as well as practical know-how how to transfer this knowledge to other players in the data ecosystem. This should be always accompanied with adequate communication.

204. At the same time, data stewardship functions should not compromise the core values of official statistics, preserving professional independence and complying with the Fundamental Principles of Official Statistics, the European statistics Code of Practice, as well as maintaining the trust and confidence of users. Communication with stakeholders, especially users, is essential in this regard.

205. The European Commission Group on Business-to-Government Data Sharing in its report titled “Towards a European strategy on B2G data sharing for the public interest”<sup>17</sup> values a “recognised data steward function [that] should be created and promoted in both the public and private sectors”<sup>18</sup> as “a key success factor in setting up sustainable and responsible B2G partnerships”<sup>19</sup>. In addition, it makes a meaningful distinction in terms of communication:

*“The use of the term ‘data steward’ should not be confused with the term as used in the context of data management/governance, where it has mostly an internal focus. While the technical data-related tasks of both functions are similar (data quality, standards, metadata, etc.), a data steward in the context of B2G data sharing has additionally an external function for data collaborations and a coordinating role to align all stakeholders around the objectives of the data collaboration. DAMA-DMBOK, Data management body of knowledge, Technics Publications, New Jersey, 2017”.*<sup>20</sup>

206. The mentioned “coordinating role” leading to enhancing collaboration between all the parties involved in data sharing process is often put at the centre of communication strategies and activities focused on data stewardship in the external context. This can be strengthened by going beyond the national framework and establishing a network or community of data stewards.

207. A crucial aspect in communication are the **target audiences**. This is true especially in the context of data stewardship which involves many actors/stakeholders dealing with different aspects of data. In drafting and applying data stewardship-driven communication strategies and plans in national contexts, a well-known notion of **personas** might be helpful<sup>21</sup>. Personas allow to perform an in-depth user/stakeholder analysis to get a better understanding of their specificities and adapt statistical products and services to them, enhance partnerships, and engage them, which is particularly important in the context of data stewardship. The information on stakeholders collected through regular interactions can be analysed qualitatively and quantitatively. The sources for this information can be a user support system, website analytics, feedback from satisfaction surveys etc. Based on the result of the analysis, stakeholders can be categorised into e.g.: politicians/ decision makers (for instance ministries assigning finances for data within the national statistical systems), policy makers (using data), businesses and companies (an important group of so called owners of privately held data sources, which integration into national and supra-national data strategies can lead to a massive expansion in data innovation and data sustainability), media, and a wider public (both using and providing data).

---

<sup>17</sup> <https://digital-strategy.ec.europa.eu/en/news/commission-appoints-expert-group-business-government-data-sharing>

<sup>18</sup> p. 7 of the B2G data sharing Commission expert group report

<sup>19</sup> p. 37 of the B2G data sharing Commission expert group report

<sup>20</sup> p. 38 of the B2G data sharing Commission expert group report

<sup>21</sup> <https://www.usability.gov/how-to-and-tools/methods/personas.html>

208. For effective relationship management and an overview of stakeholder groups, an integrated customer relationship management system is recommended. Such a system consolidates different interaction channels and key information on stakeholders, to better identify their evolving needs and expectations. When the personas to whom we communicate are established, the key messages for communication can be developed.

209. Finally, the communication of data stewardship may be focused on new products, new services and new partnerships which are generated having in mind data stewardship approach. The following section enumerates a few of such cases.

### 6.3 New products, services and partnerships as elements of data stewardship – seen from communication perspective

210. Implementing data stewardship can be linked with **innovation**: new products, new services and new partnerships for NSOs. An example of a service-related data stewardship undertaking of an NSO can be a Public Services Monitoring System (SMUP), developed by Statistics Poland (see box 10)<sup>22</sup>.

211. **The Sustainable Development Goals** can be cited as another case where NSOs play data stewardship role and communicate accordingly. The leading role of official statistics in SDGs monitoring was emphasized in the 2030 Agenda. In many countries NSOs became institutions responsible for SDG<sup>23</sup> data management, coordination of their collection, reporting and communication.

212. The role of statistics in the context of SDGs goes beyond sharing data, it often involves launching SDG National Reporting Platforms, which serve as national knowledge banks about sustainable development. Data stewardship approach manifests itself in the technical aspect – NSOs invest in API development, so that it is possible to view and download the full range of data and metadata for further processing. They assume also an education function, presenting users with wide range of methodological information and interpretative guidelines. The publications often adopt a modern interactive form, along with adjusting the content to a wide range of readers and different target groups, which allows to attract new data users. A holistic approach to communication in statistics for SDGs is manifested in NSO's presence at events dedicated to the SDGs on the national and international levels, next to the government, academia and NGOs.

213. Building the image of the institution responsible for statistical information and knowledge as the main communication message frequently results in initiatives that go beyond standard statistical cooperation. This corresponds to re-defining their place in data ecosystems thanks to their numerous actions as data stewards in the area of SDGs. These initiatives tend to focus on the specific needs of data users and enable the potential of data for practical use. An example of such an initiative can be the participation of Statistics Poland in the development of the "Impact Barometer" for Polish business (see Box 11).

---

<sup>22</sup> In Poland, there exists a legislative enabler attributing to the President of Statistics Poland certain functions of data steward. The Act on Official Statistics obliges him to prepare - in consultation with the corresponding central public administration agencies - standard classifications and nomenclatures, which are relevant for describing economic and social processes.

<sup>23</sup> See the works of [UNECE Steering Group on Sustainable Development Goals](#)

#### Box 10. Example

##### Public Services Monitoring System (SMUP) developed by Statistics Poland

The aim of the project is to develop and implement a universal, publicly available on-line system for monitoring public services, providing integrated, high-quality data that will allow to optimize the process of providing the mentioned services. The needs are defined by the local government administrations.

The project stems from a lack of such a service - highly needed by citizens who want to be able to assess and compare quantity, quality, availability and financial efficiency of public goods and services. The Project is conducted in cooperation with the Ministry Internal Affairs and Administration and organizations of local self-government – Association of Polish Cities and Association of Polish Counties. This format of the project implementation is unique – it requires close communication between government administration units, local government organizations and NSO. It is also necessary to closely cooperate with end-users and data providers at every stage of the project. A big gain of the project, besides creating a user-centred and user-friendly product/ service, is also a new model of partnership with local government administration who is at the same time data provider and data user. It can be translated in a joint defining of needs and joint design of tailored indicators to measure and monitor public services, which in turn lowers the local government administration's burden (as currently they have to report to various public administration institutions) and increases availability and usefulness of data collected so far in various structures and systems (registers) by different data owners (institutions of public administration).

Approaching this undertaking from a communication angle, there is an interesting distribution of tasks between stakeholders. The Ministry of Interior and Administration is a partner in the project. It implements a communication strategy from a practical point of view and is mainly responsible for promoting project deliverables with other users, especially through training courses to enhance knowledge of users. But the NSO is responsible for the conceptual and substantive part of communication, defining the strategy and focusing on promotion of the methodology of newly-created indicators. The main communication message is compliant with the slogan 'direction better services' and underlines the role of official statistics in providing better, more innovative solutions for users ("SMUP provides the necessary data so that everyone can benefit from better and more efficient public services"). The main messages of communication are e.g.: "public services concern all citizens, so better delivery is beneficial for all"; "public services amount to an annual expenditure of over PLN 250 billion. A 1% improvement in efficiency may translate into PLN 2.5 billion in savings"; "SMUP is a large, unique database of new indicators"; "data reported by municipalities and counties are returned to them through the SMUP as useful information for benchmarking"; "improvement of services is possible primarily thanks to reliable data guaranteed Statistics Poland". These messages clearly show how official statistics is assuming the role of data steward in practice, not only through coordinating and managing data to provide new innovative products, but also through complex and extensive guidance and education of users.

#### Box 11. Example

##### Participation of Statistics Poland in the development of the "Impact Barometer" for Polish business.

Resulting from cooperation between representatives of science, the private sector, government administration, non-governmental organizations, the Impact Barometer is a set of indicators thanks to which Polish entrepreneurs can assess their individual impact on achieving the goals of the 2030 Agenda. In the process of building the tool, Statistics Poland acted as a data steward and provided expertise in the selection of appropriate SDG measures adapted to the specificity of the business.

214. NSOs may also consider **joint communication by several countries** experiencing the same challenges. Communicating as a system of partners that address those challenges can make their voice stronger. See an example of a joint ESS communication work in respect to data held by the private sector in Box 12.

#### Box 12. Example

The *European Data Strategy* aims to create a single market for data which will allow data to flow freely within the EU and across sectors for the benefit of businesses, researchers and public administrations. Two legislative proposals have a special meaning for the official statistics:

1) the *Data Act: Proposal for a Regulation on harmonised rules on fair access to and use of data*, which “aims to maximise the value of data in the economy by ensuring that a wider range of stakeholders gain control over their data and that more data is available for innovative use, while preserving incentives to invest in data generation”. A joint NSOs approach in this field has been explored at the policy level - by means of the ESS position paper on the future Data Act proposal (from July 2021).

2) the *Data Governance Act* proposal which will create a mechanism to enable the safe reuse of certain categories of public-sector data that are subject to the rights of others. This includes, for example, trade secrets, personal data and data protected by intellectual property rights. In other terms, the latter defines data-steward-like functions, without using the term.

Joint communication activities on this topic were implemented by the Task Force on the ESS Strategic Communication. The group issued a common communication strategy, featuring communication on the need of access to privately held data seen as urgent and necessary for producing new, faster, more detailed official statistics. The ESS NSOs have also developed a concrete, translatable and reusable communication package, built on common key messages to three target audiences: policy makers, private data holders, and EU citizens (general public).

A sub-group of the Task Force created an inventory of partnerships for access to privately held data and related communication actions. Based on the key messages and insights from the inventory of partnerships, the subgroup then prepared a ‘communication toolkit’ – a set of communication material for three key target audiences – policy makers, private data holders and citizens.

215. Data stewardship roles and functions assumed by NSOs are often operationalised in a form of **data strategies**, describing how NSOs use and manage data while providing the best standard of statistical information for the public. The existence of such strategies can be a strong asset of communications, arguing the cause of data stewards and explaining to stakeholders how the information is processed, protected and what standards it is based upon. Data strategies are in place to guide NSOs data-related actions and processes, as well as to ensure the stakeholders and users that the best information, based on solid analysis and standards, is being provided to users, within a comprehensive framework, with a strong concern about safeguards for data. Data strategies usually encompass such aspects like data collection, data use and re-use, data dissemination and communication, as well as data protection and security. They translate vision and mission into practice by setting strategic objectives of NSOs – which is vital with the data stewardship-driven approach, because those objectives often concern nothing else than the roles and functions of data stewards, i.e. data integration, standard setting, enhancing metadata, spreading solid ethical principles and frameworks within data ecosystems.

216. Other elements supporting NSO’s communication in the context of data stewardship may be **experimental studies** (frontier statistics, non-official statistics). The statistical products and services based on using new methods, modern analytical tools and new data sources are often communicated as a separate, additional activity of NSOs. These types of products and services can complete a picture

of data stewardship in a national context enhancing NSOs role in data ecosystems and underlining new partnerships.

