DATA GOVERNANCE FRAMEWORK FOR STATISTICAL INTEROPERABILITY (DAFI)

modernstats by HLG - MOS

Data Governance Framework for Statistical Interoperability (DAFI)

Acknowledgement

This document is the final deliverable of the UNECE High-Level Group on Modernisation of Official Statistics (HLG-MOS) project "Data Governance Framework for Statistical Interoperability". The project was selected in 2021 "HLG-MOS Workshop on the Modernisation of Official Statistics" and conducted from 2022 to 2023.

The following team members kindly dedicated their time, and contributed their knowledge, experience, and expertise:

- Juan Muñoz (Project Lead), Silvia Fraustro and Juan Eduardo Rioja – INEGI, Mexico
- Carlo Vaccari – UNECE Project Manager
- Flavio Rizzolo and Chantal Vaillancourt – Statistics Canada
- Zoltán Vereczkei – Hungarian Central Statistical Office
- Muriel Shafir and Debbie Soffer – Israel Central Bureau of Statistics
- Emanuela Reccini and Samanta Pietropaolo – Istat, Italy
- Munthir M. Alansari – Ministry of Tourism, Saudi Arabia
- Daniel Gillman – Bureau of Labor Statistics, USA
- Edgardo Greising – ILO
- David Barraclough and Barbara Ubaldi – OECD
- InKyung Choi – UNECE

Executive Summary

Statistical organisations deal with data coming from different sources and domains. While each information (data and metadata) set possesses intrinsic value on its own, integrating them with other information holds a great potential to provide knowledge and insights to society vital to addressing the increasing number of multi-faceted challenges. Reusing the data sets already collected and produced for other statistical programmes where relevant could also further amplify their values.

Yet, exchanging and making use of data sets across various sources requires a shared understanding among involved parties on several aspects such as data semantics, representation, formatting, and more. These difficulties exist not just for the exchange and sharing between different organisations; they are significant challenges even within the same organisation.

Enhancing statistical interoperability, a capacity to exchange and make use of the statistical information with minimal or no prior communication, is crucial for improving the efficiency and quality from producers' perspectives as well as the usability and value of products for users. Furthermore, is also important to maximise the potential of traditional and new data sources and leverage new technologies such as data science.

Interoperability encompasses multiple facets – semantic, structural, syntactic and system - which are closely related and important for the smooth exchange and utilisation of information. The governance system needed to support and improve interoperability needs to consider various factors, including organisational roles, legal and business policies as well as the standards and technologies that facilitates it.

Moving forward, it is recommended to develop an interoperability strategy within the organisation and establish concrete metrics to evaluate the journey. Moreover, expanding the use of open standards and cultivating a culture of change while supporting staff in acquiring necessary skills and knowledge are pivotal in this endeavour.

Contents

Acronyms

APIs – Standardised Application Programming Interfaces

CDO – Chief Data Officer

CEMs – Common Exchange Models

CIO – Chief Information Officer

CSV – Comma-Separated Values

DCAT – Data Catalog Vocabulary

DDI – Data Documentation Initiative

FAIR – Findable, Accessible, Interoperable, Reusable

GAMSO – Generic Activity Model for Statistical Organization

GSBPM – Generic Statistical Business Process Model

GSIM – Generic Statistical Information Model

INEGI – National Institute of Statistics and Geography of Mexico

ISO – International Organization for Standardization

JSON – JavaScript Object Notation

LOD – Linked Open Data

MAF – Machine Actionable Format

NSO – National Statistical Organisations

OWL – Web Ontology Language

RDF – Resource Description Framework

SEPs – Standardised Exchange Protocols

SDMX – Statistical Data and Metadata Exchange

SKOS – Simple Knowledge Organization System

XKOS – eXtended Knowledge Organization System

XML – eXtensible Markup Language

1. Introduction

1.1. Background

The primary purpose of national statistical organisations (NSOs) is to produce high-quality information to portray the society phenomena as accurately, completely, and timely as possible. Statistical information describes different aspects of the society such as demography, economy, and environment, among others. It is used as input for the design, monitoring, and evaluation of public policies as well as in the making of a wide set of other decisions by private sector and individuals. To create a coherent picture of reality we need an interoperable set of high-quality statistics produced by a set of well-aligned information production processes.

Interoperability is gaining more and more attention due to the increasing complexity of the phenomena that statistical organisations must measure. Multi-faceted policy issues such as climate change adaptation and circular economy, among many others, involve numerous interrelated variables or factors that interact with one another, which most of the time are produced by different programmes in the statistical organisations or independent organisations in the country. On the other hand, statistical organisations have been increasingly exploring the use of big data and new data sources such as satellite images, sensors, and other technologies to meet society's expectations for improved and timely information products. Statistics derived from surveys and censuses could offer an accurate and comprehensive portrayal of society, economy and environment through their systematic data collection approach and rigorous survey methodologies. Additionally, big data represents a great opportunity for statistical organisations to generate information products in near real-time, which could help to have intercensal information or information on topics where data is not available through a traditional survey or a census such as environment statistics. The use of new data whose attributes such as type and source differ from traditional poses additional challenges for statistical organisations as they might not necessarily be interoperable with existing datasets in the organisations. Therefore, data interoperability is a necessary capability to provide a new generation of services and products that meet the emerging demands of statistics users.

This document aims to provide a reference framework that contains the core elements to implement a governance programme focused on achieving data interoperability and thus helping statistical organisations improve their data management.

1.2. Problem Statement

Institutions whose core business is the production and dissemination of statistical information must deal with data and metadata coming from different sources and domains. Each statistical information set has a value by itself, but integration of these individual information assets into a harmonised and interoperable statistical data platform (e.g., data lake) creates a synergy that amplifies the value delivered to society as described above.

However, making use of data sets from different sources (including the internal case of data sets from different units within the same organisation) requires that parties involved should have common understanding on several aspects.

First, knowing the concept to which the data pertains is important. The number corresponding to a statistical indicator represents the measurement of a concept that was defined for that indicator. Logically, we need to be sure that when talking about a specific indicator, we are referring to the same concept. For example, "work" and "occupation" may be used interchangeably in everyday language, but in the labour market, they mean very different things - occupation is a specific form of work. Other forms of work are own-account production work, volunteer work and unpaid internship work. Therefore, if we want to make use of statistical information from different sources, we need to know the concepts that the data obtained from those different sources pertains to.

Once we are sure that the common understanding of concepts is established, the next step involves determining how we will represent the numbers they refer to. To make sure that the number is well interpreted, we need to accompany it with all the information needed to understand the number correctly such as the period it is referring to, the geographic area that is covered, the units of measurement, etc. For example, a population number "126,014" can lead to different interpretations, but only when accompanied with information indicating that it refers to the population of Mexico in thousands, as counted by the National Census in 2020, we can have a proper context and accurate understanding about what the number refers to. Concept (variable), period and geographic dimensions are often considered as the minimum essential information that is needed to determine how the number (measure) can be used or compared to others. Depending on the type of variable, however, other information might be needed to understand the data. For example, if they are foreign trade variables, it will be necessary to know if they refer to an import or an export, the country of origin and the country of destination, etc. On the other hand, if the variables are about a sociodemographic subject, it would be useful to know if they refer to the total population of women or men, and perhaps even the age group to which the data refers. The information that can help us to provide a semantic context of the statistics, as in the last examples described, is called structural metadata. It is needed to ensure that we correctly interpret the numbers when we exchange and make use of them. Some of this metadata can be coded using code lists or classifications to divide the variables into categories and to have a better knowledge of the composition of each indicator. Having a common agreement on how this structural metadata will be incorporated into the information set that is being exchanged will improve the capability to interoperate with the information set.

In the current digital era, the exchange and integration of statistics is done primarily using information technologies. It is easier to achieve technological interoperability if we share the same syntax to conform structures that can be easily interpreted by the different software systems and tools used by the organisations. However, to achieve statistical interoperability it is necessary to consider other aspects.

