

Seventy-first plenary session

Geneva, 22-23 June 2023

Item 5(b) of the provisional agenda

**DATA STEWARDSHIP AND THE ROLE OF NATIONAL STATISTICAL OFFICES
IN THE NEW DATA ECOSYSTEM**

D R A F T

Prepared by the CES Task Force

The CES Bureau set up a Task Force on data stewardship in February 2021 to clarify the terms related to data stewardship and public data governance, and the tasks of NSOs that this may include in different settings. The Task Force consisted of: Austria, Belgium, Canada, Denmark, Estonia (Chair), Finland, Hungary, Ireland, Kazakhstan, Mexico, New Zealand, Netherlands, Poland, Switzerland, UK, Bank of International Settlements (BIS), Eurostat, OECD and UNECE. The members represented national statistical offices and other government agencies, central banks and geospatial agencies.

The first draft of the document was consulted electronically with all CES members in spring 2022. The feedback provided support and many constructive comments to improve the document.

The Task Force has updated the document, taking into account the comments and the latest developments. Two supplementary documents have been developed to complement the main report – a summary, which can be used as a self-standing document to explain NSOs' role in data stewardship, and a document with country examples of how NSOs in different countries are pursuing their role as data stewards.

The CES Bureau reviewed the draft 'Data stewardship and the role of national statistical offices in the new data ecosystem' in February 2023 and asked the Secretariat to send it for electronic consultation to all CES members before the 2023 CES plenary session. Subject to a positive outcome of the consultation, CES will be invited to endorse the document.

*Please provide your comments **by 10 May 2023** using the [online feedback form](#).*

Executive Summary

The rapidly expanding and evolving landscape of data production and sharing is an opportunity for National Statistical Offices (NSOs) to exercise their expertise in the larger statistical, as well as national data community. Indeed, because of this environment, NSOs must expand their role into data stewardship in order to fully carry out their mandate. As the transformation is multifaceted, there are different interpretations of how the NSO role can and should change, and what data stewardship means in this context.

In this report, the United Nations Economic Commission of Europe (UNECE) Task Force (created by the Conference of European Statisticians in 2021) has examined the issue of data stewardship through the lens of NSOs. Data stewardship is viewed as operating *in service of*, rather than *in control of* the data ecosystem and reflects an overall paradigm shift or evolution of the framework that guides how NSOs operate, moving from the production of statistics to the provision of data and data-related services.

The work of this Task Force has focused on defining and explaining the foundational concepts and frameworks relevant to data stewardship. More concrete guidance and recommendations could be developed at a later stage, as a follow-up to this report. The Task Force has considered further steps to support NSOs in implementing their stewardship role, which are summarised in Section 1.4 'Future work'.

Ultimately, the report aims to serve as a guide to statistical offices regardless of data or structural maturity. It will help offices decide whether they should extend their role as government data stewards, as well as provide inspiration and ideas to those who are already performing this function. Public data holders will also gain a deeper understanding of the advantages and services provided by NSOs in the future data ecosystem. Below is a high-level summary of the report *as a whole*, followed by a distillation of the key messages that the report communicates *specifically* with respect to data stewardship.

Report Highlights:

- The development of complex data supply chains, new technological developments, and increasing importance of infrastructure for data sharing and access requires monitoring and maintenance. This has put a **renewed focus on public trust** and subsequently, highlighted the importance of data stewardship and governance (Chapter 2).
 - NSOs can offer competencies and capabilities to contribute to better governance of data in the public sector and beyond, while bearing in mind the legal, organisational and technical aspects that shape NSOs' role in this area.
- Common definitions of data governance and data stewardship are necessary, and can be leveraged in the context of official statistics and across the public sector (Chapter 3).
 - **Data stewardship is a particularly important concept** in current data operating environments, especially in the way it promotes a holistic perspective on the management of data through clear governance accountabilities that emphasizes intergenerational guardianship and public good.
- **Data stewardship is enabled through clear and authoritative data governance.** To organise data governance activities in the public sector, three types of enabling governance models include **centralised** (top-down), **distributed** (bottom-up), and **federated** (hybrid) data governance (Chapter 4).

- The centralised model is characterised by consolidated governance authority, and the distributed model by independent governance. The federated model is a hybrid model with a central authority that enables input from agencies, and is the model best suited for an all-of-government approach to data stewardship.
- Selection and implementation of a data governance model will depend heavily on national contexts.
- The successful delivery of data stewardship relies heavily upon people’s understanding of stewardship, and especially how it relates to their **roles and functions** – both within an agency and at the system-wide level (Chapter 5).
 - The roles and corresponding competencies of a data steward can be explored through examples and case-studies from different countries, each with their own unique context and particular lesson-learned.
- Measuring data stewardship performance is relatively new territory, so a data stewardship maturity assessment model is proposed. A **data stewardship maturity model** can embed important ideas and values into the understanding of what constitutes sound data stewardship and, by extension, into organisational structure and practice, to steer agencies in the right direction in the digital data ecosystem (Chapter 6).
 - A data stewardship maturity model can help clarify the meaning of data governance and highlight the ways it is meant to operate in conjunction with, and help deliver, data stewardship. One such possible maturity assessment model is presented, based on an example implemented in New Zealand.
- Though data stewardship can represent a challenging concept to understand, the importance of data stewardship and NSOs’ changing roles in performing it can be effectively communicated using **enabling products, services, and partnerships** (Chapter 7).
- This commitment to data stewardship and addressing the changing data landscape can result in new partnerships and data sources which decrease the response burden and increase the timeliness, quality, interoperability, reusability, and fitness-for-purpose of data, contributing to **better insights and services to citizens** (Chapter 8).

Key Messages about Data Stewardship

The concept of **data stewardship** is fundamental to data best-practice (especially in an evolving ecosystem) and to the future of the work of NSOs. To provide clarity and consistency in this subject, the Task Force has developed a cohesive definition:

Data stewardship means ensuring the ethical and responsible creation, collection, management, use and reuse of data so that they are used for public good and benefit the full community of data users.

The report will convey the following key messages about data stewardship:

Data stewardship is necessary to maximize the value of data assets.

- Stewardship enables data policy implementation and the treatment of data as a strategic asset.
- Stewardship promotes sharing and reuse of data assets subject to the right ethical and cultural conditions, thereby maximizing their value.
- Stewardship promotes open data, interoperability and ethical and culturally appropriate use of data, enhancing the timeliness and efficacy of decision making and enabling data-driven public service delivery.

Data stewardship has two main scopes: intra-agency and system wide.

- Internal data stewardship can help consistently promote best practices in working with data within an agency, including improving data quality and metadata-based production.
- System-wide data stewardship requires cooperation and coordination across agencies which goes beyond various data domains. Achieving the desired availability and interoperability of data is therefore a multi-party effort.
- There is no 'one-size-fits-all' approach to data stewardship and the needs of individual NSOs must be considered.

NSOs have inherent and unique expertise to lead data stewardship in the National Statistical System (NSS) and to take on data stewardship responsibilities across the national data ecosystem.

- In their role of producing quality statistics, NSOs have always been engaged in data stewardship. This unique experience puts NSOs in a position to leverage its experience and functional expertise to take on new tasks and expanded roles related to data stewardship.
- The data stewardship role of the NSO can cover responsibilities in system-wide data stewardship related to public sector data management, depending on the national context.

Data stewardship improves data quality and relevance of NSOs and the National Statistical System.

- Data stewardship can support the delivery of high-quality data in all countries. Data stewardship brings added value to NSOs and the NSS, factoring for the national context affecting its implementation.

Data stewardship strengthens public trust in official statistics and in data management across the public sector

- Provision of data-driven public goods and services requires sound data governance, public trust and data reliability.
- Data stewardship supports the Fundamental Principles of Official Statistics which enable privacy, high quality and public trust.

Data stewardship needs governance, capabilities and resources for successful implementation.

- Data stewardship goals should be embedded in data governance - *a system of decision rights and accountabilities for the management of data, and the resulting laws, regulations, policies and frameworks that provide enforcement.*
- Data stewardship requires co-ordinated investments and capabilities, in transparent dialogue with the public and the budget decision makers to ensure progress and needed investments.
- A maturity model can be used for self-evaluating the gaps between a current state and desired end-state of data stewardship. This allows the definition of appropriate governance structures and planning investments.

Acknowledgements

This document has been prepared by the UNECE Task Force on data stewardship representing national statistical offices and other public agencies. The following persons have contributed to the report:

Tobias Thomas, Djordjevic Ana, and Thomas Burg (Statistics Austria), Nathalie Delattre (National Geographic Institute Belgium), Anil Arora, Tom Dufour, Stéphane Crete, Crystal Sowards and Courtney Cameron (Statistics Canada), Niels Ploug and Maciej Truszczynski (Statistics Denmark), Ott Velsberg, Urmet Lee, Veiko Berendsen and Raivo Ruusalepp (Statistics Estonia), Timo Koskimäki, Ville Vertanen, Essi Kaukonen and Leena Storgårds (Statistics Finland), Eszter Nagy (Hungarian Central Statistical Office), Pádraig Dalton, Patrick Kelleher and Paul Morrin (Central Statistics Office Ireland), Ms. Assem Kaïrgeldina (Bureau of national statistics Kazakhstan), Sergio Carrera (National Institute of Statistics and Geography Mexico), Mark Sowden, Craig Jones, Kevin Sweeney, Howard Hutchins-Pondand Julia Carter (Stats NZ), Erwin van Mierlo and Wim Stolk (Statistics Netherlands), Dominik Rozkrut and Olga Swierkot (Statistics Poland), Manuela Lenk and Matthias Steffen (Federal Statistical Office Switzerland), Nicola Shearman (Office for National Statistics United Kingdom), Bruno Tissot (Bank for International Settlements), Mariana Kotzeva, Emanuele Baldacci, Albrecht Wirthmann and Konstantinos Giannakouris (Eurostat), Barbara-Chiara Ubaldi and Jacob Arturo Rivera Perez (OECD), Tiina Luige and Jonathan Gessendorfer (UNECE).

The preparation of Chapters was led by the following countries:

1. Introduction – Estonia
2. National statistical offices and data stewardship – Estonia
3. Defining data stewardship and data governance – Canada
4. Data governance models and data stewardship – New Zealand
5. Roles and responsibilities for data stewardship – Ireland
6. Data stewardship maturity models – New Zealand
7. How to communicate data stewardship – Poland
8. Conclusion: A foundation for NSO data stewardship leadership – New Zealand

Courtney Cameron from Statistics Canada did the language editing of the text.

Contents

- Executive Summary 2
- Acknowledgements 5
- 1 Introduction..... 8
 - 1.1 Background..... 8
 - 1.2 Work of the Task Force 9
 - 1.3 Structure of the report 10
 - 1.4 Future work 10
- 2 National statistical offices and data stewardship 12
 - 2.1 Context 12
 - 2.2 What NSOs have to offer..... 13
 - 2.3 What is to be gained, and what is at stake?..... 16
 - 2.4 Enabling environment 17
- 3 Defining data stewardship and data governance..... 21
 - 3.1 Data stewardship and related terms..... 21
 - 3.2 The role of a data steward 25
- 4 Data governance models in the public sector and data stewardship 26
 - 4.1 Governmental Data Governance Models..... 26
 - 4.1.1 Centralised model 26
 - 4.1.2 Distributed model..... 27
 - 4.1.3 Federated Hybrid model 28
 - 4.2 Data stewardship in different data governance systems..... 29
- 5 Roles and responsibilities for data stewardship 30
 - 5.1 System-wide data stewardship 31
 - 5.1.1 Sound data stewardship – a public sector goal..... 31
 - 5.1.2 Support to other public sector data providers..... 37
 - 5.2 Roles associated with intra-agency data stewardship 40
 - 5.2.1 Generic roles required to implement data stewardship in NSOs 40
 - 5.2.2 The role of Data Steward within an NSO..... 41
 - 5.2.3 Other roles related to Data Stewardship 42
 - 5.3 Human resource requirements for the data steward role..... 43
 - 5.3.1 Secondments 44
 - 5.3.2 Partnerships 45
 - 5.4 Services outside of the NSO mandate that could be facilitated by NSO stewardship 46
- 6 Data stewardship maturity assessment model..... 47
 - 6.1 Maturity models..... 47
 - 6.2 Benefits of a maturity model for data stewardship 48
 - 6.3 Using a maturity model to assess data stewardship..... 50
 - 6.3.1 Maturity dimensions for data stewardship 50
 - 6.3.2 Reporting data stewardship model insights..... 51
 - 6.4 Supporting improved stewardship across the data ecosystem 52
 - 6.4.1 Data stewardship maturity model outcomes..... 52
- 7 How to communicate data stewardship? 54
 - 7.1 A complex concept – how to communicate it? 54

7.2	Possible messages for communicating data stewardship.....	55
7.3	New products, services and partnerships as elements of communicating data stewardship 56	
7.4	Critical Communication	57
8	Conclusion: A foundation for NSO data stewardship leadership.....	59
	Annex 1. Glossary of data stewardship concepts.....	61
	Annex 2. Data Stewardship and the Fundamental Principles of Official Statistics	72
	Principle 1: Relevance, impartiality and equal access.....	72
	Principle 2: Professional standards and ethics.....	73
	Principle 3: Accountability and transparency.....	73
	Principle 4: Prevention of misuse.....	74
	Principle 5: Sources of official statistics	74
	Principle 6: Confidentiality	74
	Principle 7: Legislation.....	75
	Principle 8: National coordination.....	75
	Principle 9: Use of international standards.....	75
	Principle 10: International cooperation	75
	Core values of official statistics	76
	1. Relevant.....	76
	2. Impartial	76
	3. Transparent	76
	4. Professionally independent.....	77
	5. Respects confidentiality	77
	6. Collaborative	77
	Data stewardship role outside the national statistical system and the fundamental principles.....	77
	Annex 3. Supporting principles of data stewardship	78
	Standards.....	78
	A Human Rights-based approach to data	79
	The FAIR guiding principles of data	80
	The CARE principles for indigenous data governance	80
	The 4Gs Model of Data Lifecycle Management	82
	The ‘once only’ principle	83
	The European Interoperability Framework.....	84
	The Necessity and Proportionality Principle	84
	References.....	86

1 Introduction

1.1 Background

1. Statistical offices operate in an increasingly interconnected, dynamic and digitized world, which brings with it new demands and new opportunities. Data has become a vital strategic asset and as such, many organizations outside official statistics create statistical output using new data sources, tools and techniques. This has also changed stakeholders' expectations for more open and detailed data that is both quickly and easily accessible. The COVID pandemic has further tested official statistics' ability to harness quality and timely data in a responsible way, accelerating a modernisation in how NSOs, public sector agencies and private organisations deal with data.

2. The data and information potential from leveraging partnerships and linking data from multiple sources, raises questions related to ethics, privacy protection, data security, accessibility, harmonisation of concepts, formats and technologies used, knowledge of data available, and how data will be managed and governed in the long-term. To respond to these questions, this report will define the term 'data stewardship' and explain its functions and competencies to enable access to and re-use of data for public benefit in a systematic, sustainable and responsible way (Verhulst, 2021).

3. To move towards a common understanding of the term data stewardship, the Conference of European Statisticians (CES)¹ set up a Task Force in 2021 to help bring clarity to this area and facilitate the work of NSOs in adapting their role to the changing situation. The following countries and organizations are members of the Task Force: Austria, Belgium, Canada, Denmark, Estonia (Chair), Finland, Hungary, Ireland, Kazakhstan, Mexico, Netherlands, New Zealand, Poland, Switzerland, UK, Bank for International Settlements (BIS), Eurostat and Organisation for Economic Co-operation and Development (OECD). The Task Force members represent statistical offices and other public institutions, and the UNECE is providing the Secretariat.

4. A draft version of the document was circulated to the countries and organizations participating in the work of the Conference of European Statisticians (CES) in spring 2022 for consultation. The present draft document incorporated the vast and varied feedback, and latest development. The CES Bureau reviewed it in February 2023, and asked the Secretariat to send it for wide consultation. Subject to a positive outcome of the consultation, the CES 2023 plenary session will be invited to endorse the document.

5. The report aims to serve as a guide to statistical offices regardless of data or structural maturity. It will help offices decide whether they should extend their role as government data stewards, as well as provide inspiration and ideas to those who are already performing this function. Public data holders will also gain a deeper understanding of the advantages and services provided by NSOs in the future data ecosystem.

6. The role of national statistical offices in the data ecosystem has been widely discussed during recent years. However, the discussions vary considerably in focus and there is no common understanding on what data stewardship is, or what the roles of national statistical offices (NSO)² in this evolving context can (or should) be. Relevant questions include:

¹ The Conference of European Statisticians is the intergovernmental body on statistics for the UN Economic Commission of Europe. Its members are the heads of statistical offices of UNECE, OECD and other countries (approximately 65 member countries).

² 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.

- What is data stewardship? What is its relationship to data governance?
- What role might NSOs play in data stewardship in the public sector, especially in cooperating with other components of the NSS, or more widely in a country's data ecosystem?
- How will NSOs respond to competition from other organisations and what strategic partnerships are needed?
- What criteria is used by NSOs when deciding how to manage their position as data stewardship subject matter experts, for the public sector? What does their leadership look like?

1.2 Work of the Task Force

7. To address the evolving digital data ecosystem and NSOs' changing role therein and to come to a common understanding of data stewardship, in 2021 the Conference of European Statisticians (CES)³ set up a Task Force. The UNECE Task Force was mandated to:

- Define 'data stewardship' and related terms.
- Develop a set of possible responsibilities of a data steward and potential considerations for shaping the NSO role(s) in this area. This can serve as a basis for developing a minimum core set of responsibilities and a maturity model of data stewardship.
- Develop a toolkit of tools and best practices 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.

8. The work follows discussions and resulting request by the CES and its Bureau for proposals for practical ways to enhance the role of NSOs in the new digital data ecosystems and in public data governance. An important background document for these discussions was a collaborative, international paper led by Estonia (United Nations, 2020) which outlined important considerations for NSOs when charting a way forward in data stewardship. The Conference noted that statistical offices would be naturally suited to provide, at minimum, an enabling and supportive role in data stewardship in their respective data ecosystems, but that many NSOs would be well-positioned to provide significant leadership in this area, based on their extensive experience with ensuring data quality, protection of privacy, and managing data and metadata.

9. When preparing this report, the Data Stewardship Task Force discussed the changing role of NSOs both theoretically and practically. With regular teleconferences including participation from over 19 NSOs and intergovernmental or international organisations, the perspectives of a wide range of NSOs were shared and considered. The Task Force collaborated to characterize and explain the digital data landscape, clarified foundational concepts related to data stewardship, established strategic directions and determined a range of feasible possibilities for the role of NSOs in the changing data ecosystem, contingent upon their national contexts.

10. The work of this Task Force has focused on defining the concepts and looking at possible tasks and responsibilities of data stewards. More concrete guidance and recommendations could be developed at a later stage, as a follow-up to this report. The Task Force has considered possible further steps to support NSOs in implementing their stewardship role, which are summarised in Section 1.4 'Future work'.

³ The Conference of European Statisticians is the intergovernmental body on statistics for the UN Economic Commission of Europe. Its members are the heads of statistical offices of UNECE, OECD and other countries (approximately 65 member countries).

1.3 Structure of the report

11. The report examines data stewardship through the lens of the NSO and takes steps to identify what role it may play in the evolving ecosystem, considering their mandates, experience and competencies. It also addresses data stewardship from a public governance perspective, providing strategic direction to NSOs and public sector agencies, and providing suggestions regarding the role of NSOs in public governance, and among other stakeholders in data stewardship. Public data holders will gain a deeper understanding of the advantages and value from services provided by NSOs in the future data ecosystem. The report makes use of the ample literature on data stewardship and data governance from the private sector and academia while firmly rooting it in the context of public sector and official statistics.

12. The report:

- Describes the wider setting of NSOs' data stewardship opportunities and why NSOs should be taking a stronger role in stewardship of public data (Chapter 2) – what strengths NSOs have that enable the extension of their mandate with respect to data stewardship, what external enablers can support this process, what concerns this may raise, etc.
- Defines 'data stewardship' and related terms, discusses how these relate to the Fundamental Principles of Official Statistics (Chapter 3, Annex 1 and Annex 2)
- Describes the current state of data governance models and principles (Chapter 4 and Annex 3)
- Describes the responsibilities and skills of data stewards (Chapter 5)
- Presents the foundation of a model to assess the maturity of data stewardship (Chapter 6)
- Discusses ways of communicating the role of NSOs in data stewardship (Chapter 7)
- Demonstrates the pathways NSOs in different countries have taken to progress their role as data stewards (document ECE/CES/BUR/2023/FEB/16/Add.1).