217. As partnerships for better statistics, a frequent case of **cooperation of NSOs with science and academia** can be cited. Partners from academia are usually responsible for co-developing methodological and data processing-related solutions to be implemented by NSOs who are acting as data stewards. An undeniable gain of such partnerships are original, advanced solutions, ready to be implemented in statistical production. This may be described as one step further than the experimental research. Opening up to cooperation with science means also gaining new perspective on research problems, which - when properly communicated - can contribute to a better, stronger image of an NSO thanks to reducing respondents' burden or widening the scope of offered products and services.

218. Another important aspect when debating on communication of data stewardship is the function of data stewards focused on broader aspect of **data governance**, connected with fighting disinformation by introducing robust ethical frameworks (stemming from the UN Fundamental Principles of Official Statistics) to a wider community of data stakeholders within the ecosystems. In the times of disinformation, the place of official statistics and data is even more crucial. Data stewardship with its more inclusive aspect may become an opportunity for ensuring trust and understanding trusted data sources – improving data literacy through communicating ethically and conveying true message in the world full of fake news.

## 6.4 Conclusion

219. As long as data stewardship may manifest in different ways and have different implications depending on the local context and the specific data community, communication plays an essential role in explaining those implications to various target audiences. Data stewardship can be communicated better when translated into a series of concrete products, services or partnerships – stemming from the holistic data governance, data management and data communication approach which is being adopted by NSOs in their data ecosystems. However, developing a common framework of data stewardship is beneficial for better communication of the roles that NSOs can play as data stewards, to ensure that data are better understood, better used and better shared. To be sure that data add value to users' lives.

## Annex 1. Glossary of data and information stewardship

The below definitions have been developed leveraging the materials from NSOs, international and intergovernmental initiatives' and organizations' resources (ISO, OECD, EUROSTAT, UNECE, SDMX, etc.) and academic publications.

### Administrative data

Administrative data describes information collected by persons, organizations or departments of government for their own purposes and subsequently communicated to NSOs. It is considered derived from an administrative source (Statistics Canada, 2016).

### Classification

Set of categories which may be assigned to one or more variables registered in statistical surveys or administrative files, and used in the production, collation and presentation of data. Categories are discrete, exhaustive, and mutually exclusive of all objects/units in the population of interest. They are defined with reference to one or more characteristics of a particular population of units of observation. A statistical classification may have a flat, linear, or hierarchical, such that all categories at lower levels are sub-categories of categories at the next level up (OECD, 2006; StatCan, 2021b, 2021c; SDMX, 2018).

### Confidentiality

Data confidentiality is an obligation to the provider of information (a physical or legal person) to maintain the secrecy of that information. It is a property of data (as opposed to privacy, which relates to data subjects), which indicates whether they are subject to dissemination restrictions and is usually based on legislative measures, preventing it from unauthorized disclosure. Confidentiality denotes an implied trust relationship between the person providing the information and the individual or organization collecting it. (OECD, 2005; StatCan, 2016; SDMX, 2018; UNECE CES, 2000).

### Data

The, often numerical, representation of information, in a manner suitable for storage, communication, interpretation, or processing by human beings or by automatic means, and from which knowledge can be drawn, including structured or unstructured forms. While there are many different types of data (digital or hard-copy data, qualitative, quantitative, etc.), **statistical data** refers to data used to produce statistics (often from a census, survey or administrative source). (OECD, 2006, 2021; StatCan, 2019, 2021b; UNECE CES, 2000).

### Data accountability

The ongoing liability and responsibility of an individual, enterprise, organization or government regarding the management of data. Data accountability ensures that appropriate technical and organizational measures and records are in place to demonstrate compliance to data related principles and policy instruments. Examples may include data protection policies, impact assessments, maintaining necessary documentation, data security measures, roles and authorities related to data protection, privacy management frameworks, and appropriate recording and reporting (Information Commissioner's Office, 2019; StatCan, 1998).



## Data audit

The process and methods of examining each phase of the data lifecycle (collection, storage, analysis, and dissemination) to assess data quality or utility, involving the assessment of key metrics to draw conclusions about the properties of a data set. Examples include internal review, statistical confidence tests, internal audit, audit by outside accountants, cross-checks with other macroeconomic accounts, etc. (OECD, 2013).

## Data catalogue

A data catalogue describes the contents of a database or a data holding, including information about the format, definitions and structures of the tables. It is a detailed and organized inventory of data and metadata assets, identifiers, and descriptions of associated items included in a register to enable the management, discovery, and retrieval of data and information. It helps data professionals collect, organize, access, and evaluate fitness of data for intended uses to support data discovery and governance (DDI Alliance, 2021; OECD, 2005; StatCan, 2021b; Wells, n.d.).

## Data curation

All the processes, including annotation, publication, and presentation of data, needed for principled and controlled data creation, maintenance, and management, together with the capacity to add value to data over time and ensure that it remains available for reuse and preservation (Miller, 2014; Treloard et al., 2007).

## Data custodian

The role(s) accountable for data assets from a technical perspective. A data custodian ensures data assets safekeeping by focusing on the information technology aspects of data management. This includes data security, custody/storage, accessibility, scalability, configuration management, availability, auditing, backing-up and restoring, standardization, restoration processes, technical standards, and policy/procedure enterprise implementation (OECD, 2005; StatCan, 2021b).

## Data dictionary

A database for the storage of metadata and a systematic approach to managing definitions. A data dictionary is a type of catalogue that describes the contents of a database or data set. Information is listed about each field in the attribute tables and about the format, definitions and structures of the attribute tables (DMBOK, 2017; OECD, 2005; UNECE CES, 2000).

## Data ecosystem

A data ecosystem encompasses data and statistical data, data subjects, along with a broad range of stakeholders, partnerships and data users that are involved in related data access and sharing arrangements, according to their different roles, responsibilities and rights, technologies, and business models. This includes the capacities, processes, policies and infrastructure used to manage data throughout its lifecycle (See the definition for the **Four G model of data lifecycle management**) and maximize its use as a strategic asset. The data governance and data stewardship activities take place in a data ecosystem, and relate to managing the interactions of four main categories of actors in the data ecosystem: data generators, data services, data business users, and end customers (European Commission, 2017; OECD, 2021; StatCan, 2020b).



## Data ethics

The knowledge that allows to acquire, use, interpret and share data in an ethical manner including recognizing legal and ethical issues (e.g., biases, privacy) (StatCan, 2020a). See also **Necessity and Proportionality Principle** and the **FAIR Guiding Principles** of data.

## Data governance

A system of decision rights and accountabilities for the management of the availability, usability, integrity and security of the data and information to enable coherent implementation and co-ordination of data stewardship activities as well as increase the capacity (technical or otherwise) to better control the data value chain, and the resulting regulations, policies and frameworks that provide enforcement. This includes the systems within an enterprise, organization or government that define who has authority and control over data assets and how those data assets may be used, as well as the people, processes, and technologies required to manage and protect data assets (Data Governance Institute, n.d.; OECD, 2001, 2019; Plotkin, 2021; StatCan, 2020, 2021b).

## Data integration

The practice of ingesting, transforming, combining, and provisioning data to create consistent, good quality and usable data from one or more diverse data sets to meet the data consumption requirements. Data integration techniques include bulk/batch data movement; extract, transform, load (ETL); change data capture; data replication; data virtualization; streaming data integration; data orchestration; etc. (IBM Cloud Education, n.d.; SAP, n.d.).

## Data management

A discipline that directs and supports effective and efficient management of information and data in an organization or public administration, from planning and systems development to disposal or long-term preservation. Data management involves the development, execution, and supervision of plans, policies, practices, concepts, programs, and the accompanying range of systems that contribute to the organizational or governmental mandates and to public good, as well as the maintenance of data processes to meet ongoing information lifecycle needs. Data management enables the delivery, control, protection, and enhancement of the value of data and information assets through integrated, user-based approaches. Key components of data lifecycle management include a searchable data inventory, reference and master data management, and a quality assessment framework (DMBOK, 2017; Government of Canada, 20219; StatCan, 2020a, 2020b, 2021b).

## Data lifecycle management

The management of data throughout the data and information lifecycle, which is characterized by four main phases outlined in Statistics Canada's "Four G" model of data lifecycle management (DMBOK, 2017; Government of Canada, 2019; Rancourt, 2019; StatCan 2020b).

*Four G model of data lifecycle management:* Used jointly with the Generic Statistical Business Process model (GSBPM), this model is based on the idea that the main activities in the data lifecycle are gathering data, guarding data, growing data, and giving data back to society in the form of information. It is a metadata-driven system that represents the business processes of National Statistics Offices (NSOs) (Rancourt, 2019). The four Gs (Rancourt, 2019) include:

1. *Gather* – Driven by data needs, this phase refers to all data ingestion, including the collecting and integrating of data assets through various systems of acquisition, as well as the policy

instruments and ethics-based legislative frameworks through which an individual, enterprise, organization, or government gains access to data and information.

*II. Guard* – The protection, pre-processing, and (de)identification of data. During this phase, special attention is paid to access rights and privileges, data audit trails are performed, data monitoring and back-up protocols are systematized and ongoing, and metadata standards and classification systems are consistently updated. The goal in this phase is to adhere to the “privacy by design” principles, ensuring that data is secure and encrypted, confidential, and with all necessary privacy protocols in place to function ethically and according to a robust trust framework.

*III. Grow* – The organization, controlled access, processing, transformation, integration, analysis, and extraction of data for various uses. During this phase, data is cleaned and verified, quality assurance is performed, data is analyzed, explanations are developed, hypotheses are tested, and simulations are leveraged for forecasting. Efforts are made to grow data by ensuring its optimization and adhering to (and continually developing) data quality frameworks.

*IV. Give* – The dissemination of information and provision of hosting services. During this phase, data and statistics are shared and published, data access and interoperability are ensured, dissemination occurs regularly and with quality and accessibility, and the appropriate metadata is made available based on strategic requirements. The goal in this phase is to increase data discoverability and be “open by design”, having sharable and open data, metadata, metainformation, and analysis.

## Data quality

Data quality refers both to the characteristics associated with high quality data (being fit for use) and to the processes used to measure or improve the quality of data. Quality assurance deals with planning, implementation, and control of activities that apply quality management techniques to data (whether statistical, administrative, or otherwise), to assure it is fit for use and meets the needs of data users. Several international organizations have their criteria defining statistical quality. These include: *relevance, credibility, accuracy, timeliness, accessibility, comparability, coherence, and proportionality*, which all contribute to the data and information’s overall quality and value (Eurostat, 2003; OECD, 2002; StatCan, 2021a; Strong and Wang, 1996).