Based on the descriptions above, we can ascertain that the core conditions to have statistical interoperability include the understanding of the concepts, having a set of structural metadata to provide a context, establishing a regular way that is well known to communicate them, and providing tools and rules to access the information.

From these conditions to have statistical interoperability, we can deduce potential problems that arise in their absence:

- We can put together indicators that refer to concepts that look similar. But if we have not previously agreed on the meaning of those concepts, we will not be sure to confirm that we are referring to the same concepts. In this case, we may not be able to compare these indicators and interoperability cannot be ensured.

- We can make mappings between classifications and transform units to put the statistics in the same contexts. But during this process, we can have problems of losing precision or it may be impossible to map data between different classifications from different parties when they have different granularities or even different conceptualisations.
- We can transform the structures of the statistical information using software. Several organisations use these kinds of tools to integrate information that reduces the effort required to put all this data under the same format, but we must be aware that these transformations can induce some errors that may be difficult to detect. In the end, we cannot guarantee the quality of the product of this transformation that gathers data coming from different sources, initially produced with different purposes and from different points of view, and potentially with undetected errors introduced by our integration processes.
- We can try to use different software tools to create an interoperable data environment, but if they don't share at least some technical specifications such as the capability to receive requests and to answer them directly, the result will add unnecessary complexity and, in many cases, it will require a lot of work to integrate or link the information from one system into others.

The stability and the scope of the agreements that must be made to achieve statistical interoperability are important to set the right conditions to build a statistical information ecosystem able to provide valuable statistical information to the policymakers and in general, to the whole society fitted to satisfy all their needs.

Providing stability means that the concepts, semantics, and structures will be kept during the different cycles of each statistical programme. Having this condition, it will be possible to have time series from different periods, providing information to build models recognising trends and scenarios for forecasting the future trends.

The scope of these agreements is fundamental for eliminating the information silos and building the statistical information platform that could answer the complex multi-dimensional information needs of the society. When a unit in charge of a certain statistical programme takes in isolation all the decisions related to the concepts, semantics, and structures without considering those used by other departments, organisations, or projects, the information produced from the program will not be interoperable, resulting in an information silo. When a division in charge of several statistical programmes establishes interoperability for the statistics produced in their programs, then at least statistics produced by the units within the division will be able to be interoperable. The integrated set of statistics will provide a better understanding of the different concepts within the statistical programs and maybe between different domains.

If the interoperability scope is extended to statistics from different domains, the value of the information will be further. This can help answer complex questions about the interplay among different domains, for example, how the evolution of certain economic activity can affect the demography and ecology of certain geographical areas. The scope can be extended to go beyond statistical organisations and encompass national systems, international, regional, or global. A wider coverage will help society to understand the context of statistics and compare it with the rest of the available areas sharing similar statistics.

Building an interoperable platform of high-quality statistical data cannot be a result of serendipity. It is necessary to establish a data governance programme to transform the data silos into a connected network of harmonised data and metadata sets that includes the structures, procedures, rules, and policies to preserve the meaning and quality of the statistical information datasets it contains.

1.3. Core Terms

Statistical organisations need interoperability to exchange information that can be used by different parties or systems. This process requires that all parties understand the numbers in the same way, ensuring that the information remains unchanged as it is moved and published in different systems and different places. To achieve this, we will need to overcome some challenges:

- Agreeing on the concepts behind the data to be exchanged to ensure that all the parties will have a common understanding of what is being exchanged.
- Establishing the process patterns and constraints that will be followed by the parties to manage, send, and receive the data using the exchange channel in a way which avoids losing or distorting the messages.
- Developing structures to arrange all the data and its related metadata in a way that the statistical information that is exchanged can be easily identified, logically contextualised, accurately integrated, and correctly analysed.
- Providing formats to reduce the diversity and complexity of the tools needed to process and publish different data and metadata sets. This includes defining the main features of the software tools to process and publish the contents of data and metadata sets related to different domains in such a way we can reduce the complexity and cost involved in developing them.

As one can see from above, interoperability can be seen from different points of view and Section 2 describes these different facets, namely, semantic, syntactic, structural, and system, in more detail.

Following this order of ideas, we can define **statistical interoperability** as the capacity to share and make use of statistical information among different parties or electronic systems without distortions of its meaning, not needing to communicate to get additional specifications or make ad-hoc adjustments for each specific case. Statistical interoperability implies achieving minimum compliance regarding the semantical, structural, syntactical, and technological aspects of the statistical data and metadata.

An organisation can achieve interoperability only if it is in control of its information. **Data governance** is defined as the exercise of authority and control (planning, monitoring, and enforcement) over the management of data assets [1]. This concept is related to the decisions that must be made to establish control and be able to manage the data. Its purpose is to ensure that data is managed properly, and according to policies and best practices.

It is important to distinguish data governance from data management. While data governance is about making decisions and establishing lines of authority and expected behaviours, the latter refers to implementing and performing all the aspects of working with data. **Data management** is defined as the development, execution, and supervision of plans, policies, programs, and practices that deliver, control, protect, and enhance the value of data and information assets throughout their lifecycles [1]. Data governance is about ruling data management.

Another related term is **data stewardship,** which is gaining importance as a role of national statistical organisations managing data assets expanding beyond those that are produced by the organisation itself, but data owned and shared by other government agencies or actors in the data ecosystem. In [2], data stewardship is considered an "approach to data governance that formalises accountability for managing information resources on behalf of others" which "is enabled through good data governance and data management."

A framework is a model that describes the structure underlying a system or concept. In this case, a data governance framework is a model that identifies the elements, structure, interactions, processes, and rules required to achieve data governance.

The **Data Governance Framework for Statistical Interoperability (DAFI)** can be defined as a model and a set of guidelines and recommendations that identify the elements, structure, interactions, processes, and rules required to establish the conditions of an information governance environment focused on facilitating the making of decisions required to align the efforts to achieve statistical interoperability.

1.4. Purpose and Scope

The purpose of this document is to provide a point of reference for the discussion regarding the interoperability in the context of statistical organisations' work, and tools that may help to create the conditions inside the organisations to align the different statistical programmes. This alignment, in turn, contributes to improving the capacity of statistical organisations to build a data and metadata platform with interoperability, ultimately enriching the understanding of reality with different pieces of information.

The target audience of this document includes mid-managers and technical experts in national and international statistical organisations who play pivotal roles in critical tasks for establishing and maintaining interoperability within their respective organisations such as data and metadata management, standards, and quality management. Moreover, the audience extends to include other stakeholders such as domain experts and analysts, as establishing interoperability in statistical organisations requires extensive discussions and consensus-building with those who are impacted by these initiatives.

It is important to highlight that the scope of this document is data interoperability. As described in the previous sub-section, data governance includes all decisions and controls related to the proper management such as security, quality, and modelling. While all these elements are crucial for data interoperability and will be mentioned throughout the document where relevant, the primary focus of this work is how to ensure interoperability.

In the statistics field data and metadata have a very strong relation, for this reason, although data and metadata governance reference different concepts, for this document when we talk about data governance, we are referring to statistical information governance, which covers both.

Lastly, it is also essential to underscore that the theme of this document centres around statistical information (data and metadata) which refers to information that statistical organisation acquires for the development, production, and dissemination of their statistics. Statistical organisations regularly use other types of information that play a significant role in their business and benefit from improved interoperability (e.g., human resource, finance, legal). However, these types of information are not in the scope of this document.

2. Interoperability in Statistical Organisations

2.1. Definition and Related Concepts

Definition

One of the early definitions of interoperability defines the concept as an "ability of a system (such as a weapons system) to work with or use the parts or equipment of another system" (Merriam-Webster) which originated from the needs of the military to make parts that can be used interchangeably. As time went on, the term began to be used in information technology in much the same way. The ISO standard ISO/IEC 2382 (Information Technology – Vocabulary) defines interoperability as a "capability to communicate, execute programs, or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units." These definitions have similarities in that they consider interoperability as a capacity or capability. This means interoperability is a condition to be met, not an activity. In other words, it is not an exchange or a function, but it promotes exchange or functionality across systems (see Annex I for more definitions).