1.4 Future work

13. When the Task Force was set up in February 2021, it was noted that the report could serve as a basis for developing further guidance and recommendations on the role of NSOs in data stewardship. In international discussions on this topic, countries have asked for practical guidance on how to implement the Data Stewardship role.

14. To provide support for practical implementation, the Task Force has considered the following as potential next steps:

- Develop Basic Principles of Data Stewardship by monitoring and analysing implementation of data policies and data governance initiatives by NSOs
- Provide more specific guidance and/or recommendations and tools for implementing data stewardship
- Develop a generic roadmap for how to increase NSOs leadership role in data stewardship (taking as a starting point a [paper by Estonia et al.](#), presented at the 2020 CES plenary session)
- Identify a core set of responsibilities of NSOs as data stewards in national data ecosystems
- Compile a knowledge base of best practices and examples of successful data stewardship implementation, including more detailed examples of how data stewardship is implemented and the data services provided in practice

- Collect examples of maturity models and other assessments of data stewardship, and on this basis, develop and test a generic maturity model for the assessment of NSO data stewardship
- Prepare guidance materials for different audiences or focusing on data stewardship in specific areas - administrative data, open data, privately held data, geospatial data, governance models etc.
- Provide a forum for exchange of experience in implementing data stewardship
- Maintain the glossary (consistent with the work of the United Nations Statistics Division's global Working Group on Data Stewardship) and facilitate the translation of key terms into other languages.

15. In undertaking any follow-up activities, it is important to consider other global and regional work on this topic, including the above-mentioned global Working Group on Data Stewardship (WGDS) with its five work streams, and the work in the Economic Commission of Latin America and the Caribbean (ECLAC) and The Economic and Social Commission for Asia and the Pacific (ESCAP) regions. Any follow-up proposals will be submitted to the CES Bureau in October 2023.

COVID-19 Impact

The COVID pandemic required many NSOs as well as their NSS counterparts to respond rapidly and accelerate modernisation (Rosolia et al., 2021). There became an urgent need for information that was collected and produced in a maximally efficient way, as important data-driven decisions, vital services and ultimately people's lives were at stake. To address this need, intergovernmental and cross-sectorial collaboration leveraged new methods of collecting and processing data, new data sharing partnerships and new technologies. 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. However, regulatory and legislative barriers, as well as the perception of risk, restricted the use of information for specific purposes.

For NSOs to meet today's demand for data and information, we must ensure that the legal, technical and administrative environment are appropriately enabling. We need to establish these more agile, timely, and effective methods and tools for data acquisition and sharing as appropriate even outside the context of crisis, to improve the flow of value to citizens. 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

16. With the exponential growth in data generated worldwide and the arrival of technologies such as artificial intelligence (AI) and edge computing, public authorities worldwide are increasingly recognizing the importance of data. In this context, it becomes vitally 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.

17. 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. This consumer data, combined with other – data (whether public or private) offer enormous potential to discover patterns, infer indicators, develop business models, or make decisions for the common good. Access to these data pools is currently restricted due to intense market competition, regulations on privacy and competition, as well as the implications associated with the costs of obtaining and sharing information. The grassroots development of joint public and private sector initiatives has made it possible to explore the benefits and challenges of sharing information held by private businesses. However, this case-by-case approach has limited the potential to produce regular statistical information for public use.

18. 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 data to be obtained from other sources and for different purposes. These advances offer the opportunity to supplement the statistics that NSOs have traditionally been producing, thereby reducing administrative cost, and response burdens.

19. These developments offer both opportunities and challenges. Government agencies and other public institutions have their own information systems and data collections with rich histories, which have led to countless data holdings that can be enriched when supplemented with alternative data. In addition to reducing burden, there is huge potential for new types of data services, more timely and granular data, new insights by linking data from different sectors, sources and topics. However, there are also huge risks that need to be mitigated. Ungoverned data could be used unethically, increased digital capacity comes with an increased need for digital security, the ‘digital divide’ could become an ‘information divide’, and of course, NSOs are challenged to remain vigilant and proactive in the protection of privacy as these developments occur. Further, in order to benefit from these developments, public agencies will have to address issues around a lack of data accessibility, interoperability and standardization; data duplication and redundancy; and the costs of linking and sharing data.

20. To overcome these risks requires governmental policies with farther foresight, to enable the focus on things like strategy, culture, ethics, roles, and the capabilities of people that can support an effective data ecosystem. The concept of 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, accepted norms and adopted standards. Fortunately, a strong culture of coordinated collaboration has been implemented within the NSS, not least in the context of the application of the Fundamental Principles.

21. Data governance establishes and makes accessible the set of norms and standards associated with data. Agreed upon data governance norms will provide clarity and assurance in the way data is overseen and managed, and will support data availability, quality, security, usability and integrity. While data governance has a strong emphasis on technical competencies, data stewardship focuses

equally on the foundational concepts, frameworks, and best-practices to support the growing maturity of data policy.

22. Data stewardship implements the policies, standards and principles outlined by data governance. It encompasses functions and competencies to enable access to, and re-use of, data for public benefit in a systematic, sustainable, and responsible way. Data stewardship is operating in service of - rather than in control of - the data ecosystem.

23. National governments are taking steps to reduce policy siloes and make public sector data strategies, projects and initiatives more coherent (OECD, 2019). Countries such as the United States⁴, the United Kingdom⁵, the Netherlands⁶ and Germany⁷ have issued National Data Strategies to cover different elements of the data ecosystem ranging from data access and sharing, open data, data for AI, and data ethics in the public sector – all under a single policy instrument. At the international level, there are ongoing efforts to promote data sharing and data access, such as the G20 Data Gaps Initiative (DGI). Its new phase, which launched in 2022, covers among other recommendations, access to private and administrative sources of information and data sharing (FSB Secretariat and IMF Staff, 2022).

24. In other cases, data-related policy instruments are made a part of broader digitalisation 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).

25. In terms of institutional governance and the creation of enshrined leadership roles, some countries have opted to define specific Chief Data Officer (CDO) positions for the whole-of-government (e.g. the appointment of a Chief Data Officer for the Government of Canada in 2022), 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 Council⁸).

2.2 What NSOs have to offer

26. National Statistical Offices are mandated to produce and disseminate relevant data and information in coordination with other parts of the NSS, in order to understand the evolution of society, the environment and the economy. This mandate comes from the United Nations’ Fundamental Principles of Official Statistics and is enshrined in each country’s statistical legislation. For decades, NSOs have developed methods and techniques to produce the insights necessary to meet this mandate. Concurrently, they consistently and collaboratively promote a culture of using data and information strategically in making decisions for the public good.

27. The United Nations’ *Fundamental Principles of Official Statistics* provide internationally endorsed principles for 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

⁴ See <https://strategy.data.gov/>

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

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

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

⁸ See <https://www.cdo.gov/about-us/>

- scientific and standardized
- quality, timely, and with as low a response burden as possible
- confidential and appropriately used
- transparent and made public.

28. 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).

29. 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) (United Nations, 2021a). These Fundamental Principles were then reaffirmed and endorsed by the UN Statistical Commission at the global level (1994) and the UN General Assembly (1994) (United Nations, 2021a). 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 (United Nations, 2014). The link between Fundamental Principles and data stewardship is explained in more detail in Annex 2.

30. NSOs have attained a certain reputation as a result of the technical strength, reliability, and timeliness they have displayed during the diligent management of the information they collect. The track record of NSOs is a testimony to their experience and competence in working with data, including confidentiality and privacy protection, data ethics, ensuring comparability, use of standards and classifications, communicating insights, supporting and growing data literacy, etc. While NSOs are seen primarily as producers of statistics, their expertise outlined above and their experiences in their role in coordinating activities within the NSS could be leveraged as an asset for the public sector and beyond. In turn, NSOs can also benefit from other public sector agencies' experience in tackling these issues in their specific domains.

31. 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, as well as 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. Capabilities of NSOs to take on responsibilities in the data stewardship landscape

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

- legislative frameworks and legal authority to collect, process and disseminate data and information (based on the Statistical Legislation)

- well established data quality frameworks including standardised metadata, reference data and master data
- long history of data protection and leading-edge methods to protect privacy and confidentiality
- modern statistical methods for data analysis, data integration and linkage, data categorization via metadata and data visualization
- growing data science expertise grounded in statistical and mathematical theory, expertise in modelling and forecasting
- 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
- storytelling and data visualization expertise from data to insights to policy to services for better outcomes
- well established culture of coordination within the NSS, and access to resources and networks with other NSS members, and a long history of data partnerships and collaboration both domestically and internationally
- collaborative workspaces (protected cloud, data collaboratives, data trusts), experience in data-access services for on-line access to data by trusted and certified external users
- statistical and data standards and classification systems to give data meaning and drive interoperability
- 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
- emerging entrepreneurial spirit and culture of continuous learning and development of innovation ecosystems.

33. NSOs are therefore well-positioned to provide leadership in data governance and stewardship, as partners in a data ecosystem whose function is the co-creation of trusted smart statistics. The relevant visions, missions, mandates, and legal responsibilities of NSOs contribute to this unique position. The increasing digitalisation of society offers opportunities for NSOs' new role as public sector data stewards and as promoters of institutional data stewardship 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 for the same information from citizens and businesses twice). NSOs can play a key role in facilitating and supporting data access and sharing.
- **Performance:** Data can be used as a key asset for public service design and delivery. NSOs augment data-driven services by collecting 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 misinformation, it is fundamentally important to channel users of information and data to trusted data sources. NSOs can play a key role as sources of trusted data and to fact-check the information provided by other sources. NSOs knowledge on the production on statistics can also help to build

knowledge and capacity across public bodies for preparing effective data visualisations and avoiding misleading data representations.

34. Beyond sound data management and standards, stewardship also requires public acceptability and trust. Social acceptance and public engagement are directly linked to privacy, confidentiality, security and transparency, and they are necessary for NSOs to be able to perform their duties and fulfil their mandates. NSO's regular commitment or adherence to the principles of open data, transparent processes, ensuring confidentiality and security, and communicating the value of citizens' sharing their data, all constitute further evidence of their expertise and opportune placement within statistical systems.

35. Data stewardship is an extension of the role that NSOs have played in coordinating the national statistical system, where the key objective is to ensure good coordination between statistical agencies within countries (Fundamental Principles of Official Statistics 8) (United Nations, 2014). The NSOs role has the potential to go deeper into the statistical system to ensure interoperability, and to go wider, to the whole data ecosystem.

36. However, the implementation of this leadership role in data stewardship is highly context-contingent and will vary depending on national factors such as legal frameworks and NSOs' mandates, maturity and use of digital technology, the nature of the data ecosystems, the extent of cooperation achieved within the NSS and public acceptance. Hence, the extent to which NSOs can provide leadership in data governance and stewardship will vary from country to country.

37. NSOs can position themselves in the data ecosystem in different ways ranging from minimal change to the NSO role, to operating as a public sector data steward with the requisite governance authority. Internationally, NSOs are all in different places with respect to their maturation journeys, with some already having established cross-government leadership roles. There is no "one-size-fits-all" solution that can be applied to all countries. In moving towards data stewardship, it is important to consider the jurisdictional and legal context and recognize that NSOs will move forward at their own pace, and taking on a stewardship role that fits their purpose and environment.

2.3 What is to be gained, and what is at stake?

38. Change sometimes involves risk, and regardless of the decision that an NSO makes regarding data stewardship, there are risks:

- If NSOs fail to participate (at all or in a leading capacity) in the process of data ecosystem transformation, not only do they risk losing relevance but there is a high risk of setting up parallel data stewardship structures which will increase instead of reducing burden and cost. The outcome could be further fragmentation of public data holdings with little or no compatibility.
- On the other hand, there is also the risk of taking on too ambitious of a data stewardship role, one which may compromise NSO's core operations and lead to reputational losses or decreased data quality.
- Further, the aspects of cyber security cannot be downplayed. Even though NSOs already apply the highest IT security standards, as data stewards, NSOs are more likely to become potential targets for cyber-attacks and need to take action to be prepared for this.

39. When successfully carried out, the data steward role makes it possible for NSOs to deliver an important and much needed service to the public while also increasing the positive effect that they can have on data stewardship and quality across the public sector. It will demonstrate that NSOs can fulfil their core responsibility, while at the same time increasing the quality and quantity of data and

official statistics. NSOs will benefit by having secured an important role into 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).

40. Sound data stewardship has societal benefit and contributes to the public good because it enables ethical operation, which creates the trust, social acceptability and public support necessary for statistical work. It minimises data misuse and enables reuse, allowing statisticians to access data already in the ecosystem. It also facilitates data- and knowledge-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.

41. 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). In several countries NSOs have already been playing a key role in furthering data stewardship (see various country case studies, Annex 4). 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.

2.4 Enabling environment

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

Legal and regulatory frameworks

43. The production of official statistics in countries and the role of NSOs is based on statistical legislation. This legislation looks different country to country and may outline direct or indirect mandates or role designations to the NSO for data stewardship - ranging from full subject-matter leadership to a particular function as a part of a wider NSS framework. However, the absence of a legal mandate should not stop NSO from exploring the benefits of taking on a stewardship role in the data ecosystem.

44. In general, legal regulations concerning data and information are rather new compared with many other regulated issues. As they reflect the digital and data landscape, they have naturally been developed according to individual countries' contexts. Only countries in the European Union have harmonized legislation on issues like data protection. In addition to being responded to differently country to country, many data issues are constantly evolving, and as such, are in a perpetual drafting process at both the national and supra-national levels. NSOs are well positioned to play an active and leading role here, providing expertise as well as consistency and ultimately efficiency to the drafting process.

45. When constructing a legal environment to enable data stewardship, it 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, other policy instruments, etc.) under which this is done, and the mechanisms (federated, centralized, platforms, trusts, etc.) by which it will be achieved.

46. The adaptation of the legal framework must be consistent with national data strategies and compliment a flexible system that allows operational adjustments without the need to frequently modify the laws. That is, it must recognize the continuous expansion of the borders of the data ecosystem, resulting 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 information reuse mechanisms must be administrative in nature, ideally designed by consensus of the actors involved (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. The legal frameworks that govern NSOs' data stewardship activities, must support third parties to carry on research and data-driven projects. It also enables the identification of new third-party sources, to create and calculate statistical indicators. The technical and legal challenge is then to ensure that this information is regularly available, while 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 agencies in the data ecosystem, in particular members of NSS, take ownership and use the Fundamental Principles of Official Statistics as a common framework, as well as the established UN 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, several pieces of legislation haven been adopted in recent years that concern data stewardship. The General Data Protection Regulation (GDPR) (EU 2016/679) protects natural persons regarding the processing of their personal data, and on the free movement of such data. The processing of such data for statistical purposes is covered by special clause in the GDPR. The recently passed Data Governance Act (EU 2022/868) regulates conditions for the re-use, within the Union, of certain categories of data held by public sector bodies (PSBs). This includes data protected both as personal data and by statistical confidentiality.

Technological developments

51. The technological environment is constantly developing and organisations across jurisdictions are facing the challenge of how to benefit from these enabling technologies. The maturity of technology implementation differs widely from country to country, but regardless, there are several data stewardship-related requirements that NSOs should prioritize: (1) achieving the appropriate level of digitalization; (2) effective registers and use of digital services (i.e., using administrative data sources); (3) governed data architecture, and master and reference data management; (4) data openness and protection, data privacy and cybersecurity; (5) harnessing artificial intelligence (AI) technologies, machine learning and big data; (6) nation-wide information systems and economic sectors or domain data spaces.

52. The level of digitalization can be seen as a basic enabler. Not all countries have been able to successfully transition from paper records to digital ones, and some are unlikely to achieve this for

some time. Other countries are now primarily operating by using digital records and reports. In the countries where digitalization is weak or fragmented, developing this capability should be prioritised.

53. Many countries have digital registers and are providing digital services to citizens, agencies and businesses. Technological innovation in registry information systems and services is an enabler for cross-national and secondary use (re-use) of data generally, or even by the NSO, for producing statistics. This capacity is further increased when overall data architecture supports semantic interoperability of data, and when reference and master data management is promoted and implemented. This is also an important enabler for the once-only principle.

54. There is always a possibility of data loss, theft or misuse. Technology enables the protection of private data and statistical confidentiality. Privacy Enhancing Technologies (PET) enable the use of individual information while ensuring data privacy. As instances and efficacy of cyber-attacks increase, countries and NSOs must ready themselves to deal with these issues as they come. To this end, the international standard ISO/IEC 27001 (2022) is establishing requirements for information security management systems, and is widely used to implement, maintain and continually improve information security management.

55. The use of AI and Big Data have given data organizations and NSOs new opportunities. The 5 V's of Big Data (characteristics of velocity, volume, value, variety and veracity) (Ishwarappa & Anuradha, 2015), combined with machine learning potential, are broadening the data stewardship role. Some concepts like data lake and data exchange between different data rooms need to be further developed and iteratively implemented to enable the data steward to perform data services.

56. Some countries (e.g. Estonia) have set up nation-wide information systems that enable a complete overview of all registers and important information systems providing data services, even providing a semantic understanding of these services. Though it has not yet been fully implemented, the solution was designed to serve as a general data catalogue. Some economic sectors have data exchange technologies to enable exchange of data in their particular domain with established semantics, like XBRL-GL. The European Union has an initiative to create data spaces for specific domains, supporting exchange of data within a domain across countries. The data space of health services is the first to be realised.

Organisational interoperability

57. The extent to which NSOs can perform a data stewardship role, and the successful implementation of this role, depends on the maturity of the data ecosystem in the whole country. This maturity is impacted by legal and technological enablers, but also organisational enablers such as existing coordination arrangements, public service departmental data strategies, national data strategies, or other frameworks or systems that provide a course of action for managing and leveraging data as a strategic asset. When such strategies are developed, NSOs should be actively involved in the discussions based on their expertise and experience coordinating activities within the NSS, managing data collections, advancing interoperability, supporting and enabling standardisation, ensuring privacy and confidentiality, maintaining data security and growing data literacy.

58. Organisational enablers help NSOs and other agencies that are part of the data ecosystem (e.g. central banks) to accomplish their mandates by advancing interoperability. Along with certain technological developments, organisational enablers advance interoperability by establishing common standards, developing data literacy, enacting data governance frameworks and processes, and facilitating the discourse and collaboration necessary to facilitate culture change.

59. These organisational mechanisms work together to provide both the course of action for managing and leveraging an organizations data assets and expertise, and the tools to make it happen. This results in resources being stewarded to protect privacy and security, maintain public trust, and ensure the optimal use and reuse of data in order to provide service and value to citizens. The required resources will depend on the scope of the data stewardship role that the NSO is ready to take on: within NSS, concerning (selected) administrative registers, within the whole public sector, or even more widely (e.g. related to data held by the private sector).

60. Concerning the public sector, NSOs should determine what is possible or beneficial based on the institutional setting, legislative context, and its own resources and capabilities, among others. For example, the NSO may set up a metadata catalogue (as in Switzerland, see Annex 4), common identifiers for people and businesses (as in Ireland), a data lake (as in Lithuania, see Annex 4), or a common data dissemination platform. For such additional tasks, NSOs can claim resources for providing a service to the public sector itself, or these tasks can be achieved in partnership with other agencies.

61. Using organizational enablers, NSOs can also leverage the respective strength of their counterparts in the NSS. In particular, central banks may be effective in supporting NSOs' public coordination efforts, drawing on 1) their contribution in several countries to the production of the financial accounts as part of the System of National Accounts (SNA) framework, which is the cornerstone of official statistics (IFC, 2020); 2) the coordinating role that they can play with other agencies (e.g. financial supervisors, market regulators) involved in the collection of official data in the financial sphere (IFC, 2021); and 3) their experience in dealing with large, granular and complex data sets ("financial Big Data") (IFC, 2019).

62. An all-of-government data stewardship role may require significant resources, both human, technical and financial, that will take time to synchronise in some countries. Even though all countries could greatly benefit from the stewardship and coordinated use of data, as described in this report, significant investments into improving data infrastructure, developing the enabling legal framework and capacity building of human resources may be required first.

63. Data stewardship is **not only for developed countries** where the government data systems are already established. For example, an NSO should seek to influence the establishment of administrative registers to ensure that they are appropriate to use for producing statistics. Proposing common approaches for data architecture, standards and classifications before the public data holdings are set up, will allow to save resources and get out more value from data in the long run.

3 Defining data stewardship and data governance

64. 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.

65. This chapter describes and explains important concepts related to data stewardship and proposes definitions for endorsement and use. The definitions have been sourced from experts from public and private sectors and academia. The proposed definition is formulated from the viewpoint of official statistics.