- A) *Relevance*: The extent to which data meets users' needs. The identification of users and their expectations is therefore necessary, because the data should reflect the needs and priorities expressed by the users of the statistical system (See also *necessity* in the **Necessity and Proportionality Principle**).
- B) *Credibility*: the nature of the authority and reliability (reputation) of the data source, the objectivity of the data and its source, and general believability of data. This is determined, in part, by the integrity of the data production or collection process and the known provenance of the data.
- C) *Accuracy*: accuracy is defined as the closeness between the estimated value and the (unknown) true value.
- D) *Timeliness*: punctuality in making open data available and accessible as well as in disseminating results, up-to-date figures, which are published frequently and on time at pre-established dates.
- E) *Accessibility*: availability, documentation, and general clarity of the data and information, for consumption by both people and machines.
- F) *Comparability*: statistics for a given characteristic have the greatest usefulness when they enable reliable comparisons of values taken by the characteristic across space and time. The comparability component stresses the comparison of the same statistics between countries in order to evaluate the meaning of aggregated statistics (see also **Interoperability**).
- G) *Interpretability*: reflects the ease with which the user may understand and properly use and analyze the data or information. May include conceptual or technical aspects.

- H) **Coherence**: the degree to which the data and information from statistical programs or administrative sources and data brought together across data sets or programs, are logically connected and complete. When originating from different sources, and in particular from statistical surveys of different frequencies, statistics are coherent in so far as they are based on common definitions, classifications and methodological standards.
- I) **Proportionality**: refers to the imperative that the data collection effort be balanced with ethical considerations (see also **Necessity and Proportionality Principle**).

(European Commission, 2015; Eurostat 2003; OECD, 2011; StatCan, 1998)

## Data modelling

The knowledge and skills required to apply advanced statistical and analytic techniques and tools (e.g. regression, machine learning, data mining) to perform data exploration and build accurate, valid and efficient modelling solutions that can be used to find relationships between data and make predictions about data (StatCan, 2020a).

## Data security

The measures taken to prevent unauthorized access or use of data. Data security enables the protection of privacy and confidentiality, as well as the maintenance of trust and social license to operate (OECD, 2004; Statistics Canada, 2021a; UNECE CES, 2000).

## Data set

A permanently and highly organized stored collection of similar data, sharing a structure, which covers a fixed period of time and relates to a single subject. Data sets can be represented physically in three forms:

- 1) **Generic Data Set**: this format allows the representation of data structured according to any Data Structure Definition;
- 2) **Structure Specific Data Set**: this format allows the representation of data structured according to a specific Data Structure Definition;
- 3) **SDMX-EDI Data Set**: a specific case of generic using the UN/EDIFACT syntax and which has limitations on what can be represented. It supports time series only (StatCan, 2019; SDMX, 2018; UNECE CES, 2000).

A data set is managed and stored as a unit, for example in the same computer file. The terms database and data set are often used interchangeably (OECD, 2002; SDMX, 2018; UNECE CES, 2000).

## Data standard

Data standards are the rules and specifications by which data are described, defined and recorded. In order to share, exchange, and understand data, we must standardize the format as well as the meaning. Examples of data standards may include data models, reference data, and identifier schemas. The use of statistical and data standards enable the integration of data over time and across different data sources, as well as reduce the resource requirements associated with many aspects of survey development and maintenance (ISO 8000-61:2016; OECD, 2005; StatCan, 2021b).

## Data steward

The role(s) accountable for the management of data assets and resources from a strategic perspective. Data stewards are responsible for ensuring that the data acquisition, entry, quality, interoperability, and overall management supports organization's needs, while also ensuring adherence to social license, legislative, and regulatory requirements. They work with stakeholders and other deliberative or

advisory bodies to develop definitions, standards and data controls, and perform key functions in the ideation and implementation of data policies that are scalable, sustainable, and significant (OECD, 2001,2018; StatCan, 2021b).

### Data stewardship

Data stewardship represents the ethical and responsible creation, collection, management, use, and reuse of data. It is expressed through long-term, inter-generational curation of data assets such that they benefit the full community of data users, and are used for public good. Data stewardship works to support the growing maturity of data policy and is applicable at all scales, from the national or data system level, to the organization or enterprise level, to the individual or dataset. Made visible through a range of internal and external functions associated with stewardship roles - including data access, security, and data quality and standards - it influences proactive and responsible data practice to help deliver data strategies, maintain trust, and promote accountability. Reflecting an appropriate level of maturity, data stewardship is enabled through good data governance and data management, which provide oversight of data assets throughout their lifecycle to ensure their proper care (New Zealand Government, 2020; OECD, 2018; Plotkin, 2021; StatCan, 2020, 2021b).

### Data value cycle (or value chain)

Refers to data-related processes through which value is created with data, including data creation, collection, validation, verification, storage, curation, enrichment, processing and analysis, access, sharing, and deletion (OECD, 2021).

### FAIR Guiding Principles

Set of data principles, which define characteristics that modern data resources, tools, vocabularies and infrastructures should demonstrate to facilitate the discovery and reuse of data by other parties. FAIR stands for:

- F** - Findable and easily searchable
- A** - Accessible and easy to use
- I** - Interoperable and more easily interpretable
- R** - Re-usable data that is easy to share and use (Wilkinson et al., 2016).

### Fundamental Principles of Official Statistics

These [Fundamental Principles](#) provide an international, intergovernmental guide for how NSOs can best provide their services to citizens (2014). The Fundamental Principles state that official statistics must be: practical and impartial; accurate, trustworthy and ethical; scientific and standardized; quality, timely, and with as low a response burden as possible; confidential and appropriately used; transparent and made public. They also state that NSOs have the responsibility to draw data from a variety of sources, to ensure that this data is assessed at the discretion of the office or agency, and to coordinate and cooperate with other organizations (nationally or internationally) to contribute to the improvement of statistical systems (United Nations Statistics Division, 2014)

### Information

Knowledge captured in any format, such as facts, events, things, processes, or ideas, that can be structured or unstructured, including concepts that within a certain context have particular meaning. Information includes **data** (Government of Canada, 2019; StatCan 2021b; UNECE CES, 2000).

## A) Enterprise Data and Information

Published and unpublished materials, regardless of medium or form, that are created or acquired because they enable and document decision-making in support of programs, services and ongoing operations, and support departmental reporting, performance and accountability requirements in an enterprise, organization, or government (Government of Canada, 2019; StatCan 2021b).

## Information life cycle

Encompasses the planning, collection, creation, receipt, capture, organization, use, re-use, dissemination, maintenance, protection and preservation, disposition, and evaluation of information (Government of Canada, 2019; StatCan 2019, 2021b). See also **data lifecycle management**.

## Interoperability

Interoperability is the ability to access and process data from multiple sources, then integrate that data for mapping, visualization, and other forms of representation and analysis. This allows systems and organizations to work together (inter-operate) towards mutually beneficial goals by sharing information and exchanging data. Syntactic and semantic interoperability are distinguished as separate types (European Commission, 2017; DDI Alliance, 2021; StatCan, 2020b). The European interoperability framework is an agreed upon approach to the delivery of European public services, which defines basic interoperability guidelines in the form of common principles, models and recommendations ([European Commission, 2017](#)).

**A) Semantic Interoperability** – Mutual understanding of the meaning of data and information in the communication process. The data is not only exchanged between two or more systems but also understood by each system.

**B) Syntactic Interoperability** – Allows two or more systems to communicate and exchange data and information, where the interface and programming languages are different.

Another approach distinguishes between four layers of interoperability:

**A) Legal interoperability** ensures the adequate legal basis for cooperation between organisations.

**B) Organisational interoperability**, business processes and their documentation are aligned with one another to ensure the efficient exchange of information. The I14Y interoperability platform plays an important role in this respect.

**C) Semantic interoperability** ensures that the content of information exchanged is preserved in its original form. To achieve this, all parties involved must know the meaning and the format of the data. “What is sent is what is understood”.

**D) Technical interoperability** describes all technical aspects of information exchange. Applications and infrastructures that link systems and services must be compatible. Aspects of technical interoperability include interface specifications, interconnection services, data integration services, data presentation and exchange, and secure communication protocols.

## Machine-actionable

Refers to data and information that is organized in a consistent way so that machines, or computers, can be programmed in accordance with the organized information. DDI provides machine-actionable metadata (DDI Alliance, 2021; StatCan, 2021b).

## Machine-readable

Refers to data and information that can be easily processed by a computer without human intervention while ensuring no semantic (dialectal) meaning is lost. Data in a data format that can be automatically

read and processed by a computer, such as CSV, JSON, XML, etc. Machine-readable data must be structured data (DDI Alliance, 2021; StatCan, 2021b).

### Master data

The consistent and uniform set of values, identifiers and extended attributes that describes the official, trusted, and shared data of the enterprise (DMBOK, 2017; Gartner, 2021).

### Master data management (MDM)

A technology-enabled discipline that ensures the uniformity, accuracy, stewardship, semantic consistency and accountability of the enterprise's official, trusted and shared master data assets. MDM allows uniformity in enterprise processes, operations, and analytics and reporting enabled by data consolidation, data governance, and data quality management (DMBOK, 2017; Gartner, 2021).

### Metadata

Metadata is the structural or descriptive information about primary data and processes that explains the measured attributes, their names, units, precision, accuracy, and data layout. Most importantly, metadata includes the data lineage that describes how the data was measured, acquired or computed. This happens under particular circumstances and for particular purposes. The set of circumstances and purposes (or perspective) for which some data are used as metadata is called the context. Metadata are data about data in some context (Gray et al., 2005; OECD, 2005, 2021; StatCan, 2021b; UNECE CES, 2000).

### Microdata

Information at the level of an individual object - statistical unit. Microdata is data on the characteristics of units of a population, such as an individual person, business, organization, or any other relevant unit of observation (e.g., transaction). They are files of records pertaining to individual respondent units. (OECD 2005; Statistics Canada, 2016).

### Necessity and Proportionality Principle

Necessity and Proportionality are concepts that have been incorporated by public administrations and national statistical systems to provide both a justification and a guide for designing strategies to gather sensitive data using surveys, administrative sources obtained from the public or private sector, or any other method. It holds that the potential benefits of the public interest pursued should be reasonably balanced against the interests of other stakeholders and data subjects.

In this principle, *Necessity* refers to the data that must be collected or produced for societal needs, and *proportionality* refers to the imperative that the data collection efforts be balanced with ethical considerations, drawing from what is known as the four-part test (composed of necessity, effectiveness, proportionality and alternatives). This is consistent with the use of this concept in the EU context, where *Necessity* is a fundamental principle when assessing the restriction of fundamental rights, *such as* the right to the protection of personal data.

*Proportionality* is a general principle of EU law, which restricts authorities in the exercise of their powers by requiring them to strike a balance between the means used and the intended aim. It is also consistent with the EU's *Data Governance Act's* focus on leveraging GDPR to balance the use of personal data with individuals' rights. The processing of data should be justified by clear and demonstrable public interest and should be necessary, relevant and proportionate in terms of detail. The cost and effort required for the supply and use of data should also be reasonable and proportionate to the public-interest benefits pursued, the interest of other stakeholders and data



subjects and the risks of harm if the data is not used (European Commission, 2020; European Data Protections Supervisor, 2020; Rancourt, 2021; StatCan 2020b; 2020c).

## **Ontology**

A formal specification of the vocabulary common to a domain or conceptualization, such as the objects, terms, definitions, and other entities that are assumed to exist in an area of interest and the relationships between them. Ontologies support the integration of data obtained from multiple sources by providing a common understanding with which to organize the documentation, collection, processing, presentation and analysis of data in a systematic manner (OECD, 2005; StatCan, 2021c; UNECE CES, 2000).