In the context of statistical organisations whose core business is the production and dissemination of information (data and metadata), interoperability thus can be considered as a **capacity to exchange and make use of the information with minimal or no prior communication**.

It is important to note that each situation for which interoperability concerns arise needs different characteristics. For different classes of objects used in the official statistics (e.g., variable, data sets, questions, questionnaires, data structures, sampling), the elements required to describe each of those classes are different. For example, a sample has a size, stages, frames, and a selection method at each stage; a question has wording, response choices, and a skip pattern; and a variable has a definition, a value domain, and the data it generates has a format and structure. If we define a technical specification as a schema organising a set of elements, the interoperability of each class (e.g., variable, data structure) depends on the requirements in a schema, for example, if descriptions of variables are to be interoperable, one needs to know the schema used to organise and format those descriptions. Similarly, if some process is interoperable, one needs to know the schema used to organise and describe the steps of the process. Thus, **conformance[1] to the appropriate technical specification is a necessary condition for interoperability**.

Relationship with standards

Standard generally refers to a documented agreement that provides rules and guidelines that are established by consensus or authority. Statistical standards are standards that are related to the production, integration, and dissemination of statistics, including processes, data, products, and services in statistical organisations. They include a set of concepts, definitions, classifications,

---

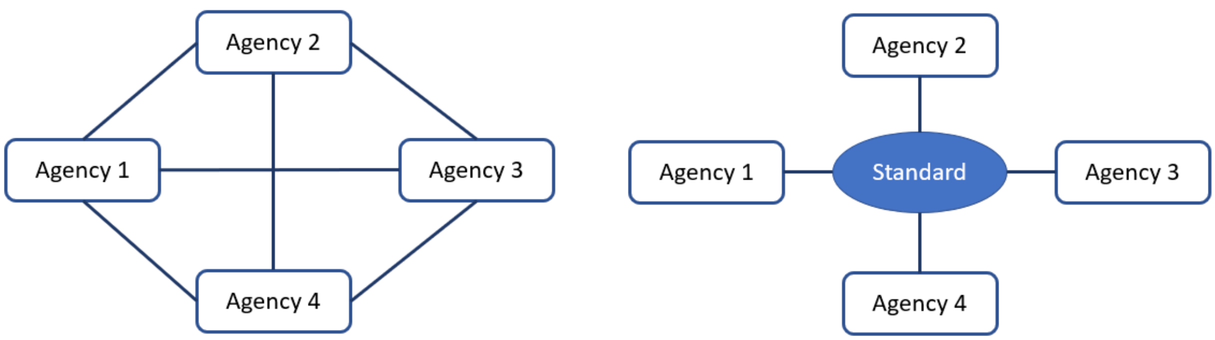
[1] Technical specifications contain normative expressions which can be divided into statement (expression that conveys information), instruction (expression that conveys an action to be performed), recommendation (expression that conveys advice or guidance) and requirement (expression that conveys criteria to be fulfilled). Conformance to a technical specification means satisfying all its requirements.

catalogues, models, methods, procedures that are created and maintained to share, exchange and understand data.

There is a close relationship between standards and interoperability. In principle, two parties can achieve interoperability through bilateral agreements once they agree on every aspect and procedure involved in the exchange (e.g., concepts used in the data, data format, data structure). However, this arrangement quickly becomes costly and inefficient when more parties are included. Another way to achieve interoperability is by making all relevant information open, thus allowing any other party to obtain and understand the data without a need to contact and communicate. However, this also creates inefficiencies as it requires additional efforts if the concepts, structure, or format used by the party is different from those used by other parties that want to make use of the data.

Adopting standards can significantly facilitate interoperability, enabling seamless data exchange not just between individual parties but automatically among any involved parties (see Figure 2.1). Therefore, standards play a crucial role in achieving interoperability efficiently.

Figure 2.1. *Interoperability through bilateral agreements vs. interoperability through adopting standard (recreated based on Figure 5-3 from [3])*



The importance of the following standards is not new for statistical organisations. The Fundamental Principles of Official Statistics states (Principle 9) that "*The use … of international concepts, classifications and methods promotes the consistency and efficiency of statistical systems at all official levels*." However, the significance of adopting standards has grown even more in recent years due to several factors.

Firstly, the landscape of standards has become much more complex. Production processes have become more granular, with each component more specialised – there are different rules, classifications, concepts, models, methods and procedures for different sub-processes and tasks within. The types of data that statistical organisations deal with have become diverse (e.g., geospatial data, unstructured data). To ensure interoperability with other domains, sectors and countries, statistical organisations should consider standards not just within their statistical field, but beyond them.

Also, the use of standards enhances the potential for data to be reused. While a statistical product may have been designed for a specific purpose, the underlying data (final, intermediate, raw) holds value for potential reuse by other programmes. The use of standards also increases the possibility of the data assets to be reused not just for current needs, but for future needs of the organisation.

For examples of standards for interoperability, see Section 3.3.

Relationship with open data and FAIR[2]

Adopting standards are crucial but not sufficient to achieve interoperability efficiently. Once we have implemented standards in our data production, integration or dissemination processes, information needs to be available in formats and through means that make it easily available and accessible to the widest range of users to maximise their value to society. If disseminated data is available without or with minimal restrictions to be reused and redistributed, we can say that this data is open. Besides, if this open data shares common concepts, classifications, or code list, and is distributed by standardised formats and means, i.e., agnostic to a specific language, technology and infrastructure, this data is interoperable with other open data sets produced by different organisations, facilitating the reuse of data that can be integrated, linked, or combined to develop other products and services.

While open data focuses on the non-restricted, freely use and share of data for anyone, interoperability facilitates the integration and linking of different sources, including open data. Therefore, interoperability and open data are both essential to reuse data, and are aligned with the FAIR (Findable, Accessible, Interoperable, Reusable) principles, which are a set of guidelines to enhance the reusability of data by both humans and machines.

As the Fundamental Principles of Official Statistics point out, high-quality official statistics plays a critical role for the analysis and decision making for many social benefits, like the mutual knowledge and exchange of data among the States and society, demanding openness, and transparency.

Open data is data that anyone can access, use and share [4]. This means that data should be freely available for use and reuse by others with no restriction, unless explicit restrictions for protection of personal data, confidentiality or property rights exist. Nowadays, open data is considered the most decisive approach to enhance data reuse by other actors to create value [5].

Many efforts around the world have been made to disseminate data through open data policies, from federal government legislations to private data exchange initiatives. However, these policies and initiatives must be supported by protocols for safeguarding confidentiality, interoperable technical standards, machine readable formats and open (user-friendly) licensing to facilitate further reusability of data, including aggregated data and microdata.

Regarding the FAIR principles, to use and reuse data, data first should be **findable**. Open data should be easy to find by both humans and computers. While using common classifications for different data domains helps humans to find data, machine readable metadata enables the automation of discovering datasets and services.

Facilitating **access** to data among different actors of a data ecosystem enables the acquisition of the full social and economic value of data. Open data regulations and technical standards can facilitate access to data, making it ready to be used, integrated, linked, and repurposed. It is recommended to align access to information and open data. To achieve this, legal frameworks could make data dissemination and nonpersonal data open by default to enable reuse and

[2] See FAIR principles overview at https://www.go-fair.org/fair-principles/

redistribution. Open access to data is crucial to achieve the benefits of widespread data use, reuse, and repurposing.

To get the most value of data, official statistics must ensure that data can be used more effectively by integrating or linking datasets. Hence, there is a need to define governance rules of data and metadata that ensure aspects like quality, common structures and means of the data to be disseminated or exchanged. **Interoperability**, as stated before, can be supported by ideally open standards, which usually are determined collaboratively by sectoral or international organisations with common needs. The adoption of common classifications, formats and tools facilitates sharing, integrating, and linking data between stakeholders. Open data[3] and interoperability foster the flow of data between participants of national data systems and enable cross-border data collaboration [5].