3.1 Data stewardship and related terms

66. Data stewardship operates in a context of a data ecosystem, therefore, we start from defining this term. The data (including statistical data), along with the data subjects, a broad range 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, 2017b; StatCan, 2019). An ecosystem includes four main categories of actors (European Commission, 2017b):

- data generators – the primary sources generating data, whether actively or passively, such as consumers, customers, enterprises, or citizens
- data service providers – the actors participating in data value creation chain; those collecting, organizing, storing, processing, retrieving, sharing, using, reusing, restricting or destructing data, 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, digital or hard-copy, qualitative, quantitative, images, text, 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. 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). Data governance is 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 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.

69. **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, and the resulting regulations, policies and frameworks that provide enforcement**. This holistic approach of data governance 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.; IFC, 2021; OECD, 2008, 2019; Plotkin, 2021; StatCan, 2019, 2021b). Data governance enables the coherent implementation and co-ordination of data stewardship activities and increases the capacity (technical or otherwise) to better control the data value chain.

70. Data management is a discipline that directs and supports effective and efficient management of information and data in an organisation, from planning and systems development to disposal or long-term preservation. **Data management involves the development, execution, and supervision of plans, 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 (DAMA, 2017; Government of Canada, 2019; StatCan, 2019, 2020a, 2021b).

71. The governance of data can be conceptualized using the “Governance ‘V’ Model” (Ladley, 2020), which is useful for the understanding of the subtle differences between data governance and data management. On the left is data governance – the authorities and policies previously discussed. The right side shows the “hands-on” management of data and information (necessary for sound data stewardship). This model emphasizes that data governance is NOT just a function performed by those who manage data (Ladley, 2020). It also serves to explain the complex relationships between data governance, data stewardship and data management.



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

72. 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 standardised, easy to find and of high quality. These concepts have a somewhat cyclical, mutually enabling relationship, but the distinction between them is important, both conceptually and practically.

73. A steward is someone who manages or looks after something on behalf of someone else. In literature the concept of data stewardship is sometimes used to differentiate it from data ownership (McGilvray, 2021). In this report, the term data stewardship emphasizes that public sector data should not be treated as if they are owned by agencies, but that the data should be managed on behalf of and for the benefit of the whole society.

74. For this context, the Task Force proposes the following definition for data stewardship:

Data stewardship means ensuring the ethical and responsible creation, collection, management, use and reuse of data so that they are used for public good and benefit the full community of data users.

75. Data stewardship:

- is an approach to data governance that formalizes accountability for managing information resources on behalf of others
- implements the policies, standards and principles outlined by data governance
- is enabled through good data governance and data management, which provide the accountability and maintenance or oversight of data assets throughout their lifecycle to ensure their proper care, high quality, data security and confidentiality
- influences proactive and responsible data practices to help deliver data strategies, maintain trust and promote accountability
- is expressed through long-term and inter-generational curation of data assets
- 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
- it deals with methods and mechanisms of 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
- is applicable at all scales, from the national or data system level, to the organisation or enterprise level, to the individual or dataset
- manages and coordinates the interactions of different actors in the system
- has two main scopes, intra-agency and system-wide data stewardship.

76. (OECD, 2018; Plotkin, 2021; StatCan 2019, 2021b; Stats NZ, 2020). See also the **open data** definition in the glossary available in Annex 1.

77. 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 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 (European Commission, 2016). In enacting and applying the principles of data governance and data stewardship, there can be some overlap. While data governance is the authority and regulatory framework that guides or mandates activities that enable the treatment of data as a strategic asset, data stewardship is the daily enactment of this governance through the implementation of the policies, standards, and principles outlined by data governance (applied data stewardship, see possible roles and responsibilities in Chapter 5). This is why data stewardship often exists as a tangible element of many organisations' data strategies. Data management is an element or mechanism of data stewardship, and encompasses practical activities like searching for data, sorting data, adding metadata, analysing and cleaning data, etc.

78. Figure 3 describes the interplay of three core concepts: data governance, data stewardship and data management. The other terms represent a set of principles for how data management should be organised and implemented, or what conventions should be followed when doing so.

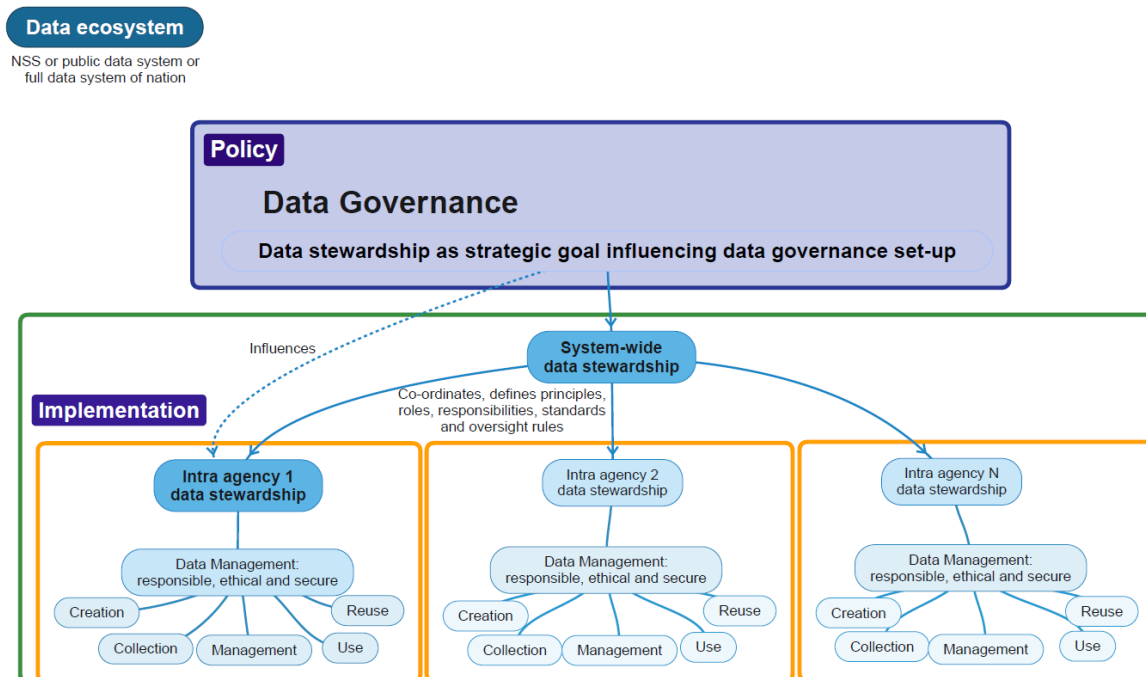


Figure 3: The relationship of data stewardship with data governance and data management in a data ecosystem

79. Data stewardship is necessary for the successful adherence to the first Fundamental Principle of Official Statistics, that “official statistics provide an indispensable element in the information system of a democratic society [...] to honour citizens’ entitlement to public information.” (United Nations, 2014). 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. If NSOs are going to adhere to the first of the Fundamental Principles in this dynamic data ecosystem, they must adapt to the developments therein. Data governance and data stewardship enable this adaptation.

80. Impartial official statistics not only help to build trust in the NSOs themselves, but they contribute to a broader goal of fostering transparency and accountability, building an open relationship between societies 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, having a slightly different relationship to issues of trust, public good, data ethics, balancing the use of personal data with individuals’ rights, inclusivity, and long-term guardianship of data and statistical information.

81. As such, 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) (European Commission, 2020; Floridi and Taddeo, 2016; StatCan, 2020).

82. The data stewardship and governance professionals working in this sphere are subject to data accountability. **Data accountability** is the ongoing liability and responsibility of an individual, enterprise, organisation or public authority regarding the management of data. This principle ensures

that appropriate technical and organizational measures and records are in place to establish 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).

3.2 The role of a data steward

83. The role assigned to the individual, team, or organisation 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 creation, acquisition, entry, quality, interoperability, and overall management supports the needs of consumers, citizens, enterprises, organizations, or governments, while also ensuring adherence to social acceptability, 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 2008, 2018; Plotkin, 2021; StatCan, 2021b).

84. There is a distinction required 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 public sector data. Further discussion on the role and responsibilities of a Data Steward can be found in Chapter 5.

85. 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.

86. 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 (Verhulst, 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 (Verhulst, 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.

87. Although the focus of this report is primarily on NSO's relationship to data stewardship roles and tasks, a wider concept of data governance points to functions in the public sector that need to be co-ordinated so that the full benefits and value of data stewardship can be utilised by everyone in the national data ecosystem. This means that NSOs must interpret its current institutional position within the public administration of the country in relation to the emerging data governance system. Chapter 4 provides an overview of the most common data governance models along with selection of country cases to illustrate possible combinations and solutions.

4 Data governance models in the public sector and data stewardship

88. As described in Chapter 3, data stewardship is enabled through clear and authoritative data governance and sound data management, which provides the necessary oversight of data assets throughout their lifecycle.

89. This chapter looks at data governance models in various contexts, according to where the locus of data stewardship responsibilities lies and the relationship between actors in this space. Three types of data governance models (centralised, distributed and federated hybrid) are presented here. While these models can be applied to a single organisation or at the sectoral level, this chapter focuses on its application in the public sector. It considers advantages and disadvantages of each model and the situational factors that contribute to their efficacy. A set of data governance principles that correspond with these data governance models are described in Annex 3.

90. Some such contextual factors, like 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) will therefore require attention when considering data governance arrangements. These data governance models should contribute to the value generated from data as an asset.

4.1 Governmental Data Governance Models

91. There are many different any data governance frameworks available. This chapter focuses on three categories of models for the organisation of data governance within the public sector. These are:

- Centralised data governance
- Distributed data governance
- Federated hybrid data governance.

4.1.1 Centralised model

92. 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/organization is accountable for data governance for the whole government (or organisation), and the data governance activity is centrally managed, funded and resourced.

93. The centralised model carries the following **opportunities**:

- It is easy to manage, thanks to the clear role of a dedicated Data Governance Lead
- It permits strong coordination through the data steward's team and more efficient decision making than in a distributed setting. This allows a more rapid response in crisis situations
- It makes it easier to focus on policy and develop guidelines, but also to control costs of data governance tasks.

94. However, there are several **challenges**:

- There is the potential for increased bureaucratic burden due to this models' linear structure, which can also lead to operational rigidity
- More time is required to accomplish data governance operations across the data ecosystem
- Tensions can arise between the locus of data governance authority and centrally mandated data requirements/standards, and the operational realities of individual agencies as they ensure the realisation of their own public mandates

- The centralised model is 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
- Centralized decision-making may hinder experimentation and creativity as well as good cooperation with other statistical agencies, who provide a substantial contribution to the NSS and have developed extensive experience.

95. While the centralised model can work well in small organizations with low data maturity, it is less suitable for an all-of government approach where multiple autonomous and experienced 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

96. In a distributed data governance model, individual agencies will have independent governance processes and are likely to have different governance models at the level of the individual agency. There will be no central authority responsible for co-ordinating data governance and stewardship. Examples of this model are described in Appendix 4, see for example Argentina or New Zealand.

97. As opposed to the centralised model, this model is characterised by a bottom-up approach. Public sector agencies own and operate independent governance processes. It is likely to lead to differing models, maturities and can contribute to difficult coordination between agencies.

98. This model's **opportunities** are:

- Empowering agency senior executives to make decisions that meet their agency's needs
- Is relatively easy to establish, as there are clear rules for responsibilities and processes and organizational data governance models will be fit-for-purpose
- Lower bureaucratic burden, as decisions are made on agency level.

99. But there are several **challenges**:

- Agencies' data governance tends to be inwardly focused – there is no external agent to explicitly coordinate inter-agency work. Some collaboration is possible, but often difficult to sustain, and reaching a consensus tends to take longer without authoritative coordination
- The distributed model can make it difficult to coordinate and commit the necessary resources from agencies, particularly as funding tends to be focused on individual agency needs and projects, rather than interagency requirements
- It is challenging to address government-wide initiatives or ensure interoperability and data sharing, because of a lack of inter-agency data standards. Agencies are therefore more siloed, which reduces the capacity for agencies to integrate data
- The model lacks a co-ordinated approach to data governance and data stewardship across agencies, so there is increased potential for redundancy and the emergence of incompatible models in siloed agencies.

100. This model works best in small or medium size agencies/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 interoperability possible. For this to function effectively at an all-of-government level, there must be some degree of coordination of data stewardship activities.

New Zealand

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. This is a very decentralised model of government administration and is closest to the distributed model of data governance, as individual government departments are operationally autonomous. There are two key governance bodies supporting the role of the Government Chief Data Steward (GCDS), the Digital Government Leadership Group and the Information Group (see New Zealand in Annex 4).

4.1.3 Federated Hybrid model

101. In this model, there is still a centralized structure which oversees the public sector's data environment but there is also bottom-up input, enabling participation from individual 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 in line with their respective public mandates. It also enables data flows from the central locus of authority to federated agencies and from the federation to the centralized structure. Examples of this model are described in Appendix 4, see for example Canada, Finland, Mexico, or Switzerland.

102. This model is characterised by a combination of centralized control and 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 organization is key.

103. The decision-making regarding standards and shared data is centralised, enabling interoperability. 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.

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

105. The **opportunities** the Federated Hybrid model offers are:

- Individual public sector agencies are responsible for their own data and metadata in line with their particular public mandates. They have the autonomy to develop standards, policies and procedures at the individual agency level, but require that these align with an all-of-government model for data sharing to ensure interoperability
- Provides the ability to focus on specific data sets at the level of the individual agency and how these datasets relate to all-of-government data
- The model allows for broad membership for working groups; these should also include those responsible for data and decision making
- Issue resolution relies on a bottom-up approach that, nevertheless, must have clear accountability and process (e.g. a new or existing committee)
- The model is relatively easy to establish.

106. The **challenges** of the hybrid model include:

- A highly skilled Data Governance and Stewardship lead position or team is required full-time
- Clear roles and responsibilities for collaboration and knowledge sharing are required
- May require strong levers, including legislation, to drive adoption of all-of-government standards, metadata and models
- Discussions at the working group level may get political. 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 public sector agencies
- Oversight of the autonomy of individual public sector agencies can be challenging and relies heavily on self-reporting.

107. The Federated Hybrid model tends to work well in large organizations with varying levels of data maturity and longstanding experience and contribution to the NSS. It is the most practical model for a cohesive, government-wide approach to data governance and stewardship.

108. The principle of the hybrid model requires establishing norms (laws, administrative provisions, etc.) that are necessary for data reuse. This 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 effectively.

Mexico

The National Institute of Statistics and Geography (INEGI) is an autonomous body, completely independent from the federal government. 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 to produce official statistics by different government agencies. The Mexican system can be classified within the Federal Hybrid model, as it has direct control of the production cycle of census and survey information to produce 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.

4.2 Data stewardship in different data governance systems

109. The centralised model can work in small organizations, where the data stewardship role would be established within this organization (explicitly or implicitly), linked with data governance and data management. However, it is not feasible to use the centralised model for a government-wide approach to data stewardship involving multiple autonomous agencies.

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

111. The federated hybrid model is best suited for an all-of-government approach, but is the most complex to implement effectively. Many different iterations and combinations of this model are possible depending on the institutional structure and 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 for the whole-of-government, and data stewards in individual agencies.

112. Annex 3 provides examples of supporting principles of data stewardship and other enabling pieces of the data governance puzzle, organized by data governance model. These concepts help facilitate governance activities and drive sound data stewardship by describing roles and some of the skills expected of data stewards (see also Chapter 5). These concepts can be implemented in any of the three models described in the previous section.

113. Annex 4 provides examples of data governance models and frameworks used by different countries.

5 Roles and responsibilities for data stewardship

114. Data stewardship deals with strategy, culture, ethics, roles, 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 or value chain. It can be exercised at the intra-agency level, and at the system-wide level (e.g. the national statistical system, the whole-of-government or public sector).

115. 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 NSS, for a particular aspect of public sector data, or as a national data steward overseeing all public sector data. Within an organisation, there must always be a person or persons (unit) 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.

116. Within an agency, 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, 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.

117. At the system-wide level, the nature and structure of the government system must be considered. 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 were discussed in chapter 4). It will be useful to also define the role of a data steward in private organizations and companies that own information assets that can be used by others.

118. In the case of NSOs, many positions within their structure are *de facto* data stewards, fulfilling different tasks and responsibilities necessary to implement data stewardship. However, the role of the system-wide data steward must be clearly designated, as this person (or organization) must 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.

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

120. Data stewards – at the intra-agency or system-wide level – form a new and essential link in the data value chain. This responsibility can be conceived as three main tasks related to collaboration, ethics and sharing (Verhulst, 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 it, so that it may be translated into meaningful impact.

121. Part of the system-wide stewardship role includes activities NSOs are already undertaking as coordinators of the National Statistical Systems. However, the new role proposed is different in that it dives deeper and covers a wider scope:

- Coordination must be done in a much more complex environment, so further detail and specificity are required using new methods and tools that enable the appropriate scaling: new data sources, including privately held data; the conceptual, methodological and technical challenges of integrating different data sources; increased attention to privacy protection, ethics, inclusivity, public acceptance, etc. New methods and approaches are needed in the whole statistical system and NSOs can and should be a leader in developing and promoting them.
- NSOs' coordination function outside the statistical system is extended, possibly even outside the public sector. This can be done by offering advice and guidance, by providing methods and tools that other agencies can use, or by taking responsibility as data stewards of (some of) the public data holdings.

122. This chapter outlines the roles for consideration for both intra-agency and system-wide 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 Generic Statistical Business Process Model (GSBPM) is outlined to show the relationship of data stewardship to the business areas.

5.1 System-wide data stewardship

5.1.1 Sound data stewardship – a public sector goal

Data strategies and data governance

123. Data stewardship roles and functions assumed by NSOs are often formalised and operationalised in data strategies, describing how NSOs manage data and metadata while providing the best standard of statistical information for the public. The existence of such strategies can be a strong communication asset, arguing the cause of data stewards and explaining to stakeholders how the information is processed, protected and what standards are used. Data strategies are in place to guide NSOs data-related actions and processes, as well as to assure 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 other 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.

124. NSOs 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 of growing importance. Therefore, the questions of data and information management, data access, privacy protection and data security are high on the respective agendas of governments and public sector departments. 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.

125. The data stewardship role that NSOs may take in the public sector can vary as much as the contexts they are situated in. From focusing solely within the NSO, to providing guidance and direction on data management for the whole NSS, or even beyond to the entire public sector. In a context where data stewardship is being applied widely, this can occur within the wider data ecosystem (including towards private data holders); concerning public data; or concerning other producers of official statistics who belong to the national statistical system. This broadest role is also relevant to NSOs in

countries where public registers do not exist, since the NSO may still set the direction for government data (e.g. New Zealand). If the focus is relegated to the intra-agency level, data stewardship activities involve a 'core' set of responsibilities for an NSO dealing with sensitive data, and will involve roles that are appropriate for the mandate of an NSO.

126. Figure 4 shows these different responsibilities and their coverage, starting from the outer circle that represents the greatest coverage – data stewardship and management for the whole public sector. Moving inward, each circle describes a successively narrower role, until the inner circle represents the coverage of only one institution - the NSO itself.

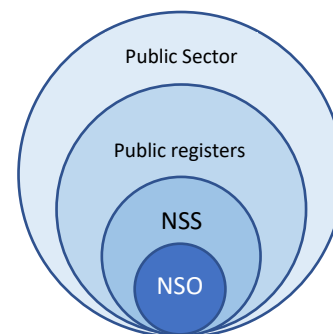


Figure 4. Possible coverage of NSO's data stewardship role

127. In cases where the data stewardship role for the whole government sector lies with another institution and the NSO has this role only for official statistics, the NSO should extensively cooperate with the public sector Data Steward. The NSO should look to play an integral part in elaborating standards, quality frameworks and other enabling instruments, given their expertise and the impact these enablers have on official statistics. Some NSOs may elect to take a less encompassing role as data steward, or some governments may decide to implement different organisational solutions than an NSO-led data stewardship program. It is then vital to enshrine the NSOs rights and responsibilities regarding data stewardship in legislation.

128. The role of the NSO in system-wide data stewardship must also align with the government's public sector data strategies and recommendations on data management. Section 4.1.3 sets out examples of hybrid approaches within this framework of shared responsibilities and government enterprise-wide alignment.

129. 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 data should be collected if there is a public sector organization that has already collected the required data (in accordance with the '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, which is facilitated by this public sector-wide alignment.

130. To achieve interoperability, it is necessary to have common guidelines, structures, metadata definitions, quality frameworks, common identifiers etc. in place. The statistical community addressed this problem decades ago by implementing the use of international concepts, classifications and methods to promote the consistency and efficiency of statistical systems (Fundamental Principles of Official Statistics 9) (United Nations, 2014).