## **Open Data**

Open data refers to data that, through agreements and arrangements, is open and non-discriminatory in its access. Open data is machine readable and can be accessed and shared, free of charge, and used by anyone for any purpose, and subject to, at most, requirements that preserve integrity, provenance, attribution, and openness (OECD, 2021).

## **Privacy**

The status accorded to data that describes the degree of protection that will be provided. Privacy applies to data subjects, while confidentiality applies to data itself, and is related to a person's right to self-determination in the way their personal data are used. Privacy can include guiding principles such as accountability, consent, security, openness, necessity and proportionality (See also Necessity and Proportionality Principle) (OECD, 2005; StatCan, 2020b; 2021a).

## **Record Linkage**

The merging that brings together information from two or more sources of data with the object of consolidating facts concerning an individual or an event that are not available in any separate record; the bringing together of two or more micro-records to form a composite record (OECD, 2006; Statistics Canada, 2016)

## **Reference data**

Reference data typically consists of a small, discrete set of values that are not updated as part of business transactions but are used to impose a consistent classification. Reference data normally has a low update frequency. Reference data is relevant across more than one business system belonging to different organizations and sectors (European Commission, 2015; StatCan 2021b).

## **Taxonomy**

A classification, or structured lists of mutually exclusive categories, according to presumed natural relationships. Taxonomies or classifications support the integration of data obtained from multiple sources by organizing the documentation, collection, processing, presentation and analysis of data in a systematic manner (OECD, 2005; UNECE CES, 2000; StatCan, 2021c).

## **Unique Identifier**

An attribute or sequence of characters capable of uniquely identifying a statistical unit or administered item, or that with which it is associated, within a specified context or within a



registration authority (OECD, 2005; Statistics Canada, 2016).

### **Variable**

A characteristic, attribute, or property of a statistical unit being observed that may assume more than one of a set of values to which a numerical measure or a category from a classification can be assigned (e.g. income, age, weight, etc., and “occupation”, “industry”, “disease”, etc.) (OECD, 2006; Statistics Canada, 2016).

DRAFT

## Annex 2 Examples of data governance and data stewardship models in countries

1. This Annex provides examples of different models and frameworks used by countries:

- the role of the Chief Data Steward or Chief Data Officer
- where they are situated within the government structure
  - one data steward for the whole government
  - many data stewards (of which one can be NSO)
  - shared responsibility for data stewardship etc.

### Argentina

2. The Argentinian National Institute of Statistics and Censuses (Instituto Nacional de Estadística y Censos - INDEC) is a decentralized public body that operates within the Ministry of Economy, which exercises the direction of all official statistical activities carried out in the country.

INDEC is responsible for:

- implementing a statistical policy for the Argentine State
- giving structure to and leading the National Statistical System (NSS)
- designing methodologies for statistical production
- organizing and running statistical infrastructure operations
- producing basic indicators and social, economic, demographic and geographic data.

3. INDEC's role does not extend into the Data Steward space. There are several entities within the National Public Administration (NPA) that regulate public and private information, data protection, access to data and the way it is produced. There is no formally constituted central Data Steward role, with the function of coordinating the national data ecosystem.

4. The functions that make up the Data Steward role are split across the agencies that have primary responsibility for stewardship of government data. INDEC does work with these agencies in implementing the Open Data plan, but there is no single regulatory framework or agency that could be described as a national or federal Data Government model or framework.

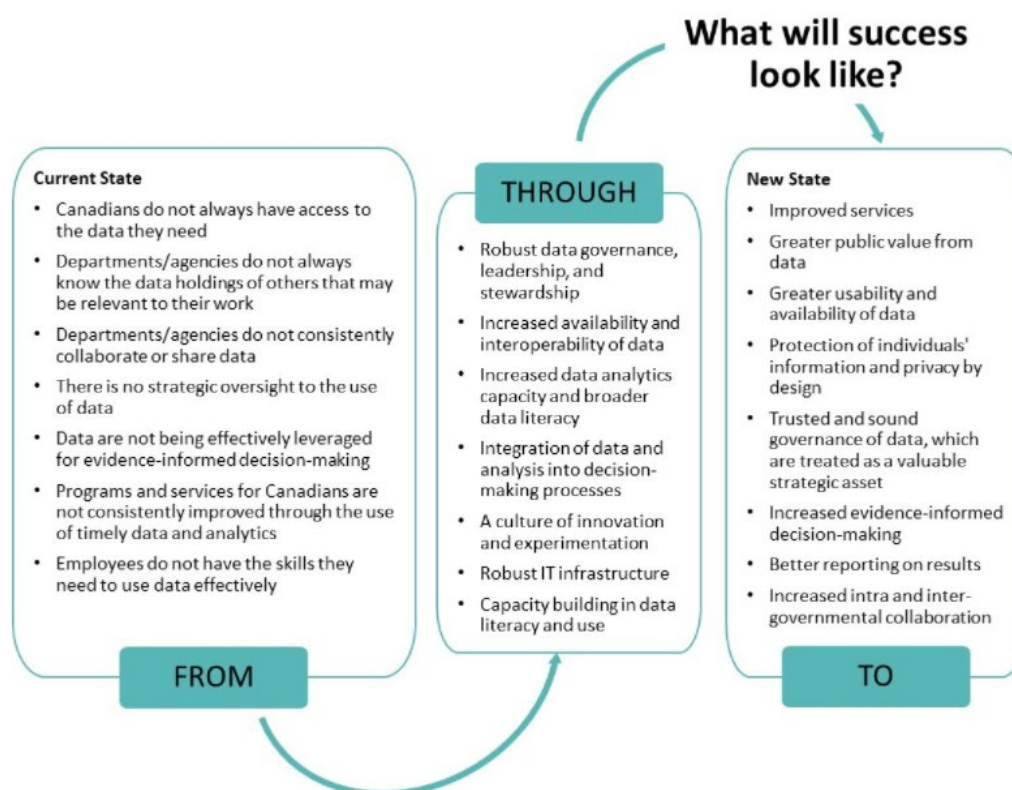
### Canada

5. The Government of Canada (GoC) has been prioritising innovation and increasing horizontality in order to facilitate a whole-of-government approach to the strategic use of data for public good. One such mechanism has been the creation and evergreening of the Data Strategy Roadmap for the Federal Public Service.

6. In 2018, the Data Strategy Roadmap for the Federal Public Service was published, a collaborative response by the Privy Council's Office (PCO), The Treasury Board (TBS) and Statistics Canada (StatCan) to a call from the Clerk of the Privy Council to develop a data strategy. While not a national data strategy concerned with all public and private data stores, this federal data strategy underpins the strategic use of data across the GoC, enables the transition to a digital government, and ensures that the entire public service can best leverage data and insights for evidence-based decision making and better outcomes for citizens. See the figure below for details about the current- and end-states that the Roadmap seeks to address and develop.

7. The Roadmap is an evergreen document currently being revised and updated to improve alignment between all levels of government, to address recent digital developments, and to account

for and describe the recent creation of a Chief Data Officer (CDO) of Canada role. The CDO will be responsible for providing GC-wide leadership for the management of data and information; developing the administrative frameworks and standards for governance and quality; enabling federal organization to leverage data; and oversight for information management and data. The evergreening of the Roadmap, the creation of the CDO and various intergovernmental work occurring around data governance and stewardship all work together to further develop Canada's digital, data, and information capacity.



## Finland

8. Finland's data governance model fits within the federated hybrid model. In the 'Government report on information policy and artificial intelligence'<sup>24</sup> (AI), submitted to Parliament in December 2018, the Government of Finland looked at information policy not only from the viewpoint of information management, but also from the perspectives of the conditions placed on the use of information, the value of the information involved, ethical principles and financial impacts.

The report:

- sets out the new information policy
- proposes extensive use of AI
- identifies the levels of AI competence required
- explores possibilities of economies in platforms and data.

9. The measures included in the report cover the collection and merging of information, information disclosure and storage, as well as information security and data protection. The measures examined in the report also include ethical issues, how to secure expertise, regulatory issues and policy-level participation in the EU and international forums.

<sup>24</sup> [Report on ethical information policy in an age of artificial intelligence](#) (5 December 2018, PDF 1.7MB)

10. The report constitutes the knowledge base and policy framework on which a roadmap with prioritised actions can be built in the future. Information policy is connected to all other policy areas because any issue will be investigated and decisions made on the basis of comprehensive information. This new policy area lays the groundwork for the development of targets for Information Management in the public and private sectors.

11. Following on from this policy, a new 'Public Administration Information Management Act' (906/2019) covering data governance, data security and data management came into effect in 2020. This Act contains provisions on the organisation and description of data management, interoperability of databases and information systems, implementation of technical interfaces and implementation of data security for the public administration sector. This Act established the Information Management Board under the auspices of the Ministry of Finance in 2020.

12. The function of the Information Management Board is to promote the implementation of information management and data security procedures laid down in the Public Administration Information Management Act and to ensure that the requirements of the Act are met. The Information Management Board is not a general authority for information management; its tasks are limited to the scope specified in the Public Administration Information Management Act.

13. The Information Management Board has prepared various general guidelines and recommendations to support data governance and data management in public organisations. Guidelines and recommendations have been prepared in the following areas:

- the handling of classified documents (in English<sup>25</sup>)
- data security and risk management
- the metadata of case processing
- technical interfaces and viewing connections
- the documentation of the data governance model used by the organisation
- the responsibilities of directors in data governance and data management

14. Statistics Finland has a place on the board representing the National Statistical System. The Statistics Finland representative provides expertise on data governance and data management in statistical work. In the statistical community there are already a number of good examples of how to improve metadata, process descriptions, total quality, common classifications, standards etc. that may be useful to share with other public organisations.

15. On the Board, there are also representatives from ministries, government institutions and municipalities. Their expertise covers areas such as basic registers, data security, ICT expertise on interoperability of databases and case management.

16. All public organisations are expected to follow these recommendations, but it is important to understand that the recommendations themselves are not obligatory. Instead, they are best practice examples of how to implement the requirements in the Act, and every organisation must implement these recommendations in a way that best suits their own operating environment and practices.

17. The Board also organises seminars and workshops on the implementation of their recommendations. The participation rate so far has been very good (around 300-500 participants per workshop). The purpose of these seminars and workshops is to share knowledge on best practice, give examples how different organisations have applied the recommendations and to support participants in understanding what is expected and what the minimum requirements for good information management are.

---

<sup>25</sup> <https://julkaisut.valtioneuvosto.fi/handle/10024/162846>

18. The other responsibility of the Board is to monitor the implementation of the Act in public organisations. In the first year, the Board collected information on the descriptions of the responsibilities of directors in Information Management and how data security and data management training is conducted in the organisations. In the following year, the surveillance focussed on documentation of the data governance model and how public organisations are implementing it.

## Mexico

19. The statistical and geographic information production system in Mexico is regulated in the Constitution and by legislation. The National Institute of Statistics and Geography (INEGI) is an autonomous body, completely independent from the federal government.