The use of open-source technologies contributes to reduce costs and facilitates adaptation to different business needs. It is considered a good practice to use open-source technologies whenever possible because it helps to the **reusability** of data and tools. Open data can be reused for research, design, evaluation of public policies, innovation, and development of different domain organisations.

2.2. Facets of Interoperability

In this document, we cover four key facets of interoperability, namely semantic, structural, syntactic and system interoperability, which collectively form a foundation that allows exchange information in an effective and efficient way. For illustration, let us consider a data set on the population dynamics which are stored in three different forms (i.e., table, CSV, and JSON):

Example 1: in table

Country	Country code	Region	Population 2000	Population 2022	Average annual population growth	Currency unit
United Kingdom	GB	Europe & Central Asia	58.9	67	0.6	Pound sterling
Canada	CA	North America	30.7	38.9	1.1	Canadian dollar
…	…	…	…	…	…	…

[3] For Open Data see "Open Data for Official Statistics: History, Principles, and Implementation a review on the principles and implementations of open data in official statistics" at https://opendatawatch.com/publications/open-data-for-official-statistics-history-principles-and-implentation/

Example 2: in CSV	**Example 3: in JSON**

Country, Country code, Region, Population 2000, Population 2022, Average annual population growth, Currency unit
“United Kingdom”, “GB”, “Europe & Central Asia”, 58.9, 67, 0.6, “Pound sterling”
“Canada”, “CA”, “North America”, 30.7, 38.9, 1.1, “Canadian dollar”
.... ``
	[
	{
	"Country": "United Kingdom",
	"Country code": “GB”,
	"Region": “Europe & Central Asia”,
	"Population 2000": 58.9,
	"Population 2022": 67,
	"Average annual population growth ": 0.6,
	"Currency unit ": Pound sterling,
	
	},
	{
	"Country ": "Canada",
	"Country code": “CA”,
	"Region": “North America”,
	"Population 2000": 30.7,
	"Population 2022": 38.9,
	"Average annual population growth ": 1.1,
	"Currency unit ": Canadian dollar,
	}
	....
	]

Semantic interoperability ensures that the information exchanged is interpreted meaningfully and accurately. It involves using standards or establishing a common set of elements needed to understand meaning such as concepts, vocabularies, classification, or code list. For example, we should be able to understand the meaning of “population” (concept) accurately as well as the meaning of its values “58.9” (millions); we need to know what “country” means and how it is represented in the data (e.g., Alpha-2-Code, Alpha-3-Code). Furthermore, to reuse the data, one might also require additional information such as when the data was obtained or what the geographic coverage was to properly contextualise the data.

Structural interoperability concerns the structure and hierarchy of information exchanged. With structural interoperability, we can understand what the values are, what are the variables, and what values are for what variables (e.g., to find the currency unit of a country, we would need to look at where the variables are and where the currencies are). In the CSV example, the first row represents the variables, and the rest of the rows are for data records; in the JSON file, each item in the inner array is paired with variable names and values in a nested structure.

Syntactic interoperability concerns the structure and syntactic consistency needed to effectively communicate [6]. It involves a common data format and common protocol to structure any data, thus the manner of processing the information will be interpretable from the structure. As an example, we might conform the data structure to specific standards to describe the format, like SGML, JSON, SDMX-ML, etc.

System interoperability, also known as technical or technological interoperability, concerns the connectivity, communication, and operation of the interacting entities, and middleware elements regarding authentication and authorisation, the use of technical standards, protocols for communication and transport, and interfaces between components required to facilitate the interaction between different systems, ensuring they can operate collaboratively. It covers the applications and infrastructures linking systems and services. It includes interface specifications, interconnection and data integration services, data presentation and exchange, and secure communication protocols.

Theoretically, it is not impossible to have one facet of interoperability without others and one may choose to focus solely on any individual facets of interoperability. However, the four facets are closely related, and the four facets are needed to exchange and make use of information smoothly. For example, different survey programmes could agree to use the same definition and code list for "economic activity" which leads to semantic interoperability, but if data sets across different programmes are still structured, stored, and encoded in different ways, the exchange, sharing and re-use of data sets would require additional mapping, transformation, and communication.

Box 2.1. Harmonisation of concept "Turnover" – Istat experience

ISTAT is currently working on the creation of a single terminology collection, which, overcoming sub-domain glossaries, centrally gathers all terminological and semantic resources, The collection can be used by various users. Standard metadata are the result of joint technical roundtable set up in ISTAT and involving researchers both in the domains and transversal structures, aimed at the integration of processes or ontologies, by sharing definitions, updating, and harmonising terms. One example concerns "Turnover": this concept, pertaining to the business statistics domain, is widely used in ISTAT. From the initial seven definitions, corresponding to as many production processes (business accounts, structural and short-term statistics on businesses, transport and telecommunications, tourist facilities, wholesale, and retail trade, etc.), the discussion within the joint table led to the wording of two harmonised definitions, reflecting the EU legislation regulating the two macro production processes on structural and short-term business statistics (EC 250/2009 and 1503/2006). An ad hoc definition was formulated for retail trade statistics due to the peculiarities of the collection process. The solicitation for terminological harmonisation may therefore come from international authorities. For example, the recent EU Regulation 2020/1197 provides definitions of concepts and variables, including "Net Turnover", which brings together under a single lemma and a single definition what had up to now been declined in several terms.

2.3. Benefits of Interoperability

Interoperability can offer many benefits for several reasons. The effective sharing and communication of data, information and knowledge among stakeholders is essential to maximise the value of data and make more evidence-based decisions in society. In the following, we explore the benefits from the perspectives of two different roles: the user, i.e., organisations or people who use data, and the producer, i.e., organisations that produce the data.

Firstly, if data is disseminated using common concepts between different domains and follows metadata standards, it allows users to locate data efficiently and promotes a more accurate use of data, providing the context and quality aspects of data sets to reduce the risk of misunderstandings and misuse. If conceptual and technical standards are implemented, data assets are more interoperable which makes it easier to find, integrate, access, link, share and/or analyse data from different sources. Besides, the use of metadata standards enables automation and good communication between different computer systems or applications.

Frequently, large part of research time of analysts is spent searching for data and executing the transformations required to integrate it with other sources. Conceptual and technical standards improve speed, efficiency, and consistency of research process, facilitating the comprehension of data and eliminating potential errors caused for non-compatible terms. Thus, interoperability enables users to better understand terms and concepts in data obtained from different sources and domains, allowing new ways of gaining insights to solve the ever-increasing data challenges that society faces.

From the producer's side, interoperability improves productivity and efficiency with the reuse of data, methodologies, tools and enables the quick access to data and information. Also, the establishment of a common language improves the quality of production processes. For example, the automation of processes to collect, integrate, process, or classify statistical data reduces the potential for human error, promoting the production of high-quality data and a better decision management.

Interoperability also plays a critical role to reduce costs and improve the quality of statistics for producers. For example, with increased data sharing and reuse of applications among stakeholders, the logical integration using common identifiers reduces redundancy and unnecessary storage expenses.

2.4. Source of non-interoperability

According to the Value Chain Analysis which conceptualise activities in an organisation, two levels of activities exist in organisations: primary activities and supporting activities. Within statistical organisations, the primary activity pertains to production activities, typically represented with the Generic Statistical Business Process Model (GSBPM). The non-Production Activity Areas of the Generic Activity Model for Statistical Organisation (GAMSO), particularly Corporate Support, can be considered as supporting activities.

Given the interoperability is by nature a cross-cutting exercise encompassing different programmes and organisational units, the standardisation activities at the corporate level are of great importance to ensure the effective coordination of various classes used in the production of statistics (e.g., code list, classification, methodology, quality indicators). The role of coordination can be carried out by a central unit or permanent committee mandated to establish, maintain, and promote the standards, or by a more loosely organised mechanism, for example, regular meetings among the stakeholders.