131. The body of knowledge regarding interoperability and data quality that has inevitably developed within NSOs, can be used to promote interoperability and the use of standard classifications, metadata structures and quality frameworks in the broader NSS, among holders of administrative data, and among other institutions and organizations within the public sector. This expertise is augmented by relationships with data holders who have compatible data and skillsets to NSOs (e.g. mapping and environmental agencies, central banks), which can be organized through partnerships (described in Section 5.3.3).

132. 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 the reuse of such data, since the networking, management and legal skills of the data steward will be essential to achieve sustainable

Australia

The [Australian Data Strategy](#) was released in December 2021 and sets the Australian Government's whole-of-government vision for data. The Strategy was delivered jointly by the then Minister for Employment, Workforce, Skills, Small and Family Business and the Minister for Superannuation, Financial Services and the Digital Economy and Women's Economic Security. The Department of Prime Minister and Cabinet (in its role as policy lead for data at the time the Strategy was developed) led its development. The Australian Statistician co-chaired the Working Group which helped shape the strategy and the ABS contributed extensively.

Canada

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 GC, 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. The strategy is currently being updated by the original contributing departments.

United States

The United States' statistical system has been the careful steward of vital data for well over two centuries. More recently, the United States government has expanded the reach of data stewardship beyond just the statistical agencies, as many other Federal agencies amass large collections of data through oversight, enforcement, regulation and other activities. The expansion can be seen in the [United States' Federal Data Strategy](#), with a vision to accelerate the use of data to deliver on mission, serve the public, and steward resources while protecting security, privacy and confidentiality. The Data Strategy is supported by Principles (guidance in areas such as Ethical Governance, Conscious Design and Learning Culture), Practices (40 practices to leverage the value of data) and Annual Action Plan (measurable activities to implement the practices) outlined therein.

access. This role is described in Section 5.2.2.

The NSO's role in Data Stewardship in the public sector

133. The national context is crucial to the role that the NSO adopts in the data ecosystem. There may already be national bodies with responsibilities for aspects of data governance and management. The NSO will be required to collaborate with these bodies to, at minimum, protect the integrity of the NSO and NSS, and ensure close coordination in the production of official statistics. The NSO may also seek to influence the ecosystem by 'marketing' its skillsets and taking advantage of any roles that are appropriate to an NSO in the broader ecosystem (see also Chapter 7). By nature of the expertise, tools and established partnerships discussed thus far, the NSO could successfully take on (some) responsibilities of a data steward for the whole public sector. However, this may not be feasible in all countries. The following section outlines potential focuses or tasks to advance and manage data stewardship at the public sector level, and some examples of the different approaches taken by countries in this regard. Figure 5 outlines different kinds of data stewardship responsibilities, and examples of the countries pursuing those tasks. Further case studies and examples of initiatives and orientations of various countries follow.

134. It should be noted that the UNECE [GAMSO model](#) (2015a), with its focus on strategy and leadership, is ideal for framing the NSO's role in the national context. The GAMSO model defines and outlines the activities that take place within a statistical organization, extending and complementing the Generic Statistical Business Process Model (GSBPM) by adding the activities necessary to support

statistical production (2015a). 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 the most impact in the national context, while 'Corporate Support' activities also include many potential externally-focused services described in this section.

135. As mentioned in the introduction of this section, below are several case studies on data stewardship implementation at the national or federal levels. The descriptions include initiatives, orientations and collaborations of various NSOs in their respective national settings.

136. NSOs are uniquely positioned to evaluate the strengths and weaknesses of the national administrative data system. The 'once only' principle is a concept used by the EU that enables efficiency and reduces response burden on citizens, institutions and companies, who are only required to provide certain information to the authorities and administrations *once*. 'Once only' requires that data gaps are identified, and the value of data collected in each silo is maximized through subsequent integration, sharing and analysis. The NSO's role in setting the direction for the system can range from setting up reference classifications to monitoring of the uptake of classifications and standards affecting statistics (e.g. common identifiers). The most evolved or advanced stewardship role would have NSOs' setting the standards for the 'Integrated Data Infrastructure', which are subsequently implemented and monitored across the public data system.

137. Due to their mandates and missions, privacy preservation and social licence are second nature to NSOs. As public sector data quality improves and demands for analysis within individual ministries increase, these discussions are becoming much more relevant to the broader public sector. These public ministries operate in a multidimensional context, where the social impact of their policies have wide-ranging effects in other domains, which contributes to the increasing need for interoperability and data linkage across these domains. While NSOs can meet some of these demands, the NSO may also need to collaborate with other partners while ensuring that overall trust in the system is maintained.

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 advancing data access and use across the public sector but are currently, relatively immature. 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 Statistics Netherlands/Centraal Bureau voor De Statistiek (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:

1. Data Architecture and Technical (which includes analytics) – CSO leads on this group
2. Data Protection and Ethics
3. Advocacy and Communications
4. Data Sharing - to oversee processes/frameworks/ templates as well as recommending approval of data sharing agreements to the main Board; CSO is not involved in this activity.

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

Finland

Statistics Finland operates in a decentralised statistical 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 the EU's *Statistics Act*, the NSO can give guidance to other offices producing statistics on data production, dissemination and quality. In Finland, this guidance is provided by the Advisory Board of Official statistics where the NSO and ONAs all participate. 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.

Another possibility for the Data Stewardship role within the NSO is to cover other institutions and organisations in the government sector. In Finland there is a Government Information Policy that directs the development of data usability. In the Programme of Prime Minister's Government 2019, the Government will add depth to the management of information policy, prioritising openness of public information. The discussions around the different roles of central government offices in the new information policy are ongoing. Discussions regarding other potential roles for the NSO, for example, with respect to data catalogues for whole government, are also ongoing. In these ways, Statistics Finland are navigating a new leadership role in advancing quality principles of data for whole government.

New Zealand

The Government Chief Data Steward (GCDS) role is held by the Chief Executive (Chief Statistician) of Statistics New Zealand (the NSO). The GCDS role leads by facilitating and enabling a collaborative and unified 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. The GCDS:

- Sets the strategic direction for government's data management
- 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.
- Leads New Zealand's state sector's response to new and emerging data issues

Netherlands

Statistics Netherlands (CBS) envisions a role for itself as a Bureau of Statistical Data Standards statistical data, mainly concerning classifications and methods relating to statistics. CBS also has a role in helping departments with implementing federated models and techniques. CBS can share its knowledge and skills of data processing through the role of a competence centre. CBS is closely involved in the development of the Interdepartmental Data Strategy. The implementation of this strategy may give substance to the potential role of CBS as a data stewardship subject matter expert within the government sector.

Estonia

Estonia has appointed a Chief Data Officer (CDO) for the Estonian government, overseeing data governance and data science in Estonian government. Among other things, the CDO's mission is to lead the development of government services through citizen-centric data governance, open data and artificial intelligence.

Figure 5. Summary of system-wide data stewardship work areas, with references to country examples in parentheses.

<p>Data stewardship as a public sector goal</p> <ul style="list-style-type: none"> • Contribute to developing and promoting data strategies, policy and principles (Australia, Canada, Estonia, Finland, France, Ireland, Lithuania, Mexico, New Zealand, Switzerland, US) <ul style="list-style-type: none"> ○ national /public sector data strategies ○ once-only, open data, FAIR, CARE principles ○ Build trust and confidence in the system
<p>Coordination and partnerships</p> <ul style="list-style-type: none"> • Coordinate standardisation and harmonisation process, support interoperability (Australia, Mexico, Switzerland) • Partnerships to improve capabilities, develop new and improve existing products <ul style="list-style-type: none"> ○ Data sharing arrangements ○ Partnership and licencing agreements ○ Guidance to other agencies (Finland, New Zealand)
<p>Quality assurance and assistance</p> <ul style="list-style-type: none"> • Assess data quality in the National Statistical System and promote and provide expertise in data quality to other public data holders (Australia, Estonia, Finland, Germany, Norway, Switzerland) • Develop and promote quality frameworks (e.g. for administrative registers, privately held data) • Quality certifications and audit
<p>Data access</p> <ul style="list-style-type: none"> • Provide research and information services <ul style="list-style-type: none"> ○ Access to microdata (Australia, Ecuador, Netherlands) ○ Surveys for other government agencies (Ireland) ○ Information services and analytical support to users (incl. policy makers) (Ireland, Netherlands) • Build data platforms and dashboards (for data sharing and dissemination) (Australia, Germany, Ireland) <ul style="list-style-type: none"> ○ Bringing together data from different areas (e.g. Data Centres) (Ireland) ○ Links with geospatial data, geospatial visualization (Croatia) • Access privately held data sources (Australia)
<p>Methods, tools and capabilities</p> <ul style="list-style-type: none"> • Standards and classifications (Australia, Estonia, Ireland, Mexico, Netherlands, New Zealand, Switzerland) • Infrastructure (e.g. for data integration, access to individual /micro data) (Australia) <ul style="list-style-type: none"> ○ Data lake (Lithuania) ○ (Meta)data catalogue (Croatia, Switzerland) ○ Common identifiers (Denmark, Ireland, Sweden) • Data integration • Data protection • Statistical methods, data science and machine learning methods (Australia, Netherlands) • Common approach to data handling (Australia) • Helping other agencies to build their skills, processes, tools and services (Ireland) • Improving data literacy (Australia, Estonia)

5.1.2 Support to other public sector data providers

Coordination and partnerships

138. In many cases, data providers are also data users, since public sector agencies, such as government ministries or central banks usually have policy functions as well as operational ones. At a minimum, information on relevant outputs should be viewed as part of the overall relationship with such agencies. A more proactive approach would involve the analysis of these agencies in the liaison groups governing data flows, where they could have an input into statistical work based on the agencies' data sources.

139. 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 effectively direct external focus. The NSO should also demonstrate its leadership role in data analysis in these outputs.

140. 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.

Australia

The Australian Bureau of Statistics (ABS) leverages partnerships with other data providers to develop new products or enhance existing ones. ABS accesses over 100 datasets for statistical and research purposes, which are predominantly public sector data assets including birth and death registrations, taxation and welfare data. Data sharing arrangements are an essential part of data partnerships, and the ABS makes use of a range of data sharing arrangements – via legislation, memorandums of understanding and licencing arrangements. For many data sources, including taxation data, legislation enables the ABS to receive data for the purpose of the Census and Statistics Act 1905 (the Act), the main legislation for collecting, compiling, analysing and disseminating statistical data. The Act governs all ABS statistical releases and specifies confidentiality requirements. In addition to legislation, memoranda of understanding are also put in place specifying terms and conditions, including review points. Licencing arrangements are used for private sector data sources.

France

The National Institute of Statistics and Economic Studies (Insee) collaborates regularly with the Interministerial Directorate for Digital Technology (DINUM), which is responsible for the state digital transformation for the benefit of citizens and state agents. Digital transformation is a government top priority and raises the important issue of the need for digital competencies. In February 2021, Insee and the Dinum were asked to lead a joint project to define a typology of needs in the different ministries, to clarify the possible role of the official statistical service, and to evaluate the recruitment processes.

Mexico

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. The Institute has the authority to collaborate with the public and private sectors, establishing agreements with companies and non-governmental organizations to obtain data that can be reused in the generation of statistical information. It is the institution that, by law, assumes the role of NSS coordinator and 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.

Data access

141. Many NSOs are considering developments to data and information services as they decide whether to become 'data and statistics' offices. Differential privacy is key to these services. NSOs provide open data and through strategic partnerships, can also facilitate other public bodies to meet their obligations for open data. Researcher access is one of the key areas for development, including the new role many NSOs have taken on of late, in facilitating access to data for COVID researchers. It is also a focus in EU legislation (see Appendix 2), 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 NSOs also provide anonymised data services to encourage data skills among inexperienced researchers.

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

Australia

ABS was instrumental in the establishment of guidelines for [Data Integration Projects involving Commonwealth Data for Statistical and Research Purposes](#) in which quality assessment and data assurance are fundamental. Many of the guidelines have now been codified in legislation via the *Data Availability and Transparency Act* (2022). ABS' statistical and data integration capability is complemented by a data access service known as the [ABS DataLab](#). The DataLab allows sophisticated analysis of detailed microdata in a secure controlled environment. The DataLab currently services approximately 400 active projects and 4,000 registered users across government but also Australia's research sector, which is a strong support of this service. Use of the ABS DataLab is growing at about 30 per cent per annum.

Croatia

Croatian Bureau of Statistics' GEOSTAT Portal contains the [Metadata Catalogue](#), which enables searching spatial data sets and services. One of the goals of the Catalogue, is to facilitate searching, analysing and sharing spatial data and to increase interoperability between the provider and users of spatial data and its services, all with the aim of meeting the requirements of the INSPIRE Directive (sets the minimum conditions for interoperable sharing and exchange of spatial data across Europe as part of a larger European Interoperability Framework and the e-Government Action Plan).

Germany

In addition to being a data producer, the Federal Statistical Office of Germany (Destatis) has broadened its role to being a data service provider (among other roles, see Germany's entry in Annex 4). In its role as data service provider, Destatis has established the data platform [Dashboard Germany](#). provides up-to date information on the economy, finance, health and mobility on this interactive portal. The Dashboard, which was developed on behalf of the Federal Ministry of the Interior, Building and Community and the Federal Ministry for Economic Affairs and Energy, already offered more than 100 indicators at its launch in 2020.

Quality Assurance (QA) assistance – Methods, tools and capabilities

143. To lead and govern the quality of the data of public organisations, the data steward needs to have several different kinds of abilities and capabilities. First, the NSO needs to have a comprehensive understanding of the public data space, the data infrastructure and data flows that exist between different public organisations – including private data (depending on the coverage of its role, see Figure

3). In other words, conceptual and architectural understanding of the National Data Infrastructure is a necessity to coordinate the national data quality improvement actions within the public sector.

144. Coordination, communication and networking skills are all needed when creating and maintaining national data quality solutions. This is long term work and requires dedicated human resources – resources that are familiar with methodologies as well as those with deep knowledge on the data in question.

145. It is also important to note that the level of data quality required depends on how the data is used. There might also be different kinds of data usages for the same data, leading to different quality assumptions and needs concerning that data. These must be discussed in detail and with the appropriate subject matter authorities. Specific topics the NSO could consider as part of the stewardship role include:

- 1) **Metadata.** This is crucial for evaluating data quality and needs to be maintained for administrative data within NSOs, as the data were not originally collected for statistical purposes and each source will have its own dimensions of quality depending on mandatory/legislative fields, coverage, etc. The NSO may also consider making the metadata publicly available to researchers, or even more broadly. As a further step, the NSO may define common metadata standards, or collaborate with the body developing the standards to create cohesion and consistency.
- 2) **Development and promotion of data standards** in the community of data owners, such as input and output **classifications**. 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. This 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 of NSOs in relation to **privately held data** to date has been access, **quality frameworks** for these data also need to be considered. The quality model that applies to public sector data may not be relevant to private sector data. ‘Volume, velocity and variety’, the defining characteristics of Big Data, will result in different but very informative, statistical products to those derived from traditional administrative sources with high and predictable population coverage levels.

Netherlands

CBS provides remote access to microdata as its data service. In the coming years CBS intends to establish even more links with government and businesses regarding AI applications. What and how this will be done is currently being worked out in a policy framework. CBS also works together with several municipalities in the form of an Urban Data Centre (UDC). The aim of the UDC is to stimulate data-driven work where CBS’s 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. Currently there are about 14 UDCs in the Netherlands. Experience with the UDCs have led CBS to continue working with the current network of UDCs, but for new social issues at the municipal level we are looking for generic solutions / instruments for all municipalities in the Netherlands. In doing so we work together with the Association of Netherlands Municipalities.

Canada

In 2021, the Government of Canada (GC) established an interdepartmental Working Group on Data and Information to advance data stewardship, leadership and governance in enterprise (GC-wide) data and information. As a part of this work, a GC-wide data and information governance framework (with clear accountabilities and principles to guide decision-making) was developed. The framework is intended to support the realization of desired outcomes such as enhanced operations; improved service, program, and policy design and delivery; and greater public value. The framework identifies key foundational enablers (People, Rules and Guidance, and Processes and Tools) that are needed to establish a holistic approach to governing data and information in the GC. These enablers are supported by a set of guiding principles and are intended to enable capabilities such as security; integration and interoperability; discovery and access; lifecycle management; data/information architecture; privacy protection; research and analysis; quality management; change management; and communications. The framework is supported by definitions which provide a common vocabulary and understanding of data and information governance in the GC.

Finland

The Government of Finland recently concluded a development programme on data usage and data dissemination (2020-2022). The main goals were 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 was divided into four subprojects: data strategy, *data quality*, opening data, and technical and semantic interoperability. Statistics Finland oversaw the data quality framework project.

The aim of the *data quality framework project* was to increase the usability and uniformity of data and extend the use of data for decision-making in society and by enterprises. The project studied the current state of data quality management and on this basis, consultations were conducted, then preliminary quality criteria were formed and published, piloted, and subsequently launched in 2022. The quality framework for the public sector included quality criteria with and metrics for public along with a couple of core models supporting the implementation at national, as well as organisational level. Since the framework's release, concrete data quality evaluation examples utilising these metrics have been published.

5.2 Roles associated with intra-agency data stewardship

5.2.1 Generic roles required to implement data stewardship in NSOs

146. Six generic key roles can be identified to fulfil all the requirements of sound data governance and stewardship within an organization, in addition to the traditional statistician/data analyst roles:

- 1) Data Steward
- 2) Data Acquisition Lead
- 3) Data Engineer
- 4) Statistical Product Owner
- 5) Domain Data Manager
- 6) Corporate Services Supporting Data Stewardship.

147. The Data Steward works closely with their colleagues in the other roles. Individuals can also hold multiple roles concurrently. The concepts and associated actions must be clear to everyone involved in working with data. This is facilitated by the Data Steward, who will be involved in the work of all other roles described here, like the hub at the centre of a wheel. NSOs may assign and distribute these roles in ways appropriate to their context, the aim in this section is to give a generic description.

148. Knowledge and expertise on data stewardship are often fragmented at NSOs. Ideally, the primary data stewardship function exists as a centrally organised role, with a mandate (responsibilities and skills) as steward of the data (including metadata). The Data Steward would not be guarding the content, which is done by the statistical product owner, but instead guarding the *processes* surrounding the data and metadata. Depending on the size of the organization, decentralised stewards who are close to the line employees may also be required in addition to a central data steward. 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, promotes it and gives advice (both solicited and unsolicited).

149. A formal approach to data management is especially important when data is collected outside of the NSO. In this case, metadata needs to be much more comprehensive to flag issues likely to affect quality in the collecting organisations. These include factors such as mandatory or statutory variables and collection channels. This role is critical as NSOs seek to move away from primary data collection towards using more secondary data, and as statistical product owners need additional safeguards to allow use of data sources that are not directly under their own control. As NSOs seek to reuse privately held data, 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 within an NSO

150. The role of the Data Steward, as it is outlined here, describes the *minimum* activities and responsibilities for the role as it exists in an organization that acquires and processes sensitive data. The NSO context is further explained in upcoming sections of the report by relating the role of the Data Steward to other roles in the organization, based on UNECE GSBPM and RASCI (Responsible, Accountable, Supporting, Consulted, and Informed) matrices. What constitutes the 'maximum list' of activities and responsibilities will depend on the degree of external engagement of the NSO.

151. At minimum, an NSO's data steward would be responsible for:

- Data description (e.g. data catalogue, data dictionary)
- Metadata quality (the responsibility of the data lies with the data holder)
- Data life cycle management
- Data ethics (the Data Steward is the connection between the Data Management officer/Data Protection Officer and the Ethics Committee)
- Data security, protection and confidentiality
- Data audits (monitoring the use of data).

152. The responsibilities of a data steward within an NSO may be categorized by networking activities (both internal and external), technical responsibilities related to data management, and ethical and legal responsibilities:

External/internal networking:

- Supporting acquisition of new data and using new data sources
- Liaising with external providers for metadata and data life cycle
- Liaising internally for quality
- Making internal and external data available based on FAIR principles.

Data management/technical

- Maintaining meta/classifications codes
- Supervising data life cycle management
- Monitoring 'once only' principle
- Overseeing quality and security processes.

Ethics and legal

- Establishing a link to ethics boards and legal services
- Ensuring compliance of Data Governance Model
- Developing rules for acquisition of data
- Enabling transparent processes.

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.

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, although some administrative and other secondary data remains in managed 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) has a strong focus on skills and roles particularly in developing the 'data engineering' function.

5.2.3 Other roles related to Data Stewardship

153. **Data acquisition** is the exploration and acquisition of new data sources (from public or private sources) for use in the statistical process, both for improving existing statistics and developing new statistics.