20. As specified in legislation, INEGI plays a dual role in the national information system:

- as a direct producer of official information and
- as the coordinator that sets norms and standards for the production of official statistics by different government agencies.

21. The Mexican system can be classified within the Federal Hybrid model: it has direct control of the production cycle of census and survey information for the production of statistical information, as well as of the statistics derived from the use of administrative records and other sources. INEGI also issues norms and standards to be followed by government agencies that generate information considered to be of national interest.

22. The Institute also has the authority to establish agreements with companies and non-governmental organizations to obtain data that can be reused in the generation of statistical information.

23. INEGI is the institution that, by law, assumes the role of custodian of all data and information used to produce statistics and establishes the responsibility for their careful handling, adhering to the principles of quality, relevance, truthfulness, opportunity, confidentiality and independence.

24. Each institution that provides information or generates statistics of national interest has a coordinator who is responsible for implementing the policies and standards determined by INEGI.

25. As can be seen, the role of Chief Data Steward does not formally exist in Mexico; however, it is INEGI that carries out the activities that this role implies within the country.

26. Those responsible for coordinating information within the different government agencies have the role of linking the data, but not of custody or direct administration of the data. By this definition they could not be considered to be true Data Stewards.

27. For this reason INEGI is currently reviewing its standards and policies to strengthen the data ecosystem through the figure of Data Stewards across the entire Mexican system.

28. INEGI has adapted the Generic Statistical Business Process Model (GSBPM) to improve information management throughout the entire cycle, in addition it has adapted its norms and standards to strengthen data management and the generation of information products and corresponding metadata.

29. In addition INEGI has developed a new information governance and architecture strategy that facilitates standardization, transparency, confidentiality and the quality of information. It has also

implemented systems with improved security schemes to better manage information and preserve its integrity.

30. One of the main challenges anticipated in the near future will be to incorporate information from companies and organizations to regularly produce statistical information. To get access this large pool of information it will be necessary to improve the regulatory framework to provide these companies and organizations assurances that their information can be processed by INEGI without affecting any other legal obligations they may have.

31. A clearly defined data governance framework and agreement on the role and responsibilities of Data Stewards will be key to moving the Mexican information system to a new level, involving more intensive use of administrative records and alternative sources of information.

## New Zealand

32. In New Zealand all government agencies are autonomous, with Chief Executives appointed by a Public Service Commissioner. While being constitutionally separate from the Executive Branch of Government they support decision-making of individual ministers under the Westminster-style system of government.

33. This is a very decentralised model of government administration, which has a number of benefits but also some limitations in terms of systemic approaches to data management. In recognition of this in 2017, the New Zealand Public Services Commissioner created the role of Government Chief Data Steward (GCDS), recognising that strong system focussed leadership was required to help New Zealand realise data as an asset. The Public Services Commissioner appointed the Chief Executive of Stats NZ to that role. In this leadership role, the GCDS recognises and champions the benefits of agencies coming together and leveraging their data assets to deliver value for New Zealanders.

34. In conjunction with the role of the GCDS, Stats NZ serves as the lead agency for data within the New Zealand government data system, by facilitating and enabling a joined-up approach to data-related opportunities and challenges. In this capacity, Stats NZ, along with the GCDS, aims to support agencies to maximise the potential of their data and ensure it is used effectively, while maintaining the trust and confidence of New Zealanders.

35. To do this, the GCDS develops data policy and principals, and has published a Government Data Strategy and Roadmap to provide clear guidance on how the data environment should operate, while ensuring a data-driven future for New Zealand. This is accomplished by:

- driving the response to new and emerging data-issues
- guiding best practice and safe innovation
- setting common data standards to ensure the consistent collection of data
- allowing for integration, comparability and production of meaningful insights and partnering with agencies and Māori to build capability
- helping agencies build the skills, processes, tools and services for the successful collection, management, use and dissemination of data.

36. New Zealand's central government data governance is currently closer to the distributed model, as individual government departments are operationally autonomous. There are two key governance bodies supporting the role of the GCDS, the Digital Government Leadership Group and the Information Group.

### The Information Group

37. The Information Group is the key governance body for the Government Data Strategy and Roadmap - responsible for setting, implementing, monitoring, and reviewing the progress made against the Strategy. In its governance of system-level participation by government agencies, the Information Group supports the GCDS, through delivery of things like implementation planning, progress reporting, identification of risks and facilitating engagements and communications. The group is chaired by the Deputy CE, Data System Leadership, at Stats NZ.

### The Digital Government Leadership Group

38. The Digital Government Leadership Group (DGLG) is co-chaired by the Government Chief Data Steward and the Government Chief Digital Officer. The role of the DGLG is to support the Government Chief Digital Officer and the GCDS to develop and improve the digital and data system across the public service and ensure the public service is aligned with the *Strategy for a Digital Public Service* and the *Government Data Strategy and Roadmap*, and other relevant strategies.

39. Treating data as a valuable asset does not mean simply gathering more data or applying it indiscriminately. It means properly stewarding and managing, and being intentional about, how we use and generate value from the data we hold.

40. Through its government data leadership role, Stats NZ administers the ongoing development of the *Data Investment Plan* which guides and informs strategic system investment in data, including the way that critical data is managed within baseline, and how we consider new investment. The *Data Investment Plan* is a cross-agency initiative and is being developed in collaboration with government agencies through a phased approach.

41. The *Data Investment Plan* has been endorsed by New Zealand's Cabinet, and the first iteration of the Plan was published in December 2021. An update to the plan to incorporate Māori data investment needs and an investment blueprint is expected in the third quarter of 2022. The Data Investment Plan is governed by the Information Group and will be monitored through a planned annual *Health of the Data System* report.

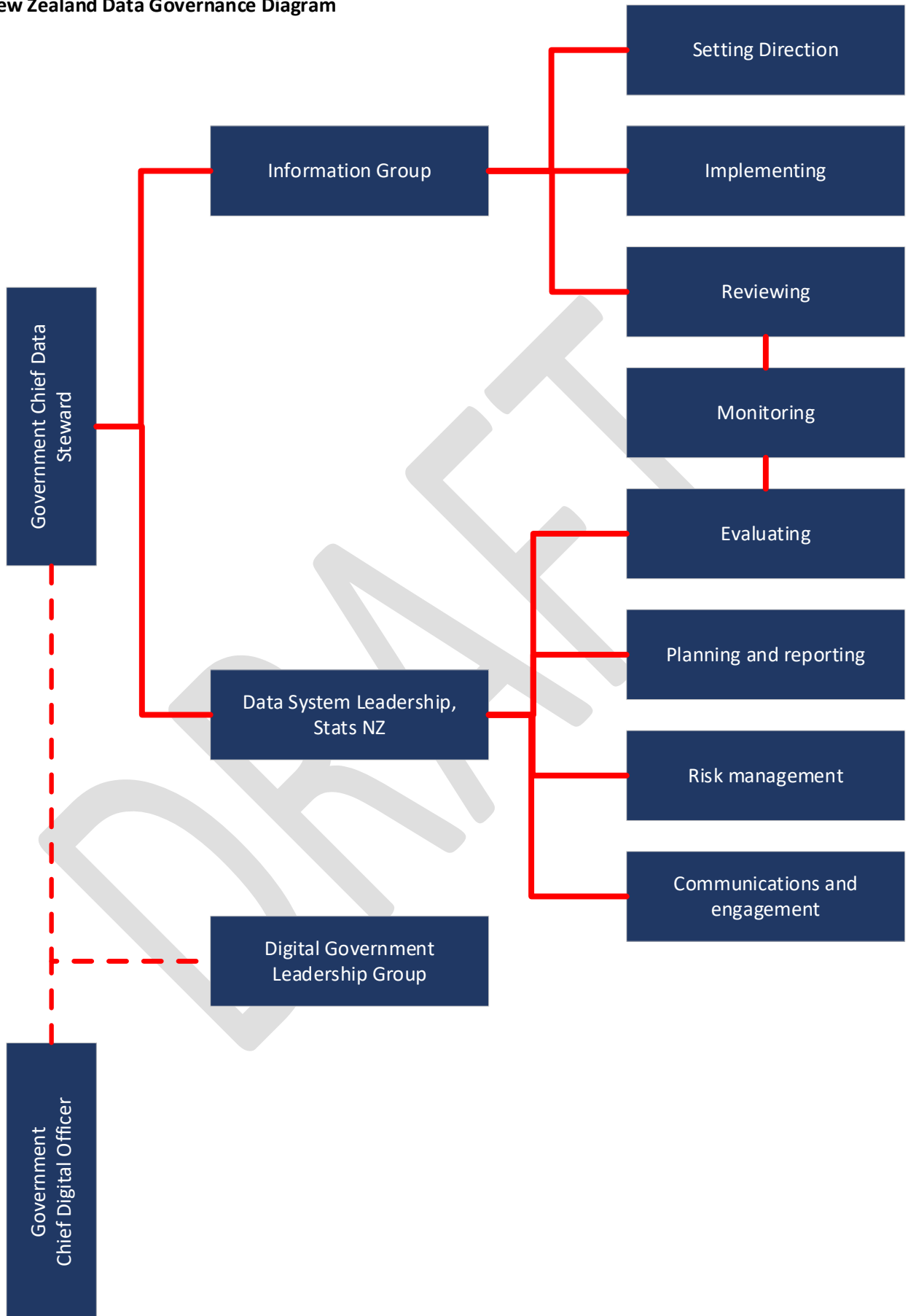
### Māori data governance

42. The Treaty of Waitangi (New Zealand's founding document) promised that the Crown would uphold the authority that indigenous Māori tribes had over their lands and taonga (prized and valued objects, resources, ideas and techniques). Data about Māori has been identified as a taonga under the terms of the treaty, and as such the Crown has an obligation to work with Māori treaty partners when considering the governance of data about Māori.

43. Governance of the New Zealand data system was not designed in partnership with Māori. Nor do we have a te ao Māori (Māori world view) lens across the wider official data system that supports both Māori and government aspirations for data. This has resulted in challenges including a lack of trust and confidence by Māori in the official data system, inadequate meaningful Māori participation in that system, including at governance levels, and missed opportunities for Māori to add value to the official data system through te ao Māori insights and innovations.



## New Zealand Data Governance Diagram



44. The Government Chief Data Steward has recognised the opportunity for government to work in collaboration with Māori partners, co-designing a Māori Data Governance model for the official government data system. This would employ a Treaty of Waitangi-based co-design process that appropriately reflects the obligations of the Crown and Māori partnership.

45. The Māori Data Governance design work sits under the auspices of a formal relationship agreement, Mana Ōrite, forged between the Data Iwi Leaders Group (DILG) representing Māori interests, and Stats NZ. The work to date has produced two reports:

- 1) Tawhiti Nuku, Māori Data Governance Co-design Report on the outcomes of the co-design process and its recommended next steps for progressing
- 2) Māori Data Governance Co-design Review, which focuses on the process of co-design.

46. Work is progressing through an Ohu raraunga (data working group) with government and Te Ao Māori membership, to continue the momentum on refining, testing and implementing the Māori Data Governance model through 2022.