It is also important to ensure the interoperability throughout the production process as decisions made during production activities could either weaken or enforce interoperability. Failing to consider the interoperability perspectives and take appropriate action could not only diminishes the potential value of the information produced but can also introduce inefficiencies during the production process. Analysis below demonstrates various sources of non-interoperability according to GSBPM Phases.

1. **Specify Needs Phase**: "this phase is triggered when a need for new statistics is identified or feedback about current statistics initiates a review" [7]. The phase includes the investigation on the practices among other national and international statistical organisations producing similar data and checking availability of existing data resources (sub-process 1.1, 1.2). The lack of research in primary concepts, code lists, classifications

and tools used in these data resource may lead to missing opportunities to increase interoperability of statistics intended to be produced to be interoperable with them. Users of the statistics produced may prioritise meeting their needs as exactly intended without considering interoperability point of view. Therefore, during consultation with users and stakeholders, the need for alignment of concepts and output needs to be communicated. When this phase is initiated to review and update existing statistical programmes, it is important to assess the impact of such changes with respect to interoperability.

2. **Design Phase**: "this phase includes the development and design activities, and any associated practical research work needed to define the statistical outputs, concepts, methodologies, collection instruments and operational processes". Design Phase plays a critical role not only in ensuring interoperability across the instance of production process but also in facilitating the overall interoperability of final statistics and any artefacts produced across the organisation. Creating variable, value domains or classifications only slightly different from existing ones just to meet the immediate needs (sub-process 2.2) would negatively impact the interoperability. Classes that could be consulted with the central repository or metadata system include conceptual classes (e.g., variable, value domain, classification, unit types) as well as those that are related to the exchange (e.g., data format, questionnaire, question statements, legal agreements, license). Given that metadata is critical to understand and make use of any data set, a lack of standardisation of the way metadata is captured and modelled in different stages will have detrimental impact on the interoperability.

3. **Build Phase**: "this phase builds and tests the production solution to the point where it is ready for use in the 'live' environment". While many design decisions are made during the Design Phase, there are several choices made at the implementation stage that could impact interoperability. For example, specific data collection systems might use different data formats or encoding, which could be exacerbated when multiple data collection modes are involved in the process. It is imperative that data dissemination methods, such as those involving APIs, are thoroughly documented to facilitate interoperability and efficient data sharing.

4. **Collect/Acquire Phase**: "this collects or gathers all necessary information (e.g., data, metadata and paradata), using different collection modes (e.g., acquisition, collection, extraction, transfer), and loads them into the appropriate environment for further processing". With statistical organisations increasingly involved with sources that are not under their direct control (e.g., administrative data, big data from the web), ensuring interoperability becomes even more challenging. Without proper documentation of the data and mapping (e.g., between code lists used by different sources), the risk of introducing non-interoperable elements during this phase significantly increases.

5. **Process Phase:** this phase "describes the processing of input data and their preparation for analysis". In this phase, various processes are applied to data and lack of availability of data provenance information would lead to non-interoperability. For example, data integration from different sources would need more mapping and transformation processes if the data sets do not share common concepts or classification. Besides, if the classification or code list associated to the variables collected are not common between different programmes, it requires validation and edition rules at each iteration, increasing risk of error and mistakes.

6. **Analyse Phase:** “in this phase, statistical outputs are produced and examined in detail”. If the processed data files are not interoperable, comparing statistics with previous cycles of the same programme or other related data would be difficult. If the concepts are not interoperable, comparisons may be even impossible. There will be additional efforts needed for carrying out in-depth statistical analyses such as time-series analysis, consistency, and comparability analysis when concepts, classifications and code lists are different.

7. **Disseminate Phase:** “this phase manages the release of the statistical products to users”. Non-interoperable data sets are more difficult to prepare and put into output systems, because formatting the data and metadata in a manual or semi-automated way is prone to error. The lack of a common classification for different domains hinders the user’s information discovery. For example, use of common standard such as the Classification of Statistical Activities (CSA) [8] could help classify information about statistical activities, data, and products by providing a top-level structure to make it easier to find information about different domains, such as demographic, economic and environment statistics.

3. DAFI Components

This section lists key elements that are important in achieving interoperability in statistical organisations, focusing on the factors that help achieve the desired interoperability. These components encompass organisational roles, legal and business policies that influence interoperability, the standards and technologies that facilitates it.

3.1. Roles and Governance Bodies

Interoperability issues within NSOs typically involve several roles and stakeholders as well as various governance bodies. It is important to note that the specific roles and responsibilities may vary depending on the structure, size, and priorities of the NSO. Some organisations may have dedicated teams or units focused on interoperability, while others may not have a dedicated unit, and distribute the responsibilities among existing roles and units. However, even if there are no dedicated roles, the functions performed by the roles mentioned can still be carried out by other roles involved in the statistical production process.

Here are some of the key roles that may be involved:

- *Chief Data Officer:* A chief data officer (CDO) is the manager dedicated to the organisation data strategy: he/she is responsible for the utilisation and governance of data across the organisation. A CDO is a senior executive who drives growth by following a data-driven approach.

- *Chief Information Officer:* A chief information officer (CIO) is the high-ranking executive responsible for the management, implementation, and usability of information and computer technologies systems of an organisation. A CIO oversees the maintenance and of the internal technology processes as a way of maximising organisation productivity and making complex tasks more achievable through automation. To navigate through continually changing landscapes, a CIO needs a diverse skillset in terms of leadership, communication ability, etc.

- *Data Governance Manager:* Data governance managers are responsible for implementing and managing the data governance framework, policies, and procedures. They work closely with various departments to ensure compliance and adherence to data governance principles.

- *Data Stewards:* Data stewards are responsible for specific sets of data and ensure the quality, accuracy, and integrity of the data, as well as its compliance with data governance policies. Internally in the organisations, data stewards are often subject matter experts in specific domains (e.g., business statistics, health statistics).

- *Data Architects*: Data architects design and develop the organisation's data architecture to support interoperability. They create data models and structures that facilitate seamless data exchange between systems.

- *IT Managers*: IT managers play a crucial role in ensuring that technical systems and infrastructure to support data interoperability. They oversee the implementation of data integration solutions and manage the data exchange processes.

- *Privacy and Compliance Officers:* These individuals are responsible for ensuring that data governance practices comply with relevant privacy regulations and legal requirements. They help manage data access, usage, and consent mechanisms to safeguard sensitive information.

- *Business Analysts:* Business analysts bridge the gap between technical teams and business users. They help define data requirements, identify data sources, and assess data quality to support interoperability initiatives.

- *Data Consumers:* These are the end-users or departments that utilise the data for decision-making and operational purposes. Data consumers play a vital role in providing feedback on data quality and ensuring that data meets their specific needs.

- *Methodologists:* Methodologists may be involved in drafting/approving data governance guidelines based on methodology best practice.

- *Statistical Standard Experts:* These experts support statistical program areas in all matters related to the development, use or implementation of statistical standards, which are key to interoperability. This could be provided in the form of supporting the development of standard concepts and value domains, e.g., classifications, definitions, etc., following established principles. They could also provide support in the use of standard models which allow the proper capture and management of metadata used to describe data.

- *Publishing Staff:* Publishing staff validates data and metadata for publication readiness. They may be involved in drafting/approving data governance guidelines. May apply to primary or secondary data.

Various governance bodies can manage interoperability issues to ensure the smooth exchange and integration of data. These bodies often oversee the implementation of standards and protocols to promote data consistency and coherence. Some governance bodies include:

- *Data Management Committee:* oversees the management and coordination of data-related activities within the NSO. It can play a key role in setting standards for data collection, storage, and dissemination, ensuring interoperability across different departments and systems. [4]

- *Standards and Methodologies Board*: focuses on establishing and maintaining standards and methodologies for data collection, processing, and analysis. It ensures that data is collected and managed using consistent and reliable methods, enabling interoperability across various statistical domains. [5]

- *Information Technology Steering Committee*: is responsible for guiding the overall IT strategy within the NSO. It oversees the implementation of IT systems and infrastructure

[4] See "Report of the Committee on Data Management," MOSPI NSO (India) at https://mospi.gov.in/sites/default/files/committee_reports/finalreportonDatamanagement01082011.pdf
[5] See ONS Statistical Quality Improvement Strategy at https://www.ons.gov.uk/methodology/methodologytopicsandstatisticalconcepts/qualityinofficialstatistics/onsstatisticalqualityimprovementstrategy

to support data management and interoperability, including decisions on the adoption of standardised technologies and platforms for data exchange and integration.