154. The person(s) responsible for data acquisition coordinates, organises and facilitates the data acquisition activities for a specific (or group of) new data sources. It may not be one centralised function, instead directly carried out by different Data Acquisition Leads (where there is a concrete need for a *specific* new data source that triggers this process). Typically, Data Acquisition Leads come from statistical subject domain departments or from departments tasked with research and innovation. Responsibility for the acquisition of a specific (or group of) new source(s) will therefore lie with those teams.

155. 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, responsible for the 'holding' and use of such data as well as the networking, data management and legal or ethical activities required to negotiate ongoing access.

156. Responsibilities of a Data Acquisition Lead include:

- Defining the specific data needs, together with internal and/or external users
- Providing leadership for the acquisition team for the specific data sources
- Searching 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 the NSO and externally (such as data owners and collaboration partners) with the goal of acquiring the targeted data source.

Netherlands

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.

Ireland

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.

157. **Data engineering**, while similar to data stewardship and acquisition, is a new speciality in NSOs which brings together aspects of technology and statistical skillsets, but with a stronger emphasis on technology or analytics training. Technology experts in NSOs need to have a strong awareness of business requirements if they are to support business goals effectively and with agility. 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, etc.)
- Provision of data extracts to data analysts in line with research goals.

158. Data engineers can be organizationally located with the Data Steward or in business areas, and ideally are separate from the software engineering and other technology functions within the NSO. The **Statistical Product Owner** (SPO) is responsible 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.

159. 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-to-1 and 1-to-many relationships between DDMs and SPOs in the organisation. DDM managers are responsible 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.

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

161. 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. A summary overview of data stewardship roles that NSO can assume in partnership with different stakeholder groups is visualised on Figure 5 (see also Figure 4 above).

5.3 Human resource requirements for the data steward role

162. The potential scope of the external and internal data stewardship roles described above requires a new approach to human resource practices in NSOs. As it stands, statisticians and data engineers increasingly come from analytics backgrounds. While the complementarity of the analytics and statistician skillsets are often neglected in the public sector, there are indeed complementary areas of expertise. Here, NSOs have an opportunity to engage in partnerships, rather than becoming involved in recruitment, training and secondments. These issues are described in this Section.

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 development of common data literacy and skills across the public service. This will need to be balanced against CSO's own staff requirements, geographical considerations and the higher turnover rate 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

163. The skills required to establish each NSO's public sector role, depends on the level of engagement described in Section 5.1 and in the various national examples.

164. 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 must convince others on the long-term direction, strategy and added value 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 and to communicate information on the progress made in the task.

165. 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). Internally, operational level skills are also required, from data management to data descriptions and data bases. In Section 5.2, these roles are described in more detail with national examples of roles and required skills. The NSO leadership will be trying to influence a range of larger and likely better funded organisations, as well as smaller, less organised data holders without any formally defined role, which requires very strong networking capability.

Technical level skills

166. A range of skills are needed to fulfil all the requirements of good data governance, from legal or ethics specialists, to data protection experts, to the statistician/data analyst level. For many statisticians or analysts, 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 a managerial role, but at the working level, rather than the managerial.

5.3.1 Secondments

167. 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 central authority. Some public sector agencies may not have the skillsets 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 skillsets to ensure that standards are adopted. There are many resultant benefits for NSOs related to data quality.

Ireland

The CSO seconds 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 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.

New Zealand

As New Zealand's NSO, Tatauranga Aotearoa Stats NZ is uniquely positioned to support the data literacy development of the entire public service, as well as supporting data-driven decision making. Stats NZ's structure, culture and systems are designed with collaboration and customers in mind, so statistician secondments and rotations to other departments are common.

5.3.2 Partnerships

168. The skills necessary to collaborate or cooperate, create partnerships, and communicate with those partners are crucial in the data stewardship role, especially if the DS role extends outside of the NSO. This requires effective negotiation and communication skills and even some legal expertise, if the partnership is documented in a Memorandum of Understanding or other agreement format. 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.

Canada

In 2022, the federal public service of Canada renewed its commitment to improving data management across the government by requesting an update to the data strategy spearheaded by Privy Council's Office (PCO), Treasury Board of Canada Secretariat (TBS) and Statistics Canada (StatCan). The different expertise and capabilities of these partners inform the responsible use of data and the continued protection of individual and organizational information and privacy. As the Government of Canada's trusted data steward, StatCan is mobilizing its data expertise by leveraging its existing data standards and data management capabilities in a culture of coordination within the national statistical system. In terms of the overall GC direction, the PCO, through the auspices of the Clerk of the Privy Council, serves as the head of the civil service of Canada and supports the Prime Minister and Cabinet by establishing the priorities and goals of the federal public service. In their role as the employer of the federal public service, the TBS is the central agency with the authority to oversee and manage the expenditures and operations of the federal public service by establishing policies, regulations and directives for organizations to follow. This partnership brings together data stewardship expertise, ministerial support, accountability, operational enablement and regulatory/policy coverage to the federal data strategy, and to the data stewardship advancement happening in Canada.

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). Until March 2020 our work was focused on collaborations involving our complementary skillsets, such as SDGs, Census mapping and near-Earth observation data. This 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 mid-March 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 be facilitated by NSO stewardship

169. It should be recognised that other agencies have a different set of requirements within their mandate for administrative data than NSOs do. Specifically, they usually collect their data for operational measurement or policy/programme development and evaluation, which can raise challenges in terms of data sharing and the compatibility with NSO's core values of confidentiality and independence. This should be considered when analysing how NSO's data stewardship role could potentially interact with their non-statistical stewardship requirements for effective data access and quality.

170. 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' as NSOs. While NSOs will not be directly involved in this analysis, if they do not provide support, it will likely be outsourced or done sub-optimally, with obvious consequences for data quality.

171. Classifications and data services (mainly visualisations) are crucial for effective Business Intelligence (BI) implementation in agencies. The audience for the BI reports is usually 'in house' and self-service, so there is less interpretation needed of the data compared with normal statistical production. The metrics provided are based on combinations of events which are supposed to represent progress – or lack thereof – in operational performance. Clearly, processes and processing times for social welfare offices, for example, compared to 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 occurs primarily through ad-hoc 'pathfinders' but can be facilitated through the data services and secondment programmes described. Also, any improvements to data for statistical purposes, such as identifiers and classifications, are strongly supportive for BI analysis.

172. Good relations between statisticians and their information technology (IT) colleagues are second nature in NSOs, as they are crucial to effective analysis. The best approach to this interface is a collaboration as described in the 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 statistical agencies do effectively, there is also an important facilitation role that NSOs can play in partnership with public sector agencies, based on the data services, training and secondments described earlier.

173. **Operational analytics** involve a range of activities in PSBs, such as operations research, fraud measurement/detection, or segmentation of clients requiring additional support and financial management (often based on similar regression techniques to those used for programme evaluation). While these activities are clearly not supported by the NSO's mandate, there are some indirect supports that can be provided via the aforementioned data services and secondment programmes. Also, any improvements to data for statistical purposes, such as identifiers and classifications, are also strongly enabling for operational analysis.

174. Modern programme evaluation attempts to control for background factors to determine whether participation in a programme is beneficial or cost-effective for clients. While there is a strong scientific basis to this analysis, it also requires some subjective decisions to obtain a clear signal about a programme's value. As with operational analytics, this is often not necessarily consistent with the NSO's mandate, but in addition to the supportive services identified for analytics, NSOs can also provide data hub/research access services and partnership options for programme evaluation, since the results involved are generally aggregates.

6 Data stewardship maturity assessment model

175. This chapter considers the use of a data maturity model to guide and assess data stewardship strategies and or programmes. It first provides a brief description of maturity models generally, and the benefits of their use. It then describes the application of a maturity model for data stewardship specifically, including the dimensions or principles of stewardship that might be reflected in such a model, and how.

176. The ways that insights derived from a maturity model assessment are reported and leveraged will contribute to its success, so options for the delivery of a stewardship maturity assessment will also be presented. Considering the potential of a data stewardship maturity model for an NSO, especially one acting in a lead steward capacity, this chapter concludes by identifying ways its use can help support improved stewardship across the data ecosystem.

6.1 Maturity models

177. The concept of maturity assessment has used by organisations since the 1980s as an intuitive way of understanding the extent to which their business practices reflect, or might develop to, desired or planned levels (Team OSTHUS, 2022; Proença and Borbinha, 2018).

178. When employed for business assessment or as a measurement tool, a maturity model is typically used to establish a baseline for practices, and based on that, offer a way to help identify, describe, implement and monitor ongoing improvements in those practices. In that regard, maturity models are employed for three purposes:

- descriptive – for describing current state
- prescriptive – for distinguishing desired maturity levels and improvement measures
- comparative – for benchmarking against others (Röglinger and Pöppelbuß, 2011).

179. While they do vary to some extent, a review of popular maturity models suggests that most are based on two consistent principles: 1) maturity represents a cumulative and progressive measure, and 2) that this measure can be expressed or scored across discrete stages or levels. The levels describe increasing maturity from non-existent or limited, to fully optimised, where the value of the assessed practice is fully realised and is operating strategically. The maturity model therefore offers an effective way to help guide the evolution of practice along a journey, to an increased maturity or improved state (Peng 2018).

180. Though maturity models are often constructed similarly and based on common concepts, it is their application in particular contexts, for a particular type of practice, or in service of an identified business need, that gives them value (De Bruin et al., 2005).

181. The promotion of data stewardship by a NSO, for instance, is one context where the use of a maturity model can help facilitate positive outcomes. In fact, maturity models have been designed with just that purpose in mind, and which offer a means of assessing the current state of data stewardship practices and identifying areas for improvement (Peng 2018).

New Zealand’s Data Maturity Assessment Toolkit

The New Zealand Government Chief Data Steward (GCDS), through Stats NZ, is currently testing the deployment of a Data Maturity Assessment Toolkit. This Toolkit will provide a set of resources to produce a comprehensive assessment of data maturity, including the theme of data stewardship. To address gaps associated with existing data maturity assessments, the Toolkit incorporates various innovations, including:

- an interview-based assessment collection
- business imperatives to serve as assessment narratives
- a values-based data practice maturity measurement
- an intuitive and qualitative results scoring standard to present to senior decision-makers.

The Data Maturity Assessment Toolkit is planned for release across the New Zealand government data system in late 2023. Once in use, it is anticipated that it will provide a valuable resource for government agencies, helping to inform decisions and target investment in data. The Toolkit will also provide the GCDS with a more accurate view of the current state of the New Zealand government data system, which will support the mandate of that leadership role to implement policy and direction, provide advice and guidance, and help lift levels of data capability.

6.2 Benefits of a maturity model for data stewardship

182. The idea of stewardship as it applies to data is still a relatively new concept, and it may not be the case that it is well understood or widely adopted throughout the ecosystem. The use of a maturity model can help address this knowledge gap by providing a practice-based view of stewardship.

183. A truly comprehensive and detailed view of the evolution or maturity of a data stewardship programme is only possible through the use of an assessment model. This is because the assessment model works to enable the identification and prioritisation of strengths and improvement opportunities for a particular practice (Proença and Borbinha, 2018). This understanding can also support a comparison of practice to an agreed standard, or to the maturity of a similar practice in other organisations. In this way, the maturity model results can be used to inform future activities things like planning or investment, in addition to the evaluation of past or present activities.

184. At the outset, the use of a maturity model can help an organisation establish where it is positioned along its data stewardship journey; representing an invaluable insight for planning, and for implementing changes and improvements to data practice.

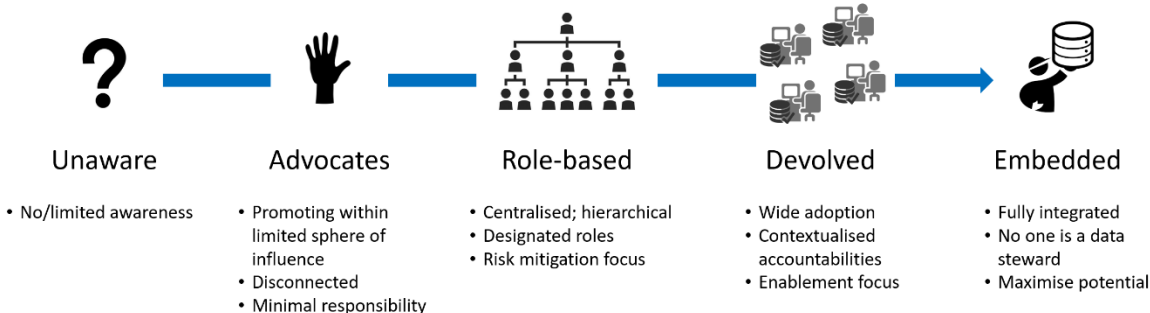


Figure 8. A maturity-based view of the data stewardship journey

185. Figure 8 provides a view of common stages, reflecting the maturity levels often used by data maturity models, experienced along a journey towards data stewardship.

186. At the earliest stage of the journey, there is likely to be no, or very limited, awareness of the concept of data stewardship and how it might benefit the organisation. In an insular environment –

one without the benefit of shared discourse, communities of practices or other partnerships outside the organization – it's possible for an organisation to remain in this state. At least, until some catalyst necessitates research and re-evaluation.

187. If the concept of data stewardship and its benefits become known, it is often only understood and promoted at first by a limited set of individual advocates. These proponents will consist of those who brought the idea with them into the organisation those who picked it up through external engagement, or recent converts of the former two. While these early adopters may be effective promoters, the extent of their advocacy at this stage is likely to be limited, perhaps only extending to their immediate team or business unit. This might reflect the nature of their position in the organisation, or the inherent challenges associated with explaining a concept like stewardship.

188. However, if these advocates are successful in spreading the word and connecting with one another, they can create a momentum that carries the idea more widely across the organisation. The communication of successful use cases involving data stewardship for instance, can serve as an effective means of establishing the idea more widely, including at some point landing it with executive leadership.

189. Once the idea of data stewardship and its value proposition reaches a 'critical mass' or establishes itself with those able to influence or steer the culture and strategic direction of the organisation, there is often a move to standardise and centralise it as an enterprise function. This represents the shift of a novel idea into familiar structures, which not only makes adoption more efficient, but also increases visibility within the organisational culture. In many cases, this transition is motivated by risk management, with centralised functions easier to monitor and control as they also contribute to a consistent and conformed approach for data across the enterprise.

190. In the case of data stewardship, this move to a centralised and often hierarchical framework takes the form of a set of designated roles, including data stewards and custodians, which are responsible for delivering and monitoring good stewardship practice throughout the organisation.

191. The 'centralised and role-based' stage in the stewardship maturity journey can deliver a significant and measurable increase in benefit realisation compared to the previous stage. Depending on the context, it may also be seen as the culmination, representing the final stage in a maturity journey.

192. However, some might find that, in the face of changing operating environments, a system of centralised data stewards delivers decreasing value over time, leading to the shift to a 'devolved' approach to stewardship. At this stage, "everyone is a data steward," and good data practice is built into all roles, thereby eliminating the need for designated stewards.

193. This evolution has the benefit of wider adoption of good stewardship practice, greater levels of data accountability, and the customisation of stewardship for specific operating contexts, where it can deliver to its full potential. It can also reflect a shift away from a strictly risk-based approach to data stewardship, to one where it is considered an enabler for the organisation.

194. For some organisations, the devolved stage can further transition over time into a fully embedded maturity stage. Here, data stewardship is so endemic to all data practice, that it's not acknowledged as a distinct characteristic. At this point, "no one is a data steward," since the concept of data stewardship and the manifestation of any role that handles or is responsible for data, are one in the same.

195. There are benefits to be realised through the adoption of data stewardship at all stages of this maturity journey, but it is the organisation's awareness and understanding of its current positioning on that journey that will offer significant value, as it helps align data investment decisions to the most appropriate state for the organisation. With that knowledge, interventions affecting data practice are most likely to deliver on their objectives and generate positive impacts.

6.3 Using a maturity model to assess data stewardship

196. To implement a data stewardship maturity model, elements of commonly employed data maturity assessments can be leveraged. The use of familiar concepts offers the advantage of a set of well-tested components with which to generate a maturity-based perspective, and from that, a clear picture of what acceptable data stewardship looks like in a particular context.

197. Popular maturity models include: the Carnegie Mellon University Capability Maturity Model Integration (CMMI) (ISACA, 2014), the Enterprise Data Management Council Data Management Capability Assessment Model (EDMC, 2015), The Open Group Service Integration Maturity Model (Open Group, 2011), Gartner's Enterprise Information Management Maturity Model (Gartner, 2014; Taylor, n.d.), the Open Data Institute Open Data Maturity Model (ODI, 2015), and the IBM Data Governance Maturity Model (Firican, n.d.; Taylor, n.d.).

198. When considering the components that are common to these models, and which might contribute to a successful data stewardship maturity model, two that stand out are the selection of the maturity dimensions to be assessed, and the approach to reporting the results of the assessment. Other elements, like the collection methodology and the measurement framework, tend to be consistent across models and therefore less amenable to the manipulation that would make them useful for a specific application like stewardship maturity.

6.3.1 Maturity dimensions for data stewardship

199. The list of dimensions or characteristics of sound data stewardship that are assessed provide a foundational structure for a maturity model. Once in place, they will have a direct influence on the maturity measurement, as well as the form and content of the assessment results, and the nature of any recommended actions included in those results.

200. While the choice of dimensions should reflect the accepted understanding of data stewardship best practices, it should also be shaped by considerations important to the organisation employing the maturity model. In this way, the results are more likely to be defensible and amenable to driving positive change in the specific operating environment where the model is used.

201. An effective way of considering the dimensions of a maturity model, including one developed for data stewardship, is to characterise them as *capabilities*. A capability can be defined simply as the ability that an organisation possesses to do what it needs or wants to do (Open Group, 2018). These typically represent a combination of people, methods, processes, information and technology.

202. The advantage of adopting capabilities as the basis for maturity model dimensions is that it provides access to a well-documented collection of accepted concepts and standardised definitions, which can be used to better encapsulate and demonstrate what constitutes good data stewardship. Capabilities allow NSOs and other agencies to structure enterprise action plans around clear competencies in a coherent way, so their use in the maturity model also supports business-based outcomes for the assessment (Open Group, 2011; 2018). This helps make them more relatable to decision-makers and other stakeholders who might benefit from the use of a maturity model, while it also supports results that are more readily actionable.

203. The Government Chief Data Steward (GCDS) in New Zealand has published a *Data stewardship framework and toolkit* (Stats NZ, n.d.), which includes the following seven key elements for effective data stewardship.

- **Strategy and culture.** Including a strategy that provides a shared vision and clear direction, and a data culture that enables strategy implementation and sustains good data stewardship practice.
- **Rules and settings.** The presence of legislation, policies, principles and sanctions providing boundaries and guiding how data should operate.
- **Roles, responsibilities and accountabilities.** The governance structures, role definitions and expectations, and leadership.
- **Data capability and quality.** The tools, processes, designs, metadata structures and platforms for managing, storing, describing and sharing data.
- **People capability and literacy.** The skills, knowledge and services for accessing, managing, analysing and communicating data and insights.
- **Influence and advocacy.** Including effective relationships and networks to endorse, promote and support good data practice.
- **Monitoring and assurance.** Including assessing environmental trends and developments, measuring stewardship performance and adapting the stewardship toolkit to respond to changing circumstances or new information.

204. These seven elements provide a view of what might constitute an effective data stewardship role. Furthermore, if aligned to business imperatives, they can be treated as a set of capabilities or competencies needed to realise data stewardship. Therefore, they constitute a sound basis from which an organisation might structure a data stewardship maturity model, while offering potential options for the assessment dimensions or principles within that model.

205. The New Zealand data stewardship framework has also been developed to characterise effective data stewardship at various scales, from the (national) data ecosystem, to an individual organisation, to a particular dataset or individual within an organisation (Stats NZ, n.d.). As such, it might also be amenable as a starting point for a NSO operating as a lead data steward.

6.3.2 Reporting data stewardship model insights

206. Once a data stewardship maturity model has been deployed, the ways that the insights gained from its use are organised and presented are no less vital to its success. To be meaningful, the insights should be delivered in a way that characterises any changes needed to improve maturity, target relevant audiences, and are easily understood by non-data practitioners, especially decision-makers. These considerations will all ensure that model outputs are genuinely actionable and therefore, more likely to generate positive impacts.

207. The use of a data stewardship maturity model and the subsequent analysis of the organisation's current state acts as a baseline and supports further interventions required to move the maturity of the practices associated with a given dimension to a higher level. For example, describing what is needed to improve data literacy across the work force, or the requirements to establish more effective data governance structures. When identifying potential improvements to data practice, they should reflect the following four components:

- what the change involves
- the potential benefits that arise from making the change
- the resources required to enact the change
- any risks and mitigations associated with the change.