### New South Wales (Australia)

47. New South Wales (NSW) is one of the 6 states and 10 federal territories that make up the Federation of Australia. Each state is a self-governing political entity with incomplete sovereignty (having ceded some sovereign rights to federation) and has its own constitution, legislatures, departments, and certain civil authorities (e.g. judiciary and law enforcement) that administer and deliver most public policies and programmes.

48. Data governance policy in NSW is currently under review which will include new definitions of data roles and responsibilities. This summary describes the current state which is one of transition. Data Custodianship Policy in NSW is administered by the Department of Customer Service. This policy took the approach of describing functions rather than labelling roles due to the variation in the use of labels for these roles in different state agencies, and as a result uses the term 'custodian' and doesn't use the term 'steward'. Custodianship is defined as formally assigning rights and responsibilities for data and information assets, including capture and management on behalf of the NSW Government. Despite this, from the documentation available it would appear that their model of data governance best fits the distributed model, due to the absence of a role equivalent to a central data steward.

49. In NSW the custodianship role and its associated responsibilities belong to the government agency which acts on behalf of the State of NSW i.e. each government agency acts as the custodian of the data and information assets and products held in their care. The Department of Customer Service has a co-ordinating role, but can only provide guidance to the autonomous agencies. The current Data Reform program has established a NSW Data Leadership Group (NDLG). This group consists of Chief Data Officers from each government Cluster. Cluster CDOs have a functional role (i.e. they perform this role in addition to their substantive role). At present, NSW does not have a NSW Government Chief Data Officer, but are working toward a hybrid governance model through the NDLG.

50. In their guidance role the DCS published a Data Governance Toolkit built on a common understanding of the benefits, obligations and best practice. The aim of this was to ensure a consistent approach to data governance across NSW Government agencies and provide practical and consistent guidance on the key components of an effective data governance program, as well as to create a shared understanding of what good data governance looks like. Compliance with the Toolkit is not mandatory, but following the guidance in the Toolkit is intended to:

- support agencies to maximise the value of data while reducing data-related risk;
- assist agencies in meeting their legislative and regulatory obligations;

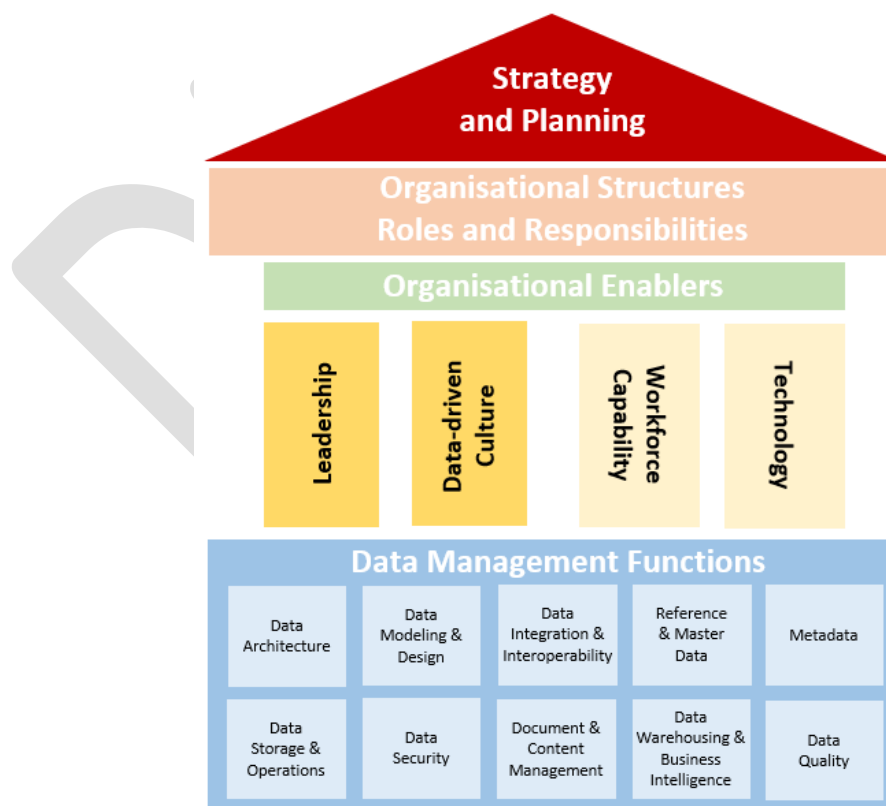
- ensure data is managed in line with national and international standards;
- facilitate better interoperability between agencies; and
- build data governance maturity at both the departmental and all-of-government levels.

### The Data Governance Model

51. The Model defines four interconnected tiers of data governance activities, each of which is critical to effective data governance in agencies. The four tiers are:

- 1) **Strategy and planning** – agencies clearly define the data governance program’s values, vision and mission and compose a business-aligned strategy for governing and managing data as an organisational asset.
- 2) **Organisational structures & roles & responsibilities** – agencies ensure accountability and decision-making authority for data-related activities to be appropriately assigned and formalised at all levels of the organisation.
- 3) **Organisational Enablers** – agencies ensure the organisational environment is an enabler of good data governance. This means ensuring there is a strong motivation (or ‘will’) to achieve good data governance by having sustained buy-in and investment from senior leadership, as well as fostering a strong organisational data culture. It also means ensuring the organisation has the requisite capability (or ‘skill’) to achieve good data governance, both in terms of workforce capabilities, as well as appropriate tools and technologies.
- 4) **Data Management** – agencies ensure their data governance program has oversight of core data management functions (e.g. data quality, storage, security, business insights etc.).

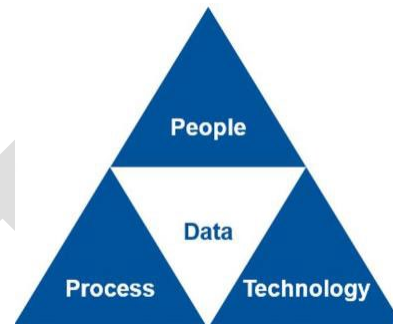
Figure 1: Data Governance Model



## Interpreting the Model

52. Each component of the Model includes a high-level summary of what the component is, why it is important, what good practice looks like (i.e. the goals), how to achieve good practice and, where appropriate, provides references to useful resources and relevant standards. The level of detail for each component has been kept to a high level and the practical elements of the framework will gradually be expanded with input from agencies once the model is in use.

The Model also aligns with Gartner's (2017) 'Golden Triangle' of 'People, Process and Technology' (with Data at the centre), which recognises that effective data governance is an ongoing effort executed by people, enabled by repeatable processes, and supported by technology. Each component of the Model therefore encompasses a mix of accountabilities relating to people, processes and technologies to support the implementation of that component.



Source: Gartner 2017

<https://data.nsw.gov.au/data-governance-toolkit-0>

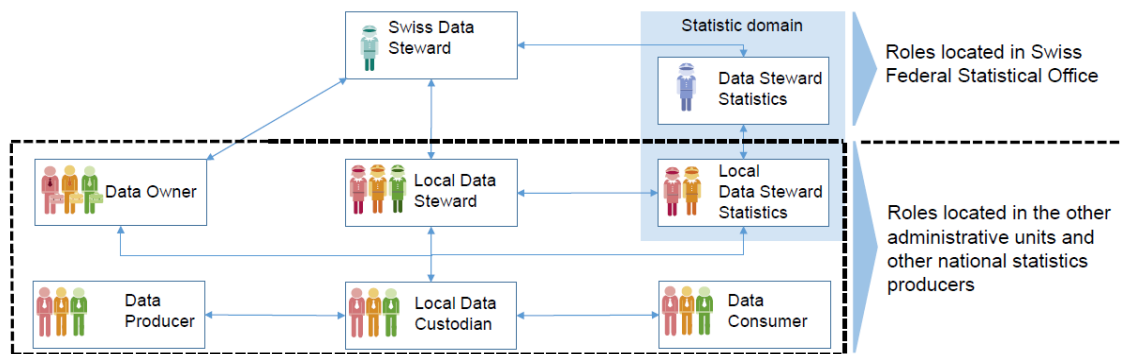
## Switzerland

53. The Swiss Confederation (the official English name of the country) is a federal republic composed of 26 cantons.

54. From the available documentation the Swiss model of data governance seems to fit the Federated Hybrid model. We have been unable to find a document that states that it is the data governance model or framework, but documentation that describes the role of Data Stewards in the Swiss government system best fits this model.

55. The Swiss government is working towards establishing the once-only principle (a concept that aims to ensure that citizens, institutions and companies only have to provide certain standard information to the authorities and administration once) while keeping data storage and ownership decentralised. This requires interoperability to achieve sharing of data across administrative units to allow the reuse of data. The first step in achieving this is the development of a nationwide metadata catalogue containing standardised descriptions of the data as well as information about provenance, usage, legal restrictions and quality of shared data.

56. The Swiss data governance model requires data stewards at different levels. The following diagram describes the roles they play.



57. There are two data Steward roles at the federal level:

- The Swiss Data Steward is centrally located in the Swiss National Statistical Institute (NSI) also known as the Swiss Federal Statistical Office (FSO) and is responsible for:
  - Coordinating the standardization and harmonization process.
  - Identifying and describing the data requirements of the various users.
  - Managing the content of metadata (data catalogue).
  - Validating the quality assurance of metadata and data in the administrative areas using data analyses.
- The Data Steward Statistics is also centrally located in the Swiss NSI and is responsible for:
  - Cooperation and coordination with international bodies
  - Enabling the use of common international classification standards.
  - Further developing guidelines for implementing of the Once-Only Principle in the statistical field.
  - Validating the quality assurance of the statistics producers.
  - Defining standards and guidelines for naming and metadata

58. At the local (Canton) level the split is maintained between the Local Data Steward (public administration offices) and the Local Data Steward Statistics (other national statistics producers), and their roles are:

- Local Data Steward (public administration offices)
  - Standardization and harmonization of data within area of responsibility
  - Validation of the quality assurance of metadata and data in the administrative areas by means of data analyses.
  - Further development of specifications to ensure interoperability
- Local Data Steward Statistics (other national statistics producers)
 

In addition to the responsibilities listed for the Local Data Steward:

  - Standardization and harmonization of statistical data within the steward's area of responsibility
  - Identify and describe the data needs of the various users in the statistical field

59. Implementation of this program

- The National Data Management (NaDB) program was launched in October 2019 by the Federal Statistical Office. This is a work in progress.
- In order to promote the re-use of data in the long term, a metadata catalogue has been established containing descriptions and information about shared data. An interoperability platform has been developed to provide a system that is available to all participating offices.
- These components anticipate the implementation of several projects which will be carried out within specified areas (wages, occupational profiles, taxes, health provisions). Drawing on previous experiences, each new component will be assessed to ensure that it is compatible with the long-term implementation of the once-only principle and will be

adapted if required. Data protection and privacy considerations as well as necessary amendments to legal frameworks will be part of the program.

## European Union: The new data governance legislation of European Union

60. The European Union (EU) has undertaken a number of actions to improve data sharing and data governance in the EU Member States.