- *Data Quality Assurance Board*: with the task of monitoring and ensuring the quality of data produced and disseminated by the NSO. It establishes protocols and procedures for data validation, verification, and quality control to maintain high data standards and promote interoperability among different datasets.
- *Inter-agency Data Sharing Task Force*: facilitates collaboration and data sharing among various government agencies and departments. It works to establish data-sharing agreements, protocols, and mechanisms that promote interoperability and seamless data exchange between different entities.[6]

In the Annex 3, we listed the roles and responsibilities taken from ISO/IEC 11179, an international standard for representing, storing, and maintaining metadata in a metadata registry. One section is specifically devoted to the roles associated with the metadata registry. While addressing the semantics of data, the representation of data and the registration of the descriptions of that data, ISO/IEC 11179 intends to promote harmonization and standardisation of data and metadata and its re-use within an organisation and across organisations.

Box 3.1. Interoperability function in INEGI

INEGI has a high-level independent area at the same level of those producing statistical and geographical information. This area is responsible for policies, programs and strategies regarding government and information architecture; security and confidentiality; quality assurance; and interoperability of statistical and geographical information. They are looking forward to the standardisation of processes, the systematic evaluation of processes, and the continuous improvement, thereby consolidating the data ecosystem produced and managed by INEGI. Likewise, this area has support from the highest level of the NSO, who constitute the high-level governance body and participate in a Quality Assurance Committee and in a Security and Confidentiality Committee. In these committees, the regulatory and administrative provisions, that the information production processes must follow, are decided; for example: the adoption of models and standards such as MPEG (which is an adaptation of the GSBPM) and its PTracking tool to implement it, as well as SDMX, CSPA and others. This high-level governance body takes an active role to ensure the change of old practices that work but create inefficiencies in the form of information silos. They decided not just the adoption of models and standards to satisfy specific needs, but the design of a whole environment to support the transition from the existing silos of data and metadata to an integrated statistical and technological environment, through the implementation of four transversal ICT platforms (data, systems, services, and ICT infrastructure) which support the entire lifecycle of Official Statistics.

3.2. Legal and Business Policy

Interoperability is essential for facilitating seamless data exchange and collaboration between different systems, organisations, and departments. Legal and policies can have a significant impact on improving interoperability as they ensure and foster adherence to common standards

[6] See "Data Sharing Working Group" Recommendations at https://resources.data.gov/assets/documents/2021_DSWG_Recommendations_and_Findings_508.pdf

and practices across diverse entities (both among different organisations and within the organisation).

In each country, there exists a Statistics Law that outlines the roles of NSOs and mandates their activities. Some of these laws may include provisions that imply a responsibility for NSOs to actively engage in and contribute to interoperability efforts (see Box 3.2 and Box 3.3 for examples from Canada and Mexico respectively).

With a growing recognition on the importance of data driving economy and improving quality of public sector services, there are increasing trend in developing centralised platform for the provision of data from public sector which often accompanied by legal decisions and data policies at the national level. Initiatives such as open government data strategy, national data strategy also pushes further for a more data governance and management across the society. For example, México's Strategic Program of the National System of Statistical and geographical information 2022 – 2046 stablishes a specific goal and a general action related to the consolidation of the interoperability between statistical and geographical programmes (www.snieg.mx/Documentos/Programas/PESNIEG_2022-2046.pdf).

In contrast to other public entities where data is often a secondary by-product, the core business of NSO is about data. With its substantial methodological and technical expertise, coupled with long proven history of managing data at scale, NSOs naturally find themselves taking on a crucial role in enhancing interoperability in the public sector broadly (see also Box 3.4 in the following section for example from Italy).

Box 3.2. Semantic interoperability as a legal requirement: Canada's Statistics Act

3. There shall continue to be a statistics bureau under the Minister, to be known as Statistics Canada, the duties of which are

> *to collect, compile, analyse, abstract, and publish statistical information relating to the commercial, industrial, financial, social, economic, and general activities and condition of the people;*
>
> *[...]*
>
> *generally, to promote and develop integrated social and economic statistics pertaining to the whole of Canada and to each of the provinces thereof and to coordinate plans for the integration of those statistics.*

Source: Statistics Act (justice.gc.ca)

Box 3.3. Information Infrastructure to achieve interoperability: Mexico's National System of Statistical and Geographical Information Law

The Law of the National System of Statistical and Geographical Information (SNIEG, by its Spanish acronym) promulgated in 2008 defines the minimum information infrastructure for each national information subsystem: demographic and social; economic; geographical and environmental; and government, public safety, and justice.

The Information Infrastructure is the set of data and methodologies that support the information production process to facilitate its interoperability and it is made up of catalogues, classifications, statistical and geographical registries, and methodologies. The use of a common Information Infrastructure facilitates the integration or linkage of information from different statistical and geographical production processes.

Sources: https:/www.diputados.gob.mx/LeyesBiblio/ref/lsnieg.htm

www.snieg.mx/Documentos/Programas/PESNIEG_2022-2046.pdf

On top of laws and initiatives at the national level, here's how legal and business policies can contribute to enhancing interoperability in NSOs:

1. Standardised data formats and protocols: Implementing policies that mandate the use of standardised data formats and protocols promotes uniformity in data representation, enabling smooth data exchange and integration within the NSO and among data providers.
2. Data sharing agreements and contracts: Establishing clear agreements that govern data sharing between various government agencies and external partners can facilitate the secure and efficient sharing of data while ensuring compliance with data protection regulations and confidentiality requirements.
3. Open data policies guidelines: Implementing open data policies guidelines encourages the responsible sharing of non-sensitive data with the public and businesses, fostering transparency, innovation, and economic development while safeguarding data privacy and confidentiality.
4. Data governance frameworks and business process integration: Developing comprehensive data governance frameworks and aligning business processes with interoperability standards can streamline data management practices and improve the compatibility and consistency of data across different systems and departments within the NSO (see Box 3.5. for the example of data governance framework from Israel).
5. Compliance with industry standards and best practices: Aligning legal and business policies with industry standards and best practices, such as those recommended by international organisations and data governance authorities, promotes the adoption of interoperable technologies and practices, facilitating data integration and exchange at both national and international levels.

By incorporating these legal and business policies, NSOs can enhance their interoperability capabilities, ultimately contributing to the overall development and advancement of national statistics and data governance.

3.3. Standards, Tools, and Technologies

Standards

The use of standards is key to ensure all types of interoperability as described in Section 2.

Standards are a set of agreed-upon and documented guidelines, specifications, accepted practices, technical requirements, or terminologies for diverse fields. They can be mandatory or voluntary and are distinct from acts, regulations, and codes, although standards can be referenced in those legal instruments.

In the world of statistics, we can think of **statistical standards**, which are standards about all aspects of the statistical production, either processes/capabilities or the data/metadata they use.