208. It is also advisable to present different visualisations of the assessment model results, so that they are easily understood by stakeholders who hold different perspectives on data stewardship. For instance, one version could present results targeted to a NSO acting as a lead data steward at the national level, providing an ecosystem-based perspective on maturity. Another version might provide targeted results to an organisation's executive leadership, reflecting a deliberately strategic view of maturity results that would be useful for their needs. Similarly, a version might be developed to target data practitioners, focussing on highly practical aspects of data stewardship that would resonate with that audience.

209. Regardless of the iterations that are developed, at least one visualization of the results of the maturity model should speak to decision-makers. This would avoid highly technical terminology, and instead employ direct, succinct, business-centric language that is familiar to those who administer or influence strategic decisions and investments.

210. Since data stewardship enables the successful delivery of business imperatives, these senior leadership-based results might be presented to show the extent to which current data stewardship practice, across each assessed dimension, is helping the organisation meet (or not meet) its obligations, expectations and responsibilities. In this way, a clear connection is made between sound data stewardship and the outcomes with which senior leaders are most concerned.

211. It is also advisable that the version targeted to senior leadership reflect a predominantly qualitative rather than quantitative view of maturity measurement. Maturity, as an assessment approach, is inherently subjective. If it is presented as an exclusively quantitative measure, using precise numeric scores, then the extent to which it can provide a realistic view of the current state of data practice, and help highlight the nature and quality of the changes that result from its use, might be diminished.

6.4 Supporting improved stewardship across the data ecosystem

212. As noted earlier, the purpose of a maturity model is two-fold: to capture a current state or a baseline measure of practice, and to identify ways in which the associated level of maturity can be improved if desired; it therefore supports both evaluation and aspiration. In the case of a data stewardship maturity model, the goal is to use the results to improve the level of data stewardship across an organisation that participates in a wider data ecosystem. In this way, the model can serve as a valuable tool for a NSO operating as a national data stewardship lead.

213. The ways in which a maturity model can contribute to improved data stewardship are numerous, and the design, administration and results reporting associated with that model can be tailored to meet the needs of a particular operating environment. This is true at the level of the organisation, but also at the national level.

6.4.1 Data stewardship maturity model outcomes

214. If used effectively, a data stewardship maturity model contributes to an enabling environment for data across the ecosystem, where stewardship is not only widely understood but is actively engaged to inform and guide practice. This will help with the realisation of consistent fit-for-purpose data and contribute to a range of positive outcomes, including:

- A shared understanding of data stewardship
 - increasing levels of trust and confidence
 - lowering levels of cross-organisation friction and increasing productive engagement
 - reducing risk associated with inadvertent access to and use of data

- An objective measurement of current data stewardship practice
 - an understanding of where data-related risk and opportunity resides
 - a view of where data investment is beneficial
 - a view of where good data practice resides and is available to leverage
 - an understanding of what effective change and investment looks like
- An increase in data value
 - underpinning better relationships between organisations
 - assisting with improved service provision
 - accelerating and broadening innovation opportunities
 - increasing transparency and levels of trust and confidence.

7 How to communicate data stewardship?

215. 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 strategies. The chapter includes some examples and success stories regarding the communication of a holistic approach to public data governance and data management across countries.

7.1 A complex concept – how to communicate it?

216. Over the last decades, the environment where official statistics has been operating is changing rapidly. It can be described as heterogeneous, because of the proliferation of new data sources and data stakeholders. Data ecosystems represent interacting data producers and users whose common characteristics can be described as “difference and diversity” in all data-related activities: production, exchange and use.

217. Users not only search for reliable information, but they are increasingly looking for a fast solution. They are frequently not interested in who is at the beginning of the information chain, i.e. who stands behind data. Instead, they expect to have quick access to extensive and comprehensive information, in an easily approachable format, accompanied with hints on how to interpret it (so called “data stories”). Users want to be assisted, treated with attention and care. They also want to be guided through the maze of data, standards, research and definitions. Data stewards are expected to help with all of this.

218. Thanks to their long-standing practice with data and statistics, NSOs are well placed to take a lead in raising the awareness of data stewardship concepts and requirements among agencies in the data ecosystem as well as various users of data. Communicating complex concepts to specialists in the field and the general public requires different approaches. 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 expertise in transferring this knowledge to other players in the data ecosystem. This experience must be communicated with clarity and using accepted best-practices in communication.

219. When talking about data stewardship and communicating its value, the first aspect that comes to mind is purely semantic: what does it mean? To this end, the current report provides a common definition (see Chapter 3). However, the reality of how the stewardship role is implemented in countries looks very different due to the national contexts (see Chapter 5). An added difficulty is the need to translate the term into national languages. In some languages, the same term is used for the core concepts, but this does not convey the important distinction between ‘data stewardship’, ‘data governance’ and ‘data management’ (see Section 3.2).

220. The communication of data stewardship should be as simple as possible, starting with essential information only, and building in more detailed and theoretical information over time. A practical solution to communicating data stewardship effectively can be to focus on what the purpose of data stewardship is, who performs the related activities and what these are. The communication can also address new products and services that data stewardship engenders and inspires.

221. Failing to communicate the role properly brings about considerable risk. For example, a lack of specificity regarding 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. A 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 on the role of the data steward(s).

222. The 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 – often the ones assuming the stewardship responsibilities.

7.2 Possible messages for communicating data stewardship

223. The NSOs have rightfully recognised the **importance of the communication of their products and roles in the society** in which they operate.

224. Data stewardship-related communication may focus on the **roles which the data steward can or should play in the data ecosystems**, related to data governance and coordination, acting as data curator,⁹ and assisting and facilitating data use. Communication **can also focus on challenges** that data stewardship can help to address in national data systems. An example 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 (2022).

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

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

226. At the same time, it has to be communicated that data governance and data stewardship functions must 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.

227. The “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 collaboration can be strengthened by going beyond the national framework and establishing a network or community of data stewards.

228. **Target audience** is a crucial aspect in communication. This is particularly true 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, the well-known and widely utilised concept of ‘**persona**’ might be helpful.¹⁰ Personas allow in-depth user/stakeholder analysis to be done to get a better understanding of their characteristics and adapt statistical products and services to them, enhance partnerships, and engage with 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 source 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, for example:

⁹ See the roles of Data Steward described in Chapter 5. Responsibilities and skills for ‘Data Stewardship’.

¹⁰ <https://www.usability.gov/how-to-and-tools/methods/personas.html>

politicians/decision makers (i.e. 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), scientific community, media and the wider public (both using and providing data).

229. 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 are established, the key messages to be communicated to each persona can be developed.

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

Canada

StatCan is currently drafting a series of NSO personas for internal use, to gain a better understanding of employees and their characteristics and contexts. This will enable the adaptation of messaging around data literacy, data strategy, and ultimately data stewardship to these different groups with greater efficiency and efficacy. While still in development, some of these draft personas include Information Technologist, Computer Scientist, Methodologist, Analyst, Researcher, Communicator, Executive, etc. These personas will enable the disaggregation of internal data by persona (particularly, data literacy assessments), but will also facilitate tailored messaging around data stewardship and will advance the cultural change necessary for StatCan to continue to adapt to the accelerated change in this digital data ecosystem.

7.3 New products, services and partnerships as elements of communicating data stewardship

231. The implementation of 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 is the Public Services Monitoring System (SMUP), developed by Statistics Poland (below)¹¹.

232. **Sustainable Development Goals (SDG)** can be cited as another case where NSOs take the role of data steward and communicate it accordingly. The leading role of official statistics in SDGs monitoring was emphasized in the UN's 2030 Agenda for Sustainable Development (2015b). and in many countries, NSOs have become the institution responsible for SDG¹² data management.

233. 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. The 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 also assume an education function, presenting users with a 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 can attract new data users. A holistic approach to communication in statistics for

¹¹ In Poland, legislation gives the President of Statistics Poland certain data steward functions. The Official Statistics Act 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.

¹² See the works of [UNECE Steering Group on Sustainable Development Goals](#)

SDGs is also manifested in the presence of NSOs at events dedicated to the SDGs on the national and international levels, alongside public authorities, academia and NGOs.

234. Having the main communication message building the image of the NSO as the institution responsible for statistical information and knowledge, frequently results in initiatives that go beyond standard statistical cooperation. This effectively re-defines their place in data ecosystems as a result of their numerous and significant 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 is the participation of Statistics Poland in the development of the "Impact Barometer" for Polish business (see Box 11).

235. NSOs from several countries experiencing the same challenges may also consider making a **joint communication strategy**. Communicating as a group to address shared challenges can make their voice stronger. See an example of a joint ESS communication work in respect to data held by the private sector below.

236. Other elements supporting NSOs communication in the context of data stewardship may be **experimental studies** (frontier statistics, non-official statistics). Statistical products and services based on 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.

237. As an example of partnerships for better statistics, **the 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 acting as data stewards. These partnerships tend to produce original, advanced solutions, ready to be implemented in statistical production. Cooperation with science also means gaining new perspectives on research problems, which - when properly communicated - can contribute to a better, stronger image of the NSO as a result of reducing respondents' burden or widening the scope of offered products and services.

238. Another important aspect to consider when looking at the communication about data stewardship is the function of data stewards focused on **data governance**. The introduction of robust ethical frameworks (stemming from the UN Fundamental Principles of Official Statistics) to a wider community of data stakeholders within the data ecosystems helps to fight dis- and misinformation. In these times of disinformation, the role of official statistics and data is even more crucial.

7.4 Critical Communication

239. Provided that data stewardship is tailored to, and allowed to reflect, the local context and the specific data community that it is meant to serve, communication will play 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 communication of the roles that NSOs play as data stewards. This will ensure that data are better understood, better used and better shared, and ultimately to ensure that data adds value to people's lives.

Public Services Monitoring System (SMUP) developed by Statistics Poland

The aim of this 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 the optimization of these services.

The project is conducted in partnership with the Ministry of Internal Affairs and Administration and local governments, with the requirements defined by the local government administrations. The project format is unique, requiring close communication between central government, local governments and the NSO, and the close cooperation with end-users and data providers. This partnership with local government administrations (who are at the same time data providers and data users) is a considerable benefit resulting from the project, in addition to the creation of a user-friendly product/service. A Public Services Monitoring System lowers the burden of the local government administrations (as currently they must report to various institutions) and increases availability and usefulness of data in various systems and registers maintained by different institutions in public administration.

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

The result of cooperation between representatives of science, the private sector, government administration and non-governmental organizations, the Impact Barometer is a set of indicators thanks to which Polish entrepreneurs can assess data about 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.

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*. This “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 words, 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.

8 Conclusion: A foundation for NSO data stewardship leadership

240. This report has proposed that the adoption of a role as a national data steward could represent an opportunity for a National Statistical Office (NSO) to reconsider its position within the national data ecosystem and enhance its standing and relevance.

241. It has described what the resulting leadership role would involve, including current contexts, a proposed stewardship definition, data governance and its relationship to stewardship, stewardship roles and responsibilities and effective ways to communicate data stewardship. If it is to contribute, a data stewardship maturity model should align with and support all those leadership topic areas.

242. **Current contexts and NSO leadership.** As noted in Chapter 2, the criticality of data for the successful operation of public sector agencies has increased tremendously, bringing with it the development of complex data supply chains, leveraging of the latest technological developments, and reliance on sharing and access infrastructure that needs to be monitored and maintained. This has put a renewed focus on the maintenance of public trust and subsequently, highlighted the importance of data stewardship and governance.

243. A data stewardship maturity model can help develop and embed the understanding of these concepts, creating a picture of what 'good' looks like, and help target change that can lead to concrete improvements. That knowledge will prove invaluable for a NSO and its ability to operate effectively in a national data ecosystem leadership role.

244. **Stewardship definition.** Data stewardship is a particularly important concept in current data operating environments, especially in the way it promotes a holistic perspective on the management of data that extends beyond the needs of a specific user. As data is increasingly democratised and represents a more prevalent means of cross-agency coordination and cooperation, that broader level of awareness becomes increasingly important.

245. **Data governance and stewardship.** As described in chapters 3 and 4, there is a particularly strong relationship between data governance and data stewardship. A data stewardship maturity model can help clarify the meaning of data governance and highlight the ways it is meant to operate in conjunction with, and help deliver, data stewardship. It does this through a consideration of data governance as a capability and its role as a dimension of data stewardship, contributing to what good stewardship looks like.

246. **Stewardship roles and responsibilities.** The successful delivery of data stewardship relies heavily upon people's understanding of stewardship, and especially how it relates to their roles and functions. As noted in chapter 5, this includes those with designated stewardship roles as well as others who have responsibility for data and need to assume a level of data accountability in their jobs.

247. **Maturity model.** A data stewardship maturity model is uniquely positioned to embed important ideas and values into the understanding of what constitutes sound data stewardship and, by extension, into organisational structure and practice, to steer agencies in the right direction in this digital data ecosystem. Challenging and complex issues like data sovereignty, representation, and inclusivity can be reflected in stewardship maturity measures, so that agencies using the model are motivated to understand and adopt them. It can increase the general understanding of data stewardship across the organisation, while also helping to facilitate the adoption of data accountabilities for those with data responsible roles. Assessment results can be used to highlight the current state of data stewardship in the organisation, which can then influence the description of data

responsibilities and competencies for various positions in the agency. This can be particularly helpful for an organisation attempting to develop its data capacity, either through recruitment of new staff or the review of existing roles.

248. [Communicating data stewardship](#). Chapter 7 describes various ways that data stewardship, which can represent a challenging concept to understand, can be effectively communicated. Recommendations include a focus on describing the roles a data steward can play in the data ecosystem, a description of the challenges that good data stewardship can address, and the support of a consistent, systematic and sustainable set of data stewardship roles that sit within an NSO operating as a data stewardship lead. A data stewardship maturity model can serve as an effective communication tool, providing a comprehensive, easily understood, and consistent understanding of data stewardship-related capabilities, a view of what good stewardship looks like, and ideas about how to improve practice associated with data stewardship roles.

249. Taken together, these insights can offer an effective and highly practical means of explaining and conveying the value of data stewardship across an organisation, the public sector, or the national data ecosystem. By helping to define and communicate a consistent view of data stewardship, a maturity model for instance can facilitate successful messaging, advocacy and uptake, making it an especially valuable tool for NSOs within their national data contexts. The results are new partnerships and data sources which decrease the response burden and increase the timeliness, quality, interoperability, reusability and fitness-for-purpose of data, contributing to better insights and services to citizens.

Annex 1. Glossary of data stewardship concepts

The glossary has been compiled with the aim to help with the understanding of the current report. It has been developed in consultation with the global Working Group on Data Stewardship (WGDS).

The below definitions have been developed leveraging the materials from NSOs, international and intergovernmental initiatives', organizations' resources (ISO, OECD, EUROSTAT, UNSD, UNECE, SDMX, etc.) and academic publications. While it reflects the best knowledge currently available, some definitions can be further developed as the work in related subject areas matures.

Administrative data

Administrative data describes data and information collected primarily for administrative purposes by organizations, government agencies or other public entities, which is subsequently communicated to governments or NSOs. Administrative data sources are data sets including administrative registers of persons and legal entities; the records of ministries, departments and specialized agencies, such as tax returns, social services records and customs data; or data of regional or local administrations. In contrast to statistical data sources, administrative data sources are not created in response to the need for statistical data but as a part of a government function, such as the provision of services or taxation. In some cases, statistical agencies participate in the design and/or collection of administrative data. In addition, statistical agencies may be involved at different stages of the production process of administrative data, with the aim of ensuring that the data will be usable for statistical purposes (StatCan, 2016; United Nations, 2019).

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, 2008; StatCan, 2021b, 2021c; SDMX, 2018; United Nations, 2003).

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, 2008; StatCan, 2016; SDMX, 2018; United Nations, 2000).

Data

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 official statistics (often from a census, survey statistical register or administrative source) by government agencies or other entities working on behalf of the government. There are different types of censuses, such as population and housing censuses, business censuses and agriculture censuses, among other

types. Sample surveys and statistical registers can cover different units, for example individuals, households and businesses (OECD, 2008, 2021; StatCan, 2021b; United Nations, 2000; United Nations, 2019).

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, 2016).

Data asset

Any entity that is comprised of data and that is strategically used to generate value – whether financial or otherwise. For example, most NSOs prioritize the treatment of data as an asset in order to produce data and information that can be of use to citizens and to the development of the country. A database is a data asset that is comprised of data records. A data asset may be a system or application output file, database, document, or web page. A data asset also includes a service that may be provided to access data from an application. For example, a service that returns individual records from a database would be a data asset. Similarly, a web site that returns data in response to specific queries would be a data asset. Data assets include data from all sources, including surveys, administrative data, record linkage projects, big data and web scraping (National Institute of Standards and Technology, n.d.; StatCan, 2019).

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, 2008).

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, 2008; 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, 2008; 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 (DAMA, 2017; OECD, 2008; United Nations, 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, 2017b; OECD, 2021; StatCan, 2019).

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** 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, 2008, 2019; Plotkin, 2021; StatCan, 2019, 2020a, 2021b).

Data integration

The practice of ingesting, transforming, combining, consolidating, and provisioning data to create consistent, good quality and usable forms (either physical or virtual) 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. (DAMA, 2017; IBM Cloud Education, n.d.; ONS, 2021; 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 (DAMA, 2017; Government of Canada, 2019; StatCan, 2019, 2020a, 2021a).

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 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 (see Annex 3, p. 85) (DAMA, 2017; Government of Canada, 2019; Rancourt, 2019; StatCan 2019).

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) and the statistical production process, to assure data is fit for use and meets the needs of data users. Different users may have different needs that must be balanced.

Several international organizations have a set of criteria defining statistical quality. These include: *relevance, credibility, accuracy, timeliness, accessibility, comparability, interpretability, coherence, and proportionality*, which all contribute to the data and information's overall quality and value (Eurostat, 2003, 2020; OECD, 2002; StatCan, 1998, 2021a; Strong and Wang, 1996; United Nations, 2019).

- 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 independence, 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, 2020; OECD, 2011; StatCan, 1998; United Nations, 2019)

Data modelling

Data modelling is the process of discovering, analysing, and scoping data requirements, and then representing and communicating these data requirements in a precise form called the data model. This process is iterative, may include a conceptual, logical, and physical model, and is used to find relationships between data and make predictions about data (DAMA, 2017, StatCan, 2020a).

Data reuse

The analysis of existing statistical, qualitative or quantitative data collected by one's organization or another, for a new research purpose. To optimize the reuse of data, data and metadata should be well described in order to replicate and/or combine for different purposes (Wilkinson et al., 2016; National Library of Medicine, 2021).

Data security

The definition, planning, development, and execution of security policies and procedures used to provide proper authentication, authorization, access, and auditing of data and information assets. Data security enables the protection of privacy and confidentiality, as well as the maintenance of trust and social acceptability to operate (DAMA 2017; OECD, 2008; Statistics Canada, 2021a; United Nations, 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. 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, 2008; SDMX, 2018; United Nations, 2000).

Data sources

A specific data set, metadata set, database or metadata repository from where data or metadata are available. There are three types of data sources according to their purpose and by the entity responsible for their compilation: statistical data sources such as surveys; administrative data sources; and other data sources. In general, other data sources include data sources associated with the term “big data” unless already included, in some instances, in statistical or administrative data sources. (OECD, 2008; United Nations, 2019).

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, standardized formats and meanings are needed. Examples of data standards include data models, reference data, identifier schemas, and statistical standards. The use of data standards enables 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, 2008; StatCan, 2021b). See also **Statistical standard** and **statistical data standard**.

Data standard, statistical

Statistical data standards are the statistical standards about how data are collected, described, recorded, managed, organized, represented, or formatted. It enables consistent and repeatable description (e.g. definitions), representation (e.g. permitted values, format), structuring (e.g. logical model), and sharing (e.g. exchange model) of data (Standards Council of Canada, 2020; StatCan, 2022).

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 acceptability, 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, 2008, 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 (OECD, 2018; Plotkin, 2021; StatCan, 2020, 2021b; Stats NZ, 2020).

Data strategy

Promotes and provides a course of action for organizing, governing, analysing and leveraging an organizations data assets and expertise to steward resources, protect privacy and security, maintain public trust, and ensure the optimal use and reuse of data in order to provide service and value to citizens. Often emphasizes the use of data as a strategic asset and structures an action plan and set of planned activities around pillars, themes, or guiding principles. These typically include areas of governance or leadership, data stewardship and management, inclusivity, culture or literacy, and technical infrastructure (DalleMule & Davenport, 2017; StatCan 2019; Unites States Government, 2020).