61. The following paragraphs describe the 2020 data strategy and the proposal for a new data governance act. This combination of strategy and legislation aims to put in place a federal hybrid model where the European Union has a legislated centralised structure and framework for data governance, and the EU Member States make their own decisions on how to implement these regulations. Under this structure public organisations have the autonomy to decide how to manage their own datasets within the regulations.

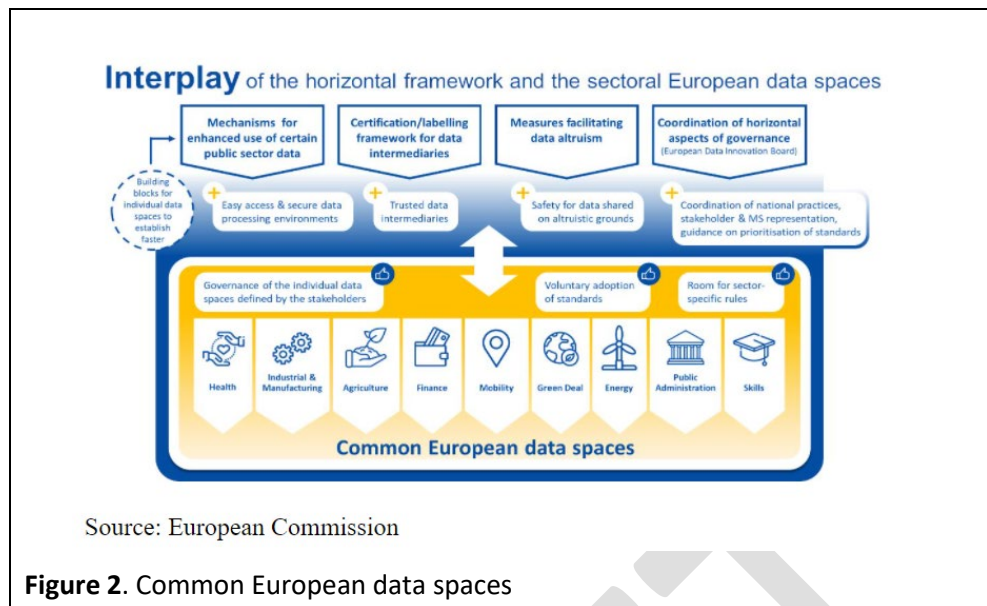
62. The 'European Strategy for Data'<sup>26</sup> was published in February 2020. The objective of this strategy is to make sure that the EU becomes a leader and role model for a society empowered by data.

63. The EU will create common European data spaces to ensure that more data become available for use in the EU. The aim of this is to create a single market for data, to unlock unused data and support the flow of data freely both within the European Union and also across sectors for the benefit of businesses, researchers and public administrations. Figure 1 describes the aims of the strategy. Figure 2 shows how the common European data spaces will work.



<sup>26</sup> <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0066&from=EN>





64. To fulfil the aims of the strategy, the EU is building a legal framework that will cover data protection, fundamental rights, safety and cybersecurity. In November 2020, the European Commission proposed the *Regulation on European data governance*<sup>27</sup> (DGA) which is the first of a set of measures announced in the 2020 European Strategy for Data.

65. This draft Regulation will empower users to stay in control of their data and encourage the creation of common European data spaces in key sectors. These sectors include health, the environment, energy, agriculture, transport, finance, manufacturing, public administration, and skills (see Figure 2).

66. To facilitate data sharing across the EU and between sectors, the draft Regulation aims to strengthen mechanisms that increase data availability and foster trust in intermediaries. The Regulation:

- i. introduces conditions under which public sector bodies may allow the re-use of certain data they hold, notably data which are protected on the grounds of commercial confidentiality, statistical confidentiality, protection of intellectual property rights of third parties or the protection of personal data.
- ii. imposes obligations on providers facilitating the sharing of personal and non-personal data.
- iii. establishes a 'register of recognised data altruism organisations' in order to increase trust in the operations of registered organisations that facilitate the voluntary sharing of data for the common good.
- iv. creates a formal expert group, namely the 'European Data Innovation Board'.

67. The Commission proposal for the Regulation is currently being considered in trilogue<sup>28</sup> negotiations between the Council of the European Union, European Parliament and the European Commission. After the Regulation is approved, there will be an 18-month transition period.

<sup>27</sup> Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on European data governance (Data Governance Act) COM/2020/767 final:  
<https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020PC0767&from=EN>

<sup>28</sup> Trilogue definition: informal tripartite meetings on legislative proposals between representatives of the Parliament, the Council and the Commission

68. This Regulation will also change the practices of National Statistical Institutes (NSI)<sup>29</sup> as well as other producers of statistics. At this time, it is too early to say what impact this will have for EU NSIs.

## OECD work on data governance in the public sector

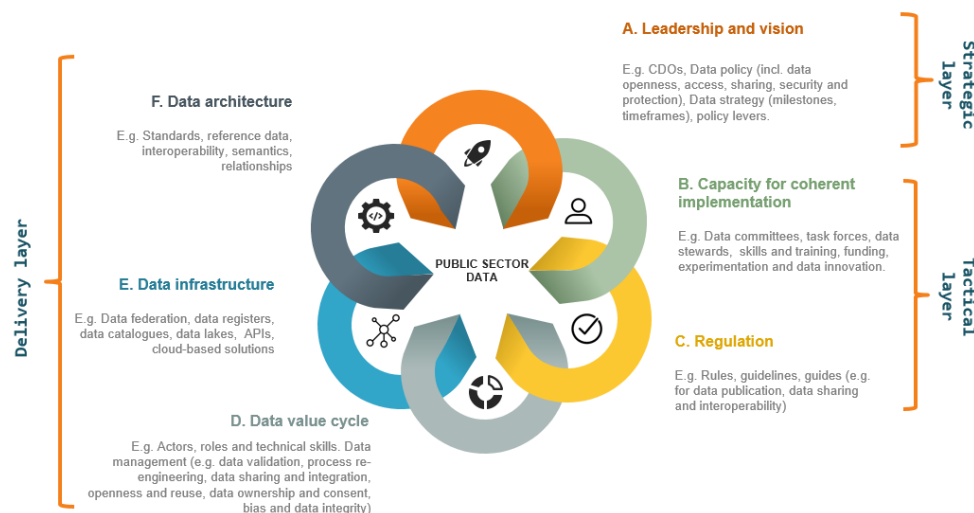
69. The OECD has developed a model for data governance in the public sector as means to showcase the core elements countries can take into consideration when designing and deploying data projects and initiatives. As presented in the 2019 OECD Report *The Path to becoming a data-driven public sector*<sup>30</sup>, the framework aims to bring greater clarity and structure to the definition and implementation of the concept of data governance at the national level across OECD member and partner countries.

70. The model is based on the extensive OECD work on digital government and government data and additional research carried-out by the OECD Secretariat. Earlier versions of the model can be found in previous OECD digital government reviews, namely the 2017 OECD Digital Government Review of Norway, the 2019 OECD Digital Government Review of Sweden, the 2019 OECD Digital Government Review of Peru, and the 2019 OECD Digital Government Review of Argentina.

71. As described in the 2019 Report, we cite, the model intends to highlight the equal and strategic relevance and the value of all organisational, policy and technical aspects for the success of data governance. It identifies a range of non-exclusive data governance elements and tools, and organises them in six different groups (a – f). These six groups are then arranged under three core layers of data governance (Strategic, Tactical and Delivery) using the three traditional data governance categories as guidance (Strategic, Tactical, Operational) as discussed and/or presented in Ghavami, 2015; DAMA, 2017; and the BARC's 9-Feld-Matrix [see Grosser (2013) and BARC (2019)]. The model is also based on additional research including Ladley (2012) and Sen (2019):

- **Strategic layer [including (a) Leadership and Vision]:** Some of the data governance elements in this layer include national data strategies, and leadership roles. It is worth noting that the model considers data strategies as an element of good data governance.

The OECD Model for data governance in the public sector



<sup>29</sup> National Statistics Institute (NSI) is a term used by the EU that is equivalent to the term National Statistic Office used by the UN.

<sup>30</sup> For more information see: OECD (2019), *The Path to Becoming a Data-Driven Public Sector*, OECD Digital Government Studies, OECD Publishing, Paris, <https://doi.org/10.1787/059814a7-en>.

This argument lays on the fact that data strategies enable accountability in relation to responsibilities and can help define leadership, expectations, roles and goals. The strategic layer also highlights how the formulation of data policies and/or strategies can benefit from open and participatory processes, thus integrating the inputs of actors from within and outside the public sector towards greater policy ownership.

- **Tactical layer [including (b) Capacities for Coherent Implementation and (c) Legal and regulatory frameworks].** It enables the coherent implementation and steering of data-driven policies, strategies and/or initiatives. It draws upon the value of public sector skills and competences, job profiles, communication, coordination, collaboration as instruments to improve the capacity of the public sector to extract value from data assets. It also highlights the value of formal and informal institutional networks and communities of practice as levers of public sector maturity and collective knowledge. This layer also comprises data-related legislation and regulations as instruments that help countries define, drive and ensure compliance with, the rules and policies guiding data management, including data openness, protection and sharing.
- **Delivery layer [including (d) the integration of the data value cycle, (e) data infrastructure, and (f) data architecture].** The delivery layer allows for the day-to-day implementation (or deployment) of organisational, sectoral, national or cross-border data strategies. It touches on different technical and policy aspects of the data value cycle across its different stages (from data production, openness and re-use), the role and interaction of different actors in each stage (e.g. as data providers), and the inter-connection of data flows across stages. In this light, each stage is inter-connected but has specific policy implications in relation to the expected outcomes. For instance, data sharing initiatives (e.g. the production of good quality, standardised and inter-operable government data) can contribute to data re-use by external actors in latter stages (e.g. as open government data). The adoption of technological solutions (e.g. cloud-based data hosting services, APIs, data lakes) takes place in this layer for it supports of those policy goals defined in the strategic layer. It also relates for instance to the need for reengineering legacy data management practices and processes or retrofitting and adapting legacy data infrastructures. Data interoperability and standardisation also take place at this level.

72. The OECD underlines that the elements used to exemplify the plethora of policy instruments, arrangements, initiatives and/or tools that can be used by countries to deploy their data governance frameworks is not exhaustive. Thus, OECD countries might opt for adopting different data governance elements and tools that better fit into their national context and public sector culture in line with the proposed three layers and the six underlying categories presented in the model.

73. More broadly, the OECD has been working on developing common principles for data governance. In 2021, the OECD Council adopted the Recommendation on Enhancing the Access to and Sharing of Data<sup>31</sup> as a means to define a set of guidance principles for data governance, including across sectors.

74. Also in 2021, the OECD launched the Good Practice Principles for Data Ethics in the Public Sector<sup>32</sup> as an action-oriented tool to support countries in the implementation of this emerging area of work.