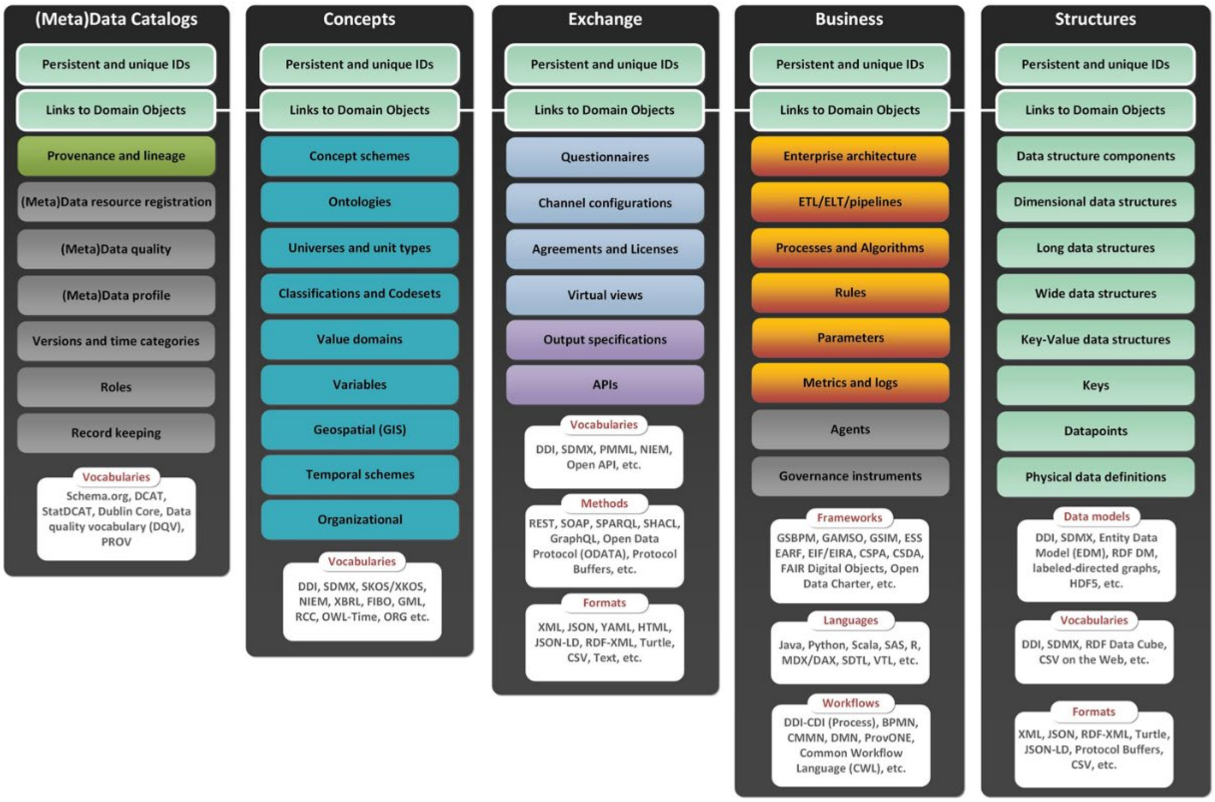
A **statistical data and metadata standard** is a statistical standard about how data and metadata are managed, organised, represented, or formatted. This includes information about processes (designs and plans of statistical programmes and each step in the statistical process), capabilities to produce statistics, data, and metadata itself, the meaning of data and the terms used in relation to data and its structure. 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.

Examples of statistical data and metadata standards:

- Statistical Data and Metadata eXchange (SDMX)
- JSON, CSV, XML, and other standard recommended data formats
- Data Documentation Initiative (DDI)
- Data Catalog Vocabulary (DCAT)
- Generic Statistical Information Model (GSIM)
- Generic Statistical Business Process Model (GSBPM)
- ISO 11179 - Information technology —Metadata registries

Applicability of standards within the statistical processes

For interoperability purposes, standards need to be applied to information that is required to advance the statistical processes. The use of the GSIM groups of concepts, exchange, business, and structures, tied together with (meta)data catalogues for management purposes, provides a framework in which multiple information domains and the applicable standards can be described. The following image provides this overview:



We must remind the meaning of the four GSIM groups:

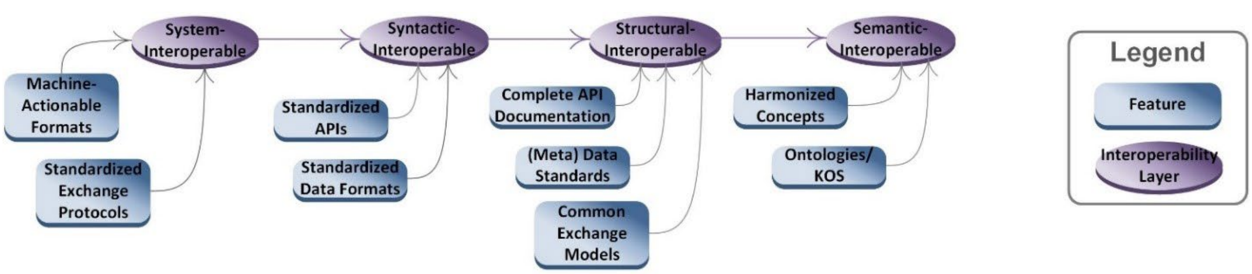
- The **Concept** Group is used to define the meaning of data, providing an understanding of what the data are measuring;
- The **Exchange** Group is used to catalogue the information that comes in and out of a statistical organisation via *Exchange Instruments*. It includes information objects that describe the collection and dissemination of information;
- The **Business Group** is used to capture the designs and plans of *Statistical Programs*, and the processes that are undertaken to deliver those programs. This includes the identification of a *Statistical Need*, the *Business Processes* that comprise the *Statistical Program* and the *Assessment* of them;
- The **Structure Group** is used to structure information throughout the statistical business process.

For (Meta)Data Catalogues and for the four GSIM groups outlined above, there are several tools/standards that can be used to describe the information in a consistent manner, starting from Vocabularies to Methods, Formats, Frameworks, Languages, Workflows and Data models.

- **Vocabularies** are organised collection of terms and relationships used to describe one or more domains pertinent to the production of statistics.
- **Methods** are standard technical means of accessing and exchanging data and metadata.
- **Formats** are physical representations for data and metadata.
- **Frameworks** are compendium of principles, reference architectures, best practices and high-level documentation intended to inform the production of statistics.
- **Languages** are programming languages for the validation, analysis, processing and transformation of data and metadata.
- **Workflows** are standard process models that capture data and metadata processing at different levels of detail.
- **Data models** are standard structure specifications for the representation of data and metadata.

Enablers for interoperability

The following schema highlights the most important enablers that allow you to more easily reach the different facets (layers) of interoperability seen in chapter 1. Here we report both the image and a brief description of the enablers, including some enablers absent from this scheme, but generally useful for achieving interoperability.



For system interoperability:

- Use Machine Actionable Formats (MAFs) - MAFs are designed to make it easier for machines to access, share, and use data, their benefits:
    - MAFs can help to **increase automation** by making it easier for machines to perform tasks such as data extraction, cleaning, and analysis. This can free up NSO staff to focus on more strategic and value-added tasks.
    - MAFs can help to **improve data governance** by making it easier to track and manage the use of data, helping to ensure that data is used in a responsible and ethical manner.
    - MAFs can help to **enhance data security** by making it easier to encrypt and protect data, mitigating the risk of data breaches and other security threats.

- Use Standardised Exchange Protocols (SEPs) - are protocols that define how data should be exchanged between different systems and technologies. In official statistics, the most relevant example is SDMX, but also JSON and CSV. SEPs benefits:
    - SEPs can help to **reduce the costs** associated with data dissemination and use by making it easier to share and reuse data.
    - SEPs can help to **improve data quality** by reducing the risk of data duplication and errors.
    - SEPs can help to **increase transparency and accountability** by making it easier to track the provenance and usage of data.

For syntactic interoperability:

- **System** interoperability (see above)
- Standardised APIs: Standardised Application Programming Interfaces (APIs) can play a crucial role in improving interoperability in official statistics by providing a uniform and consistent method for different systems to communicate and exchange data. Some examples of standardised APIs that are relevant to official statistics include **SDMX-API**, for the exchange of statistical data and metadata in accordance with the SDMX standard, but also **CKAN API**, used in the publication of Open Data.
- Standardised formats: standardised formats are powerful tools that can be used to improve interoperability in official statistics. By using standardised formats, NSOs can make their data more accessible and usable to everyone. Some examples of standardised formats: **CSV, JSON, SDMX**; In addition to the above, NSOs can use standardised formats to publish data in different types of media, such as tables, charts, and maps. NSOs can also use standardised formats to publish data in different languages.