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, 2021a). See also **Appendix 2**.

Generic Statistical Business Process Model

The Generic Statistical Business Process Model (GSBPM) defines and describes the set of business processes needed to produce official statistics. It provides a standard framework and harmonised terminology to help statistical organisations to modernise their statistical production processes along the statistical value chain. The GSBPM can be used for integrating data and metadata standards, as a template for process documentation, for harmonising statistical computing infrastructures, and to provide a framework for process quality assessment and improvement (United Nations, 2019, Stats NZ, 2020a).

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; United Nations, 2000).

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, 2017a; DDI Alliance, 2021; StatCan, 2019). 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, 2017a).

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 (DAMA, 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 (DAMA, 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, 2008, 2021; StatCan, 2021b; United Nations, 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 2008; StatCan, 2016).

National Statistical Office (NSO)

The leading statistical agency within a national statistical system. National statistical office and national statistical institute mean the same thing. In general, the NSO has a coordination role within the national statistical system, and is responsible for the development, production, and dissemination of official statistics across multiple statistical domains. (United Nations, 2019).

National Statistical System (NSS)

The ensemble of statistical organizations and units (statistical agencies) within a country that develop, produce and disseminate official statistics on behalf of the national Government (and other levels of government) (United Nations, 2019).

Necessity and Proportionality Principle

This principle holds that the potential benefits of the public interest pursued should be reasonably balanced against the interests of other stakeholders and data subjects, where *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). (European Commission, 2022; European Data Protections Supervisor, 2020; Rancourt, 2021; StatCan 2019; 2020b). See also The Necessity and Proportionality Principle, Annex 3.

Official Statistics

Statistics that describe, on a representative basis, economic, demographic, social and environmental phenomena of public interest. Official statistics are developed, produced and disseminated as a public good by the members of the NSS in compliance with the Fundamental Principles of Official Statistics and accepted quality frameworks such as the United Nations National Quality Assurance Frameworks Manual for Official Statistics (UN NQAF), as well as other internationally agreed statistical standards and recommendations. (United Nations, 2003, 2019,).

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, 2008; StatCan, 2021c; United Nations, 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, 2018; 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, 2008; StatCan, 2019; 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, 2008; StatCan, 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).

Source Data

Data collected on a regular basis (from respondents, administrative entities and other data providers) by members of the national statistical system to be edited, imputed, aggregated and/or used in the compilation and production of official statistics (OECD, 2008; United Nations, 2019).

Statistics

Statistics are numerical information relating to an aggregate of data on units or observations; the science of collecting, analysing and interpreting such data. In general, the term statistics refers to an output of a statistical production process and the term data refers to input and possibly throughput in the statistical production process (the term data includes microdata which, depending on the context, can be also an output) (OECD, 2008; United Nations, 2019).

Statistical Standard

Statistical standards define and establish uniform specifications and characteristics about all aspects of statistical production, either processes/capabilities, the data/metadata they use, products and/or services. They constitute a comprehensive set of statistical concepts, definitions, classifications and models, methods and procedures used to achieve the uniform treatment of statistical issues within or across processes and across time and space. Statistical standards are a subset of data standards, so data standards will include the same content as you would see in a statistical standard, but additionally include end to end guidance to make implementation easier (; OECD, 2008; StatCan, 2022; Stats NZ, 2022; United Nations, 2019).

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, 2008; United Nations, 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, 2008; StatCan, 2016).

Variable

A characteristic, attribute, or property of a statistical unit or population 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.). A variable combines a concept and a unit type in order to define the characteristic of measure (OECD, 2008; StatCan, 2016; United Nations, 2013).

Annex 2. Data Stewardship and the Fundamental Principles of Official Statistics

Developed and adopted by the Conference of European Statisticians, the United Nations Fundamental Principles of Official Statistics¹³ 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.

The Fundamental Principles¹⁴ provide an international, intergovernmental guide for how NSOs can best provide their services to citizens. 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 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).

FPOS constitute the basis for enabling trust and form a solid basis for all ethical and quality-related conceptual documents in official statistics. Using these principles contributes to the production of high quality official statistical information, which is crucial in analysis and informed policy decision-making processes. This section aims to analyze the relationship between each of the Fundamental Principles and data stewardship.

Principle 1: Relevance, impartiality and equal access

“Official statistics provide an indispensable element in the information system of a democratic society, serving the government, the economy and the public with data about the economic, demographic, social and environmental situation. To this end, official statistics that meet the test of practical utility are to be compiled and made available on an impartial basis by official statistical agencies to honour citizens' entitlement to public information.”

National statistical offices are not the only producers of official data. The role of NSOs as data stewards includes coordinating statistical production. Data stewards also need to ensure that their statistics and those produced by other public entities meet the existing standards of relevance, impartiality and equal access. This involves fostering a common understanding of the data assets within the NSS (people, technology and processes) to improve the quality of the data.

The data steward is expected to standardize the conceptual and methodological frameworks of official statistics production and ensure that the data outputs are available on an impartial basis. This coordination role also translates into benefits for the statistical production of the NSO. By accessing to new data sources, such as administrative records provided by public and private institutions, NSOs can enrich the relevance and timeliness of their outputs. Furthermore, relevance can be guaranteed through the steward's coordination, by identifying clear achievable needs, avoiding the duplication of

¹³ https://unece.org/sites/default/files/2022-06/ECE_CES_2022_29-2208645E_0.pdf

¹⁴ 2022 marks the 30th anniversary of the Fundamental Principles. The anniversary campaign includes useful materials explaining and promoting the Principles. For more information, see: <https://unece.org/FPOS30>

information generated by different stakeholders in the data ecosystem, and promoting evidence-based public policy.

As statistics are increasingly based on combined sources of data, it is important to ensure that the use of “new data sources” is consistent with Principle 1. Validation of information and crosschecks are indispensable. The more actors operate within the emerging data ecosystems, the richer they become but, at the same time, more trust is needed. This trust is engendered on the solid basis of FPOS - the shared standards which ensure transparency and resilience.

Principle 2: Professional standards and ethics

“To retain trust in official statistics, the statistical agencies need to decide according to strictly professional considerations, including scientific principles and professional ethics, on the methods and procedures for the collection, processing, storage and presentation of statistical data.”

Being based on professional considerations and ethics helps to retain trust in official statistics as the public and users can rely on the fact that the processes carried out by the NSO are conducted according to an existing normative framework and are not moved by “political” drivers. When broadening functions towards a data stewardship role, the national statistical offices should ensure that the data provided by other stakeholders complies with pre-established professional standards. These standards guarantee that the personnel working in all producing entities of official statistics, particularly the top management, have the accurate profile to do so. This prevents the NSS from malpractices and assures the quality of the data while keeping away biased information.

24. When NSOs guide the adoption of ethical and professional standards in the statistical environment, this gives an opportunity to target equity and inclusion-related goals within official statistics. This could help to make visible populations that are often, for financial or methodological reasons, underrepresented in official statistics, guaranteeing an inclusive approach in statistics (e.g. indigenous communities, LGBTIQ+, among others). The involvement of underrepresented population groups along the data value chain is also important. In doing the above, the data steward can highlight the need to transform standards to better collect, produce and disseminate data.

25. Nowadays when the boundaries of official and non-official statistics are blurred in the public’s opinion, the proliferation of new sources of data calls for other actors to comply with the highest ethical standards ensured by the FPOS. Mass media, social media, private stakeholders, NGOs, citizens - they all constitute the modern sources of information. Having other stakeholders align with professional standards and ethics will result in benefits for both producers and users of data.

Principle 3: Accountability and transparency

26. “To facilitate a correct interpretation of the data, the statistical agencies are to present information according to scientific standards on the sources, methods and procedures of the statistics.”

27. Being transparent and open to scrutiny is fundamental to foster trust in official statistics. Since the adoption of the FPOS, NSOs have been improving the transparency of their work, publishing metadata, publication calendars, being subject to audits and external evaluations. However, as data stewards the NSOs should not only make themselves but also others accountable. Data stewards are needed to promote transparency and become scrutinizers of information produced by other government agencies. Metadata, integration methodologies of administrative records, anonymization processes, etc. must be publicly available. NSOs’ expertise in this matter is paramount. In the data-driven world different professions dealing with data may have common goals. For example,

statisticians and responsible journalists, in their pursuit for the truth, have common goals and foundations (very close to those set out in the FPOS), yet they often speak different languages. Data stewardship should foster closer links between different professional groups dealing with data for the common goal of informing and thereby empowering through the use of unbiased, solid, verified information.

28. Moreover, sharing professional and ethical standards with the wider audience contributes to accountability and transparency.

Principle 4: Prevention of misuse

29. “The statistical agencies are entitled to comment on erroneous interpretation and misuse of statistics.”

30. Data stewards must promote data literacy to support the correct interpretation of information. It is the steward’s responsibility to educate specialized and non-specialized audiences to better understand the scope and limitations of official statistics. NSOs have the responsibility to unlock the value of data, for public decision-makers, civil society organizations, private sector, and the public, particularly vulnerable communities. Misusing statistics may lead to misguided decisions, jeopardizing people’s agency and their trust in official statistics. This responsibility is also transferred to new official statistics producers, and stewards should also encourage them to promote data literacy.

Principle 5: Sources of official statistics

31. “Data for statistical purposes may be drawn from all types of sources, be they statistical surveys or administrative records. Statistical agencies are to choose the source with regard to quality, timeliness, costs and the burden on respondents.”

32. The new data processing technologies have allowed to expand the sources from which statistical information can be drawn. NSOs as coordinators of national statistical systems should encourage collaboration schemes that stimulate access and integration of different data sources to improve timeliness and quality of statistics, and reduce costs and respondent burden. The data stewards, on one hand, should widen access to so called non-traditional sources. Those sources can include administrative records (in many cases already accessible to NSOs by law), big data with all its variations, and privately held data. Fostering the use of more data sources, NSOs should help shape data governance to facilitate sharing and collaboration.

33. On the other hand, NSOs as data stewards should empower other producers to use new sources of information, to both reduce respondent burden and increase quality of outputs. Capacity-building of other producers in using new sources will help promote the FPOS compliance.

Principle 6: Confidentiality

34. “Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes.”

35. Confidentiality plays a key role in maintaining trust of both data providers and users. The data steward needs to ensure that these good practices, particularly anonymization techniques, are fully followed by all entities that produce official statistics, and when using all different types of data sources.

36. Data stewardship implies data protection to maintain trust, which in turn means using the collected data only for statistical purposes, preventing its use for other purposes.

Principle 7: Legislation

37. “The laws, regulations and measures under which the statistical systems operate are to be made public.”

38. Formal governance schemes include the mandates of any public institution and shape expectations of stakeholders in data ecosystem. FPOS are legally mainstreamed into national contexts through national legislations. From the data stewardship perspective, NSOs as leaders of the national statistical systems are called to promote adherence to statistical legislation.

39. Additionally, the data steward should create the governance and coordination mechanisms in accordance with the legislation.

Principle 8: National coordination

40. “Coordination among statistical agencies within countries is essential to achieve consistency and efficiency in the statistical system.”

41. Data stewardship is closely linked with coordination of the statistical system, which is essential to achieve consistency and efficiency. Coordination involves ensuring the harmonious work of the different players in the data ecosystem, guiding and creating collaborations, and assuring compliance with standards. To be able to execute this role, a national data strategy and a clear legislation with the explicit role of the NSO as data steward would be considerable assets.

42. National coordination as part of data stewardship should ensure efficiency of the processes, guarantee data relevance, promote data integration, avoid duplication and raise awareness about the importance of data as a strategic asset. The data steward should link the data produced by state entities, creating relationships between official entities at the national level, and between those at national and subnational levels.

Principle 9: Use of international standards

43. “The use by statistical agencies in each country of international concepts, classifications and methods promotes the consistency and efficiency of statistical systems at all official levels.”

44. The use of international concepts, classifications and methods in official statistics is essential to retain trust. As data stewards, NSOs need to integrate international standards into the practices of all the producers of data to promote the consistency and efficiency of statistical systems at all official levels. For instance, data stewards could assist other entities providing data to comply with quality of international standards for the use of administrative records, or anonymization processes.

Principle 10: International cooperation

45. “Bilateral and multilateral cooperation in statistics contributes to the improvement of systems of official statistics in all countries.”

46. International cooperation in statistics contributes to the improvement of official statistical systems in all countries. Assuming the responsibility of data stewards means to support other actors in the national statistical systems to participate in bilateral or multilateral cooperation. In this way,

data stewardship supports the mechanisms of sharing and collaboration to strengthen the national data ecosystem and its relationship with international actors. In addition, stewards themselves are called to share and collaborate to exchange experience and identify best practices on operationalizing data stewardship and adapting it to national contexts.

47. There is also an on-going international discussion on how the position of an NSO as head of the NSS is subject to change and evolve in the international context.

CORE VALUES OF OFFICIAL STATISTICS

To help the implementation of FPOS and their translation into everyday statistical practice, a Conference of European Statisticians' Task Force has developed a list of 'core values of official statistics and the related behaviours' (2022).

These core values and behaviours may be helpful in fulfilling the data steward role, especially towards actors outside NSS where the FPOS do not apply.

The core values are those common values that have guided the development of the official statistics community, including the formulation of key standards such as the Fundamental Principles of Official Statistics, regional and national codes of practice and quality frameworks. They have always been implicitly present in the background of statistical work, but are now being explicitly elaborated. The core values are intended to be complementary to the Fundamental Principles and other standards, providing an additional element to help define and communicate the concept of official statistics. Looking forward, it is intended that the core values will help to guide future decisions on the development of official statistics at the national and international levels.

The core values are reflected standards and practices of official statistics, including the United Nations Fundamental Principles of Official Statistics. The core values provide an ethical framework to support decision-making and interactions with governments, society and other stakeholders. The aim of communicating these core values is to promote trust in official statistics and the organisations that produce them, by helping to articulate what sets official statistics apart from other data.

The official statistics community upholds the following core values of official statistics:

1. Relevant

The official statistics community ensures relevance through:

- An **engaged, responsive, user-centric** approach, based on clear and regular **communication** with all stakeholders;
- A **proactive, innovative** and **agile** approach to the development of new products, services, tools, methods and sources;
- Applying appropriate **standards** and **quality frameworks** to produce **accurate, timely, coherent** and **comparable** statistics.

2. Impartial

The official statistics community:

- Is **objective**, in all aspects of our work, acting with **fairness** and **integrity** to serve the **public good**;
- Ensures **equal access** for all users through all relevant channels;
- Produces **inclusive** statistics reflecting all parts of society.

3. Transparent

Official statistics, methods, processes, products and quality reports are communicated to the public through appropriate channels, and are **open** to scrutiny.

4. Professionally independent

Credible, trustworthy and authoritative official statistics are based on:

- Freedom from external interference;
- The application of **professional expertise** and scientific principles;
- **Countering misuse** and misinterpretation of official statistics.

5. Respects confidentiality

The official statistics community:

- Protects privacy by ensuring that data collection is limited to what is **necessary and proportionate**;
- Upholds the **confidentiality** of data on individual people, households and businesses;
- Acts in an **ethical** way, following professional ethics standards.

6. Collaborative

Official statistics are strengthened by sharing ideas and experiences within the international statistical **community** and with other stakeholders.

The Task Force also developed related behaviours that are intended to encourage official statistics organisations to act in ways that inspire the confidence and trust of users, whilst upholding strict professional and ethical standards. They provide practical guidance on what it means to implement the Core Values and Fundamental Principles. The core values and behaviours are planned to be agreed upon by the Conference of European Statisticians in June 2023.

DATA STEWARDSHIP ROLE OUTSIDE THE NATIONAL STATISTICAL SYSTEM AND THE FUNDAMENTAL PRINCIPLES

The links as explained above pertain mainly to the role of a statistical office as a data steward in the national statistical system (NSS). The activities of NSOs and NSS are based on FPOS and therefore, the data stewardship role in NSO and NSS should be fully in line with the FPOS.

The situation is different when an NSO has some data stewardship role outside the statistical system, e.g. concerning administrative registers and other public sector data. FPOS state that ‘Individual data collected by statistical agencies for statistical compilation [...] are to be strictly confidential and used exclusively for statistical purposes’. The data in the public sector may be used for administrative purposes and (some of) the Fundamental Principles do not apply. Therefore, the statistical offices have to be very careful what kind of stewardship functions to undertake, in both what they do and how this is perceived by other stakeholders and the public.

What statistical offices can do:

- Promote, explain and raise awareness about FPOS;
- Advocate for the core values and behaviours of official statistics;
- Advocate for and help develop similar principles and values to be used in the public sector data work, and in the whole national data ecosystem.

Annex 3. Supporting principles of data stewardship

250. This Annex 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 by describing roles and some of the skills expected of data stewards. The concepts described below include: the Fundamental Principles of Official Statistics, standardization, the FAIR Guiding Principles of data, the Four G Model of data lifecycle management, CARE model, the ‘once only’ principle and the European Interoperability framework.

Standards

251. As professionals of the international data community, NSOs 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*.

252. 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 and registers. 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 and businesses.

253. 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. 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.

254. 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.

A Human Rights-based approach to data¹⁵

The Office of the UN High Commissioner for Human Rights has prepared general guidance on a Human Rights-Based Approach to Data, with a focus on issues of data collection and disaggregation. This approach helps bring together relevant data stakeholders and develop communities of practice that improve the quality, relevance and use of data and statistics consistently with international Human Rights norms and principles. The guidance draws from internationally agreed principles related to statistics and data, including the FPOS. A preliminary set of key principles have been formulated under the following headings:

- **Participation** of relevant population groups in data collection exercises, including planning, data collection, dissemination and analysis of data
 - Consider a range of processes that facilitate and encourage participation
 - Clearly communicate how participatory processes are conducted and the outcomes of these exchanges
 - Ensure that the views of vulnerable or marginalized groups, and groups who are at risk of discrimination, are represented
 - Maintain knowledge holdings and institutional memory in relation to information gathered through participatory processes
- **Disaggregation** of data allows data users to compare population groups, and to understand the situations of specific groups. Disaggregation requires that data on relevant characteristics are collected
 - More detailed data than national averages is key in identifying and understanding inequalities
 - Data should be disaggregated by key characteristics identified in international human rights law
 - Collection of data to allow disaggregation may require alternate sampling and data collection approaches
 - Birth registration is foundational for robust data sets that allow accurate disaggregation
- **Self-identification.** For the purposes of data collection, populations of interest should be self-defining. Individuals should have the option to disclose, or withhold, information about their personal characteristics
 - Data about personal characteristics should be provided by the individuals to whom the data refers (at the individual's discretion)
 - Data collection activities should be conducted in accordance with the human rights principle of 'doing no harm'
- **Transparency.** Data collectors should provide clear, openly accessible information about their operations, including research design and data collection methodology. Data collected by State agencies should be openly accessible to the public
 - Official Statistics are part of the public's right to information
 - Information about how data is collected should be publicly available
 - Data should be disseminated as quickly as possible after collection
- **Privacy.** Data disclosed to data collectors should be protected and kept private, and confidentiality of individuals' responses and personal information should be maintained

¹⁵

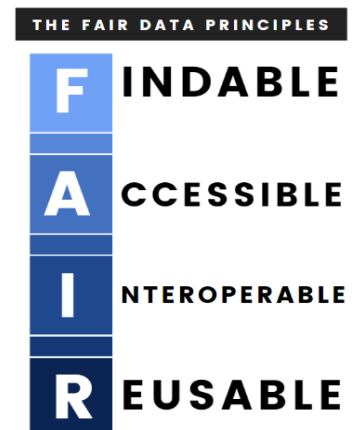
<https://www.ohchr.org/sites/default/files/Documents/Issues/HRIndicators/GuidanceNoteonApproachtoData.pdf>

- Privacy and confidentiality must be considered alongside access to information
- Information that identifies individuals or discloses an individual's personal characteristics should not be made public as a result of data dissemination
- Data collectors must have robust data protection mechanisms and procedures
- When personal data is released, this should only be done with the permission of the individual concerned (or their appropriate representatives)
- **Accountability.** Data collectors are accountable for upholding human rights in their operations, and data should be used to hold States and other actors to account on human rights issues
 - Data can, and should, be used to hold human rights actors to account
 - National Statistical Offices are human rights duty-bearers and are accountable for respecting, protecting and fulfilling human rights

The FAIR guiding principles of data¹⁶

255. 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.

256. The FAIR 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).



257. 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.

The CARE principles for indigenous data governance¹⁷

258. The increasing convergence of technology infrastructure and digital connectivity has raised the value of data across the globe, and as such have important implications for Indigenous Peoples' ability to exercise their individual and collective rights to self-determination. Indigenous Peoples are often excluded from decision-making fora and their knowledge when such knowledge exists only as part of an oral tradition.