<sup>31</sup> For more information see: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0463>

<sup>32</sup> For more information see: <https://www.oecd.org/digital/digital-government/good-practice-principles-for-data-ethics-in-the-public-sector.htm>

## References

- Canadian Government. (2019). *Policy on Service and Digital*. Retrieved from <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32603>
- Canadian Government. (2021). *Canada's Digital Charter: Trust in a digital world*. Retrieved from [https://ic.gc.ca/eic/site/062.nsf/eng/h\\_00108.html#:~:text=Digital%20Charter%20Implementation%20Act%2C%202020%20In%20November%202020%2C,protection%20of%20personal%20information%20in%20the%20private%20sector.](https://ic.gc.ca/eic/site/062.nsf/eng/h_00108.html#:~:text=Digital%20Charter%20Implementation%20Act%2C%202020%20In%20November%202020%2C,protection%20of%20personal%20information%20in%20the%20private%20sector.)
- Data Governance Institute. (n.d.). *Governance and Decision Making*. Data Governance Institute. Retrieved from: <https://datagovernance.com/governance-and-decision-making/>
- Data Documentation Initiative Alliance. (2021). *DDI Alliance Glossary*. DDI Alliance. Retrieved from: <https://ddialliance.org/resources/ddi-glossary>
- Data Management Association (DAMA). (2017). *DAMA-DMBOK: Data Management Body of Knowledge (2<sup>nd</sup> Ed.)*. Basking Ridge, NJ: Technics Publications.
- Economic Commission for Europe of the United Nations (UNECE). (2000). *Terminology on Statistical Metadata In Conference of European Statisticians Statistical Standards and Studies*. (53), Geneva.
- European Commission. (2015). *Methodology and tools for metadata governance and management for EU Institutions [PDF]*. ISA Programme. Retrieved from: [https://joinup.ec.europa.eu/sites/default/files/custom-page/attachment/methodology\\_and\\_tools\\_for\\_metadata\\_governance\\_and\\_management\\_for\\_eu\\_institutions.pdf](https://joinup.ec.europa.eu/sites/default/files/custom-page/attachment/methodology_and_tools_for_metadata_governance_and_management_for_eu_institutions.pdf)
- European Commission. (2017). *European Political Strategy Centre, Enter the data economy : EU policies for a thriving data ecosystem*. Publications Office 21:11. Retrieved from: <https://data.europa.eu/doi/10.2872/33746>
- European Commission. (2017). *European Interoperability Framework*. Luxembourg: Publications Office of the European Union. Retrieved from: [https://ec.europa.eu/isa2/sites/default/files/eif\\_brochure\\_final.pdf](https://ec.europa.eu/isa2/sites/default/files/eif_brochure_final.pdf)
- European Commission. (2020). *Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on European data governance (Data Governance Act)*. Retrieved from: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020PC0767&from=EN>
- European Commission. (2021). *Towards a European strategy on business-to-government data sharing for the public interest, Final report prepared by the High-Level Expert Group on Business-to-Government Data Sharing*. Luxembourg: Publications Office of the European Union. Retrieved from: <https://op.europa.eu/en/publication-detail/-/publication/d96edc29-70fd-11eb-9ac9-01aa75ed71a1>
- European Data Protection Supervisor. (2020). *Necessity and Proportionality*. European Data Protection Supervisor. Retrieved from: [https://edps.europa.eu/data-protection/our-work/subjects/necessity-proportionality\\_en](https://edps.europa.eu/data-protection/our-work/subjects/necessity-proportionality_en)

- Fellegi, I. P. (1995, October). Characteristics of an Effective Statistical System [Paper presentation]. Morris Hansen Lecture, Washington Statistical Society 1995, Washington, DC, United States.
- Floridi, L. and Taddeo, M. (2016). What is data ethics?. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2083), p.20160360.
- Gray, J., Liu, D. T., Nieto-Santisteban, M., Szalay, A., DeWitt, D. J. & Heber, G. (2005). Scientific data management in the coming decade. *SIGMOD* 34:4 (34–41).
- Gartner. (2021). *Information technology: Gartner Glossary*. Gartner. Retrieved from: <https://www.gartner.com/en/information-technology/glossary/master-data-management-mdm>
- IBM Cloud Education. (2020). *IBM Cloud Learn Hub*. IBM Cloud Education. Retrieved from: <https://www.ibm.com/cloud/learn/data-catalog>
- Information Commissioner's Office, UK. (2019). *Data sharing: a code of practice*. Guide to Data Protection. Retrieved from: <https://ico.org.uk/for-organisations/guide-to-data-protection/>
- Ladley, J. (2020). *Data Governance: How to Design, Deploy, and Sustain and Effective data Governance Program* (2<sup>nd</sup> Ed.). London, UK: Academic Press.
- Miller, R. J. (2014). Big Data Curation *In* 20th International Conference on Management of Data (COMAD). Hyderabad, India.
- New Zealand Government. (2020). Data stewardship: managing New Zealand's data better to change lives. A Data Stewardship Framework for NZ. Retrieved from: <https://www.data.govt.nz/assets/Data-stewardship-framework-and-toolkit-Nov2020.pdf>
- Organisation for Economic Co-operation and Development. (2001). *Glossary of Statistical Terms*. OECD. Retrieved from: <https://stats.oecd.org/glossary/index.htm>
- Organisation for Economic Co-operation and Development. (2002). *Measuring the Non-Observed Economy: A Handbook*. Paris: France: OECD Publications.
- Organisation for Economic Co-operation and Development. (2011). *Quality Framework for OECD Statistical Activities* [PDF]. Retrieved from: <https://www.oecd.org/sdd/qualityframeworkforoecdstatisticalactivities.htm>
- Organisation for Economic Co-operation and Development. (2017). The role of national statistical systems in the data revolution, *In* Development Co-operation Report 2017: Data for Development, Paris, France: OECD Publishing. Retrieved from <https://doi.org/10.1787/dcr-2017-8-en>.
- Organisation for Economic Co-operation and Development. (2018). Governing open data for sustainable results, in *Open Government Data Report: Enhancing Policy Maturity for Sustainable Impact*, OECD Publishing, Paris. Retrieved from: [https://read.oecd-ilibrary.org/governance/open-government-data-report/governing-open-data-for-sustainable-results\\_9789264305847-4-en#page1](https://read.oecd-ilibrary.org/governance/open-government-data-report/governing-open-data-for-sustainable-results_9789264305847-4-en#page1)

Organisation for Economic Co-operation and Development. (2019). Data Governance in the Public Sector

*In The Path to Becoming a Data-Driven Public Sector*, OECD Digital Government Studies, OECD Publishing, Paris, <https://doi.org/10.1787/059814a7-en>.

Organisation for Economic Co-operation and Development. (2021). Recommendation of the Council on

Enhancing Access to and Sharing of Data. OECD Legal Instruments. Retrieved from: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0463>

Plotkin, D. (2021). Data Stewardship: An Actionable Guide to Effective Data Management and Data Governance (2<sup>nd</sup> Ed.). London, UK: Academic Press.

Rancourt, E. (2019). The scientific approach as a transparency enabler throughout the data life-cycle. *Statistical Journal of the IAOS*, 35(549-558).

Rosenbaum, S. (2010). Data Governance and Stewardship: Designing Data Stewardship Entities and Advancing Data Access. *Health Services Research*, 45:5-2 (1442-1455).

Standards Council of Canada. (2020). *What Are Standards?* Standards Council of Canada. Retrieved from

<https://www.scc.ca/en/standards/what-are-standards>

Statistics Canada. (1998). Quality Guidelines (3<sup>rd</sup> ed.). Ottawa, Canada: Authority of the Minister responsible for Statistics Canada. Retrieved from:

<https://www150.statcan.gc.ca/n1/en/pub/12-539-x/4194542-eng.pdf?st=2Q91Jgsf>

Statistics Canada. (2019). *Measuring investment in data, databases and data sciences: Conceptual framework*. 13-605-X. Retrieved from: <https://www150.statcan.gc.ca/n1/pub/13-605-x/2019001/article/00008-eng.htm>.

Statistics Canada. (2020a). *Data Literacy Competencies*. Statistics Canada. Retrieved from:

<https://www.statcan.gc.ca/en/wtc/data-literacy/competencies>

Statistics Canada. (2020b). *Statistics Canada Data Strategy: Delivering insight through data for a better*

*Canada* [PDF] Retrieved from [https://www.statcan.gc.ca/eng/about/datastrategy/statistics\\_canada\\_data\\_strategy.pdf#:~:text=The%20Statistics%20Canada%20Data%20Strategy%20%28SCDS%29%20will%20be,of%20highquality%20data%20and%20information%20to%20support%20evidence-based](https://www.statcan.gc.ca/eng/about/datastrategy/statistics_canada_data_strategy.pdf#:~:text=The%20Statistics%20Canada%20Data%20Strategy%20%28SCDS%29%20will%20be,of%20highquality%20data%20and%20information%20to%20support%20evidence-based)

Statistics Canada. (2020c). *Principles of Necessity and Proportionality*. Statistics Canada's Trust Centre.

Statistics Canada. Retrieved from: <https://www.statcan.gc.ca/en/trust/address>

Statistics Canada. (2021a). *Statistics Canada's Approach to Data Stewardship* [PDF]. Unpublished internal departmental document.

Statistics Canada. (2021b). *Enterprise Information and Data Management Glossary* [PDF]. Unpublished internal departmental document.

Statistics Canada. (2021c). *Introduction to the National Occupational Classification (NOC) 2021 Version*



1.0. Statistics Canada. Retrieved from:  
<https://www.statcan.gc.ca/en/subjects/standard/noc/2021/introductionV1#a1>

System Analysis Program (SAP). (n.d.). *Business and Technology Fundamentals: Data management Glossary*. SAP. Retrieved from: <https://www.sap.com/insights/data-management-glossary.html>

Treloard, A., Groenewegen, D. and Harboe-Ree, C. (2007). The Data Curation Continuum: Managing data objects in institutional repositories. *D-lib Magazine*, 13:9/10. Retrieved from: <http://www.dlib.org/dlib/september07/treloar/09treloar.html>

United Nations General Assembly. (2014). Resolution adopted by the General Assembly on 14 January 2014 -- 68/261. Fundamental Principles of Official Statistics. Sixty-eighth session. A/68/L.36 and Add.1. Retrieved from <https://unstats.un.org/unsd/dnss/gp/FP-New-E.pdf>

United Nations Statistics Division (UNSD). (2021). Fundamental Principles of National Official Statistics. United Nations. Retrieved from <https://unstats.un.org/fpos/>

United Nations Statistics Division. (2021). UN Global Platform: Data for the World. Retrieved from <https://unstats.un.org/bigdata/un-global-platform.cshtml>

Wells, D. (n.d.). What is a Data Catalog? *The Data Intelligence Blog* by Alation. Retrieved from <https://www.alation.com/blog/what-is-a-data-catalog/?nowprocket=1>

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. Retrieved from: <https://www.nature.com/articles/sdata201618#citeas>

United Nations General Assembly. (2014). Resolution adopted by the General Assembly on 14 January 2014 -- 68/261. Fundamental Principles of Official Statistics. Sixty-eighth session. A/68/L.36 and Add.1. Retrieved from <https://unstats.un.org/unsd/dnss/gp/FP-New-E.pdf>

United Nations Statistics Division. (2021). Fundamental Principles of National Official Statistics. United Nations. Retrieved from <https://unstats.un.org/fpos/>