¹⁶ <https://www.go-fair.org/fair-principles/>

¹⁷ <https://www.gida-global.org/care>

82. The increasing convergence of technology infrastructure and digital connectivity has raised the value of data across the globe, and as such have important implications for Indigenous Peoples’ ability to exercise their individual and collective rights to self-determination. Indigenous Peoples are often excluded from decision-making fora and their knowledge when such knowledge exists only as part of an oral tradition.

83. The UN Declaration on the Rights of Indigenous Peoples (UNDRIP) reaffirmed Indigenous rights to self-governance and authority to control their Indigenous cultural heritage embedded in their languages, knowledge, practices, technologies, natural resources, and territories (i.e. Indigenous data). Indigenous data, which include data collected by governments and institutions about Indigenous Peoples and their territories, are intrinsic to Indigenous Peoples’ capacity and capability to realise their human rights and responsibilities to all of creation.

84. The current movement toward open data and open science does not fully engage with Indigenous Peoples’ rights and interests. The FAIR principles (above) primarily focus on characteristics of data that will facilitate increased data sharing among entities but ignores power differentials and historical contexts. The emphasis on greater data sharing alone creates a tension for Indigenous Peoples who are also asserting greater control over the application and use of Indigenous data and Indigenous Knowledge for collective benefit.

85. This includes the right to create value from Indigenous data in ways that are grounded in Indigenous worldviews and realise opportunities within the knowledge economy. The CARE Principles for Indigenous Data Governance (Caroll et al., 2020) are people and purpose-oriented, reflecting the crucial role of data in advancing Indigenous innovation and self-determination.

86. These principles complement the existing FAIR principles encouraging open and other data movements to consider both people and purpose in their advocacy and pursuits. As data stewardship “represents the ethical and responsible creation, collection, management, use, and reuse of data”, and 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.” the addition of the CARE principles to the FAIR principles ensures that the public good aspect of data stewardship is kept visible.



The CARE model compared to FAIR principles.

87. The CARE Principles:

Collective Benefit	Data ecosystems shall be designed and function in ways that enable Indigenous Peoples to derive benefit from the data.
Authority to Control	Indigenous Peoples’ rights and interests in Indigenous data must be recognised and their authority to control such data be empowered. Indigenous data governance enables Indigenous Peoples and governing bodies to determine how Indigenous Peoples, as well as Indigenous lands, territories, resources, knowledges and geographical indicators, are represented and identified within data.

Responsibility	Those working with Indigenous data have a responsibility to share how those data are used to support Indigenous Peoples’ self-determination and collective benefit. Accountability requires meaningful and openly available evidence of these efforts and the benefits accruing to Indigenous Peoples.
Ethics	Indigenous Peoples’ rights and wellbeing should be the primary concern at all stages of the data life cycle and across the data ecosystem.

The 4Gs Model of Data Lifecycle Management

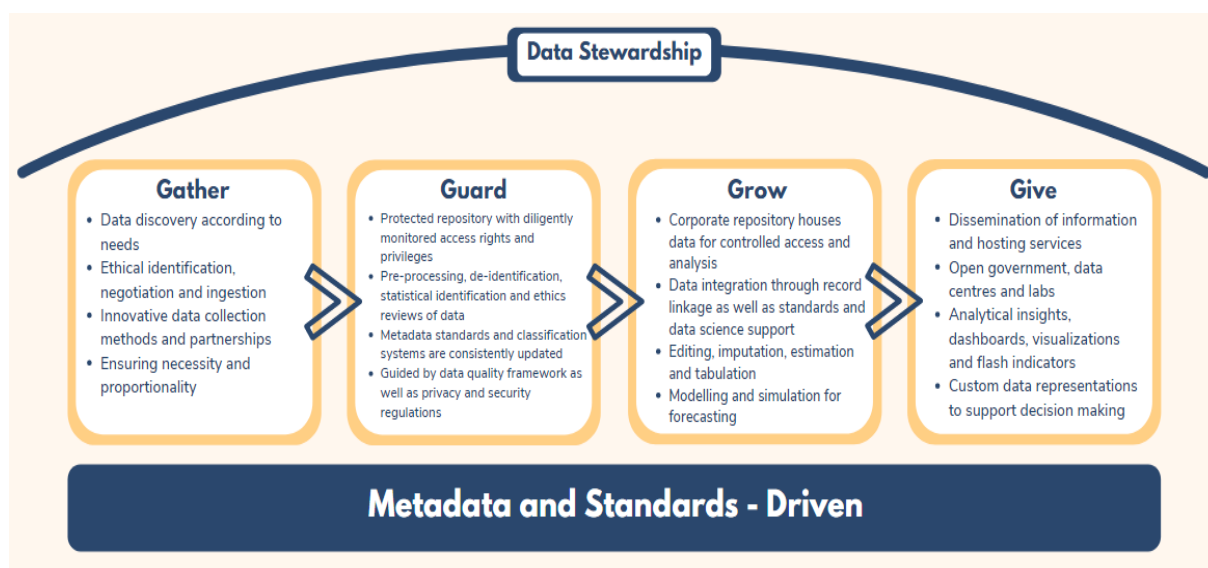
88. To advance National Statistical Systems (NSS) globally, a commitment to sound stewardship is required throughout the data lifecycle. To achieve this, Eric Rancourt of Statistics Canada developed the “Four Gs” model of data life-cycle management – *gather*, *guard*, *grow*, and *give* (2019).

89. 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.

90. 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.

91. 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 analysed, 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.

92. 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.



93. Organizing data stewardship activities along these data lifecycle phases and aligning them to the overall role of the NSO – which indeed, is to gather, guard, grow, and give data – has been extremely valuable. It has allowed NSOs 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.

The 'once only' principle

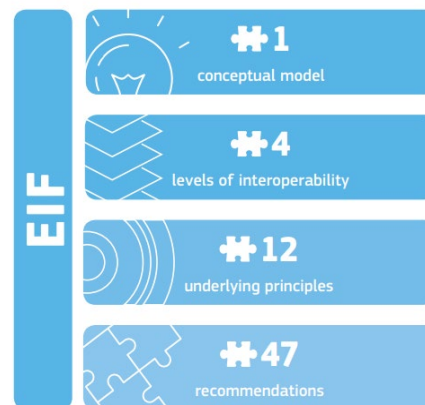
94. The once-only principle is an e-government concept that aims to ensure that citizens, institutions, and companies only have to provide certain standard information to the authorities and administrations once. By incorporating data protection regulations and the explicit consent of the users, the 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*). "[European eGovernment Action Plan 2016-2020](#)". *Digital Single Market*).

95. 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.

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

The European Interoperability Framework

97. 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)

98. 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.

99. 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

100. 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.

The Necessity and Proportionality Principle

101. 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, where *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, 2022; European Data Protections Supervisor, 2020; Rancourt, 2021; StatCan 2019; 2020b).

References

- Australian Bureau of Statistics (2021). DataLab. Australian Bureau of Statistics.
- Australian Government (n.d.). A Guide for Data Integration Projects Involving Commonwealth for Statistical and Research Purposes. National Statistical Service.
- _____ (2021). *Australian Data Strategy: The Australian Government's whole-of-economy vision for data*. Department of the Prime Minister and Cabinet.
- _____ (2022). Data Availability and Transparency Act 2022: An Act to authorise the sharing of public sector data, and for related purposes. Federal Register of Legislation.
- Carroll, S.C. and others (2020). The CARE Principles for Indigenous Data Governance. *Data Science Journal*, 19: 43, pp. 1–12.
- Central Statistics Office (Ireland) (2023). Ordnance Survey Ireland. Central Statistics Office
- Croatian Bureau of Statistics. (n.d.). GEOSTAT Portal: Connecting Spatial Data with Statistics. Croatian Bureau of Statistics
- DalleMule, L. and T.H. Davenport (2017). What's Your Data Strategy?. Data Management. *Harvard Business Review*, (May- June).
- Data Governance Institute (n.d.). Governance and Decision Making. Data Governance Institute.
- Data Documentation Initiative Alliance (2021). DDI Alliance Glossary. DDI Alliance.
- Data Management Association (DAMA) (2017). *DAMA-DMBOK: Data Management Body of Knowledge (2nd Ed.)*. Basking Ridge, United States of America: Technics Publications.
- de Bruin, T. and others (2005). Understanding the main phases of developing a maturity assessment model. In *ACIS 2005 Proceedings - 16th Australasian Conference on Information Systems*.
- Destatis – German Federal Statistical Office (n.d.). Dashboard Deutschland. Destatis.
- Enterprise Data Management Council (EDMC) (2015). Data Management Capability Assessment Model. EDMCouncil.
- European Commission (n.d.) European data strategy: Making the EU a role model for a society empowered by data. European Commission.
- _____ (2015). Regulation (EU) 2015/759 of the European Parliament and of the Council of 29 April 2015 amending Regulation (EC) No 223/2009 on European statistics
- _____ (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

_____ (2017a). European Interoperability Framework. Luxembourg, Belgium: Publications Office of the European Union.

_____ (2022). Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act).

European Commission, Directorate-General for Communications Networks, Content and Technology (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, Belgium: Publications Office of the European Union.

European Commission, Eurostat (2003). *Assessment of quality in statistics - Definition of Quality in Statistics*, Working Group, Luxembourg, October 2003.

_____ (2020). *Quality assurance framework of the European statistical system: version 2.0*. Luxembourg, Belgium: Publications Office of the European Union.

_____ (2022). *Privately held data communication toolkit*. Luxembourg: Publications Office of the European Union.

European Commission, European Political Strategy Centre (2017b). Enter the data economy: EU policies for a thriving data ecosystem. Luxembourg, Belgium: Publications Office of the European Union, 21:11.

European Commission, Interoperability solutions for public administrations, businesses and citizens (ISA) Programme (2015). *Methodology and tools for metadata governance and management for EU Institutions* [PDF]. Brussels, Belgium: European Commission.

European Data Protection Supervisor (2020). Necessity and Proportionality. Annual Report 2020. European Data Protection Supervisor.

Fellegi, I. P. (1995). *Characteristics of an Effective Statistical System* [Paper presentation]. Morris Hansen Lecture, Washington Statistical Society 1995, Washington, DC, United States, October.

Financial Stability Board Secretariat and International Monetary Fund Staff (2022). *G20 Data Gaps Initiative (DGI-2) — Progress achieved, lessons learned, and the way forward*. The Financial Crisis and Information Gaps, June.

Firican, G. (n.d.). IBM Data Governance Maturity Model. LightsOnData.

Floridi, L. and M. Taddeo (2016). What is data ethics?. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2083), pp. 20160360.

Gray, J. and others (2005). Scientific data management in the coming decade. *Special Interest Group on Management of Data*, 34:4 (34–41).

Gartner (2014). Gartner's Enterprise Information Management Maturity Model. Gartner Research. Gartner.

_____. (2021). Information Technology. Gartner Glossary. Gartner.

Government of Canada (2018). *Report to the Clerk of the Privy Council: A Data Strategy Roadmap for the Federal Public Service*. Ottawa, Canada: Her Majesty the Queen in right of Canada.

_____ (2019). *Policy on Service and Digital*. Ottawa, Canada: Her Majesty the Queen in right of Canada.

_____ (2021). *Canada's Digital Charter: Trust in a digital world*. Ottawa, Canada: Her Majesty the Queen in right of Canada.

Government of Ireland (2019). *Public Service Sector Data Strategy 2019 – 2023*. Office of the Government Chief Information Officer. Department of Public Expenditure and Reform.

IBM Cloud Education. (2020). IBM Cloud Learn Hub. IBM Cloud Education.

Information Commissioner's Office of the United Kingdom (2019). *Data sharing: a code of practice. Guide to Data Protection*.

Information Systems Audit and Control Association (2014). *Capability Maturity Model Integration (CMMI)*. CMMI Institute.

International Organization for Standardization (2016) *Data quality — Part 61: Data quality management: Process reference model* (ISO standard no. 8000-61:2016).

_____ (2022). *Information security, cybersecurity and privacy protection — Information security management systems — Requirements* (ISO standard no. 27001:2022).

Irving Fisher Committee on Central Bank Statistics (IFC) (2019). *The use of big data analytics and artificial intelligence in central banking*. IFC Bulletin, no. 50, May.

Irving Fisher Committee on Central Bank Statistics (IFC) (2020). *Using financial accounts*. IFC Bulletin, no. 51, February.

Irving Fisher Committee on Central Bank Statistics (IFC) (2021). *Issues in data governance*. IFC Bulletin, no. 54, July.

Ishwarappa, K. and J. Anuradha (2015). *A Brief Introduction on Big Data 5Vs Characteristics and Hadoop Technology*. *Procedia Computer Science*, 48. 319-324.

Ladley, J. (2020). *Data Governance: How to Design, Deploy, and Sustain and Effective data Governance Program (2nd Ed.)*. London, United Kingdom: Academic Press.

McGilvray, D. (2020). *Executing Data Quality Projects. Ten Steps to Quality Data and Trusted Information (TM)*. London, United Kingdom: Academic Press.

Miller, R. J. (2014). *Big Data Curation* In 20th International Conference on Management of Data (COMAD). Hyderabad, India.

National Institute of Standards and Technology (n.d.). *Glossary – Data Asset*. Computer Security Resource Centre. Computer Security Resource Centre. United States Government.

National Library of Medicine (2021). *Data Glossary*. Network of the National Library of Medicine. United States Government.

- Office of National Statistics (ONS) (United Kingdom) (2001). *ONS launches Integrated Data Service to boost government collaboration on data sharing*. Office of National Statistics.
- _____ (2020). *Data Strategy: Data Principles*. Office of National Statistics.
- Open Data Institute (ODI) (2015). *Open Data Maturity Model*. Open Data Institute.
- Organisation for Economic Co-operation and Development (2002). *Measuring the Non-Observed Economy: A Handbook*. Paris, France: OECD Publications.
- _____ (2008). *Glossary of Statistical Terms*. OECD – Better Policies for Better Lives.
- _____ (2011). *Quality Framework for OECD Statistical Activities* [PDF]. Paris, France: OECD Publications.
- _____ (2016). *Open Government Data Review of Mexico: Data Reuse for Public Sector Impact and Innovation*. OECD Digital Government Studies. OECD Publishing, Paris, France: OECD Publications.
- _____ (2017). *The role of national statistical systems in the data revolution, In Development Co-operation Report 2017: Data for Development*. Paris, France: OECD Publications.
- _____ (2018). *Governing open data for sustainable results, in Open Government Data Report: Enhancing Policy Maturity for Sustainable Impact*. Paris, France: OECD Publications.
- _____ (2019). *Data Governance in the Public Sector In The Path to Becoming a Data-Driven Public Sector*, OECD Digital Government Studies. Paris, France: OECD Publications.
- _____ (2020). *Digital Government Index: 2019 results*. OECD Public Governance Policy Papers, No. 03. Paris, France: OECD Publications.
- _____ (2021). *Recommendation of the Council on Enhancing Access to and Sharing of Data*. OECD Legal Instruments.
- OSTHUS (2022). *The importance of a Data Management Maturity model for organizations*. OSTHUS, March 28.
- Peng, G. (2018). *The State of Assessing Data Stewardship Maturity – An Overview*. *Data Science Journal*, 17, p.7.
- Plotkin, D. (2021). *Data Stewardship: An Actionable Guide to Effective Data Management and Data Governance (2nd Ed.)*. London, UK: Academic Press.
- Pöppelbuß, J. & M. Röglinger (2011). *What makes a useful maturity model? A framework of general design principles for maturity models and its demonstration in business process management*. *19th European Conference on Information Systems, ECIS (19)*.
- Proença, D. & J. Borbinha (2018). *Maturity Models for Data and Information Management*. *In Digital Libraries for Open Knowledge* (pp.81-93). New York, United States: Springer.

- Rancourt, E. (2019). The scientific approach as a transparency enabler throughout the data life-cycle. *Statistical Journal of the International Association for Official Statistics*, 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).
- Rosolia, A., S. Stapel-Weber and B. Tissot (2021). New developments in central bank statistics around the world. *Statistical Journal of the International Association for Official Statistics*, vol. 37, no. 4 (1055-1060).
- Standards Council of Canada. (2020). What Are Standards? Standards. Standards Council of Canada.
- Statistics Canada. (1998). *Quality Guidelines (3rd ed.)*. Ottawa, Canada: Authority of the Minister responsible for Statistics Canada.
- _____ (2016). *Statistics Canada Policy on the Use of Administrative Data Obtained under the Statistics Act*. Ottawa, Canada: Her Majesty the Queen in Right of Canada.
- _____ (2019). *Statistics Canada Data Strategy: Delivering insight through data for a better Canada* [PDF]. Statistics Canada.
- _____ (2020a). Data Literacy Competencies. Data Literacy Training. Statistics Canada.
- _____ (2020b). Principles of Necessity and Proportionality. Statistics Canada's Trust Centre. Statistics Canada.
- _____ (2021a). *Statistics Canada's Approach to Data Stewardship* [PDF]. Unpublished internal departmental document.
- _____ (2021b). *Enterprise Information and Data Management Glossary* [PDF]. Unpublished internal departmental document.
- _____ (2021c). Introduction to the National Occupational Classification (NOC) 2021 Version 1.0. Definitions, Data Sources, and Methods. Statistics Canada.
- _____ (2022). Managing statistical data with standards. Welcome to Information and Data Management. Statistics Canada.
- Statistics New Zealand | Tatauranga Aotearoa (n.d.). A Data Stewardship Framework for NZ. Data.gov.nz - Discover and Use Data.
- _____ (2020). Data stewardship: managing New Zealand's data better to change lives. A Data Stewardship Framework for NZ.
- _____ (2021). The Government Data Strategy and Roadmap. Data.gov.nz - Discover and Use Data.
- _____ (2022). Stats NZ's statistical standards will become data standards. Data.gov.nz - Discover and Use Data.
- Strategic Data and Metadata eXchange (SDMX) (2018). SDMX Glossary Version 2.1. Strategic Data

and Metadata eXchange

System Analysis Program (SAP) (n.d.). *Business and Technology Fundamentals: Data management Glossary*. SAP.

The Open Group (2011). *The Open Group Service Integration Maturity Model*. The Open Group.

Treloard, A., D. Groenewegen and Harboe-Ree, C. (2007). The Data Curation Continuum: Managing data objects in institutional repositories. *D-lib Magazine*, 13:9/10.

United Nations, Department of Economic and Social Affairs, Statistics Division (2003). *Handbook of Statistical Organization (3rd ed.): The Operation and Organization of a Statistical Agency*. New York, United States of America: United Nations.

_____ (2019). *United Nations National Quality Assurance Frameworks Manual for Official Statistics*. New York, United States of America: United Nations.

_____ (2021a). *Fundamental Principles of National Official Statistics*. United Nations Statistics Division.

_____ (2021b). *UN Global Platform: Data for the World*. UN Big Data.

United Nations, Economic Commission of Europe (2000). *Terminology on Statistical Metadata* In Conference of European Statisticians Statistical Standards and Studies (53). Geneva, Switzerland: United Nations.

_____ (2013). *Variable. Generic Statistical Information Model (GSIM)*. UNECE Statswiki.

_____ (2015a). *Generic Activity Model for Statistical Organisations (GAMSO)*. Geneva, Switzerland: United Nations.

_____ (2020). *Implementation of the new role of national statistical offices at the time of expanded possibilities* In Conference of European Statisticians, sixty-eighth plenary session, ECE/CES/2020/10.

_____ (2022). *Sustainable Development Goals. Statistics for SDGs Home*. UNECE Statswiki.

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.

_____ (2015b). *Resolution adopted by the General Assembly on 25 September 2015 -- 70/1. Transforming our world: the 2030 Agenda for Sustainable Development*. Seventieth session. A/RES/70/1.

_____ (2022). *Core values of official statistics and their mapping with Fundamental Principles* In Conference of European Statisticians, seventieth plenary session, ECE/CES/2022/2.

United Nations, High Commissioner for Human Rights (2018) *A Human Rights-based Approach to Data*. Geneva, Switzerland.

United States Government, Office of Management and Budget, the CDO Council and the General Services Administration (2020). Federal Data Strategy: Leveraging data as a strategic asset. Federal Data Strategy Action Plan 2020.

Verhulst, S. G. (2021) *Data Stewardship Re-Imagined — Capacities and Competencies*. Presentation to the United Nations World Data Forum 2021.

Wells, D. (n.d.). What is a Data Catalog? The Data Intelligence Blog by Alation.

Wilkinson, M. and others (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018.

Wang, R.Y. and D.M. Strong (1996) Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, 12, 5-33.

* * * * *