For structural interoperability:

- **Syntactic** interoperability (see above)
- (Meta)Data standards: data and metadata standards are instrumental in improving interoperability in official statistics by establishing common formats, structures, and definitions for data and accompanying metadata. They facilitate the seamless exchange, integration, and understanding of data across different systems and organisations. Examples of metadata standards that are relevant to official statistics:
    - **SDMX:** SDMX standard is a set of international standards for the exchange of statistical data and metadata. It is used by NSOs around the world to publish and exchange data.
    - **DCAT:** DCAT is a standard for describing datasets. It is often used to describe NSO data because it is designed specifically for datasets.
    - **DDI**: is an international standard for describing surveys, questionnaires, statistical data files, and social sciences study-level information.
    - **GSIM**: the Generic Statistical Information Model is a reference framework for describing the information objects that are used in the production of official statistics. GSIM is often used as a conceptual model including all Metadata needed to describe the statistical processes.
- Common Exchange Models: Common Exchange Models (CEMs) can significantly enhance interoperability in official statistics by providing standardised formats for the exchange of data between different statistical systems. These models ensure that data can be

seamlessly shared and interpreted across various statistical agencies, facilitating the integration and comparison of data from different sources.

- Complete APIs Documentation: complete API documentation is a comprehensive set of documents that explains how to use an API. This is an essential tool to use APIs effectively: by providing complete API documentation, NSOs can make it easier for their staff and users to understand and use APIs. This can lead to improved data collection, processing, and dissemination.

For semantic interoperability:

- **Structural** interoperability (see above)
- Harmonised concepts: Harmonised concepts play a crucial role in improving interoperability in official statistics by ensuring that statistical data is consistently defined and interpreted across different systems and organisations. Harmonisation involves establishing a shared understanding of key concepts and variables, enabling seamless data exchange and comparison.
- Ontologies/KOS: Ontologies and Knowledge Organization Systems (KOS) improve interoperability in official statistics by providing structured vocabularies and frameworks for organising and categorising data and knowledge. These systems help establish common conceptualisations and relationships between data elements, enabling more effective data integration, sharing, and analysis.

See Annex 3 for more details on these standards as well as the existing applications which are based on these standards.

Box 3.4. The Italy case

Semantic interoperability: the activities that Istat is carrying out in the context of the National Recovery and Resilience Plan

One of the investments foreseen by the EU National Recovery and Resilience Plan (PNRR in Italian) includes the creation of the "National Digital Data Platform" (Piattaforma Digitale Nazionale Dati, PDND) that will enable the exchange of information between public administrations and will promote the interoperability of information systems and public databases. The creation of PDND will be accompanied by a project aimed at guaranteeing Italy's full participation in the European initiative of the Single Digital Gateway (SDG), which will allow harmonisation among all Member States and the complete digitalisation of a set of procedures/services of relevance (e.g., request for a birth certificate, etc.). Inside PDND at https://www.interop.pagopa.it/ you can find many documents, among them one containing "Infrastructure Guidelines technology of the PDND for the interoperability of information systems and databases."

National Data Catalogue

Istat, in collaboration with the Department for digital transformation, activated the National Data Catalogue on 30 June 2022 for the semantic interoperability of the information systems of public administrations. The Institute has achieved its first objective on schedule by publishing the access portal to the National Data Catalogue at the link www.schema.gov.it, which will make it possible to develop and increase interoperability between data of national interest.

The investment involves the creation of a National Data Catalogue, with the aim of providing a common model and standard and promoting the exchange, harmonisation and understanding of information between public administrations, within the context of the National Digital Data Platform. The Catalogue will make available controlled vocabularies and classifications capable of making access to different information bases more functional.

To manage the project, the establishment of an Implementation Committee for the governance and direction of the agreement is envisaged, in which the Department for Digital Transformation at the Presidency of the Council of Ministers and Istat participate, but which is also open to other possible public entities, such as the Agency for Digital Italy (AgID) and National Research Council (CNR). For the development of the project plan, which provides for a budget of 10.7 million euros, an important commitment of highly skilled human resources is required, to be recruited through new hires. For Istat, specifically, the selection of a contingent of up to 25 full-time people is envisaged, with technical, thematic, methodological, and legal skills.

METAstat: the new Istat Metadata System

The importance of interoperability, which can be pursued primarily using statistical models and standards existing at an international level (GSBPM, GSIM, etc.), has clearly emerged in the context of the National Data Catalogue. At Istat level, the contribution to interoperability can be achieved by having two fundamental infrastructures available: a complete and transversal metadata system together with ontologies and controlled vocabularies.

Istat is currently working on the creation of METAstat, the new institutional system for the documentation of metadata, processes, and statistical products. It will consist in three core modules (controlled terminology collection; structural metadata; referential metadata), currently independent one from the others. It will integrate the different Istat systems containing data (and consequently metadata), with the aim not only of improving their performance and aligning them in their common aspects but also of adding the necessary functions to assist production processes to the current documentary aspects. Indeed, METAstat is designed not to be a passive catalogue of metadata, to be fed ex post, but must have an active role in providing production services with the concepts (represented by metadata) on which to structure the data to be produced (metadata driven). It will enter the production processes already in the design phase of the survey and will have to be integrated into the production processes. METAstat is intended to provide an active support to simplify and automate production processes, as well as to increase the reliability, consistency and timeliness of the data produced (quality principles). In this way the sharing (internal or external) of the data produced will be simplified and facilitated, because these data are already structured on shared and certified metadata from birth.

Both National Data Catalogue and METAstat need to manage the semantic heritage common to administrative procedures and statistical production processes, respectively.

Consultation, reuse, and reporting will be the characterising METAstat functions, also in a metadata driven and interoperability perspective.

It is clear how crucial the aspects of governance and shared rules are before the development of the system: the definition of an appropriate Istat metadata governance, with impacts on relations with and between data systems that make use of metadata, represents an essential task of the project.

Box 3.5. Israel Governance Framework

The 'data governance framework' includes three main components:

- Principles for maintaining connectivity, privacy, quality, and trust, that are intended to create a common understanding, alignment, and coherence of the organisational efforts in the field of data.
- Access/permission rules that define who can access which data, in accordance with the new operating concept and the limits of responsibility established in this framework.
- Guiding principles for the architecture of the data, in setting up the data lake and other systems alongside development and updating processes as part of the Israel Central Bureau of Statistics (CBS) ongoing work.
- The governance framework for the CBS anchors the criteria, regulations, and standards in the following aspects:
- Entity model: Rules and principles of an entity model map (core vocabularies) for the CBS that also includes needs.
- Meta-data layer: Rules and principles for the architecture of the data, connectivity and information flow, storage, information retrieval, information storage, retrieval.
- Permission & rights management: Access permissions and compartmentalization of data, who can access the data, confidentiality, data anonymization, including access to information for researchers, permissions, data catalogues, dictionaries.
- Quality: includes quality assurance management and methodological elements for quality.

Entity model

At the CBS we chose to begin with a conceptual entity model that essentially represents the top level of data concepts in the organisation. The model will include subjects of content, connections, and topics on which the CBS activity is based.

Each topic and primary entity (core vocabularies) constitute a content world with unique features that can be identified in a distinct manner and for which the organisation is interested in its information and representing it in the database.

Metadata layer

Metadata set and regulate the rules, principles, quality, and architecture of the data.

The CBS is currently in the process of establishing a centralised metadata management system that aims to document and manage all types of metadata (structural, technical, descriptive, reference) at all the stages of the GSBPM – the business processes needed to produce official statistics (See the Metadata Flow Chart below)

Such metadata management system will be implemented within the data lake to promote both general and specific instructions for collecting, organising, and preserving metadata elements and turning them into a driving factor in the business processes to produce data in the CBS and dissemination for different users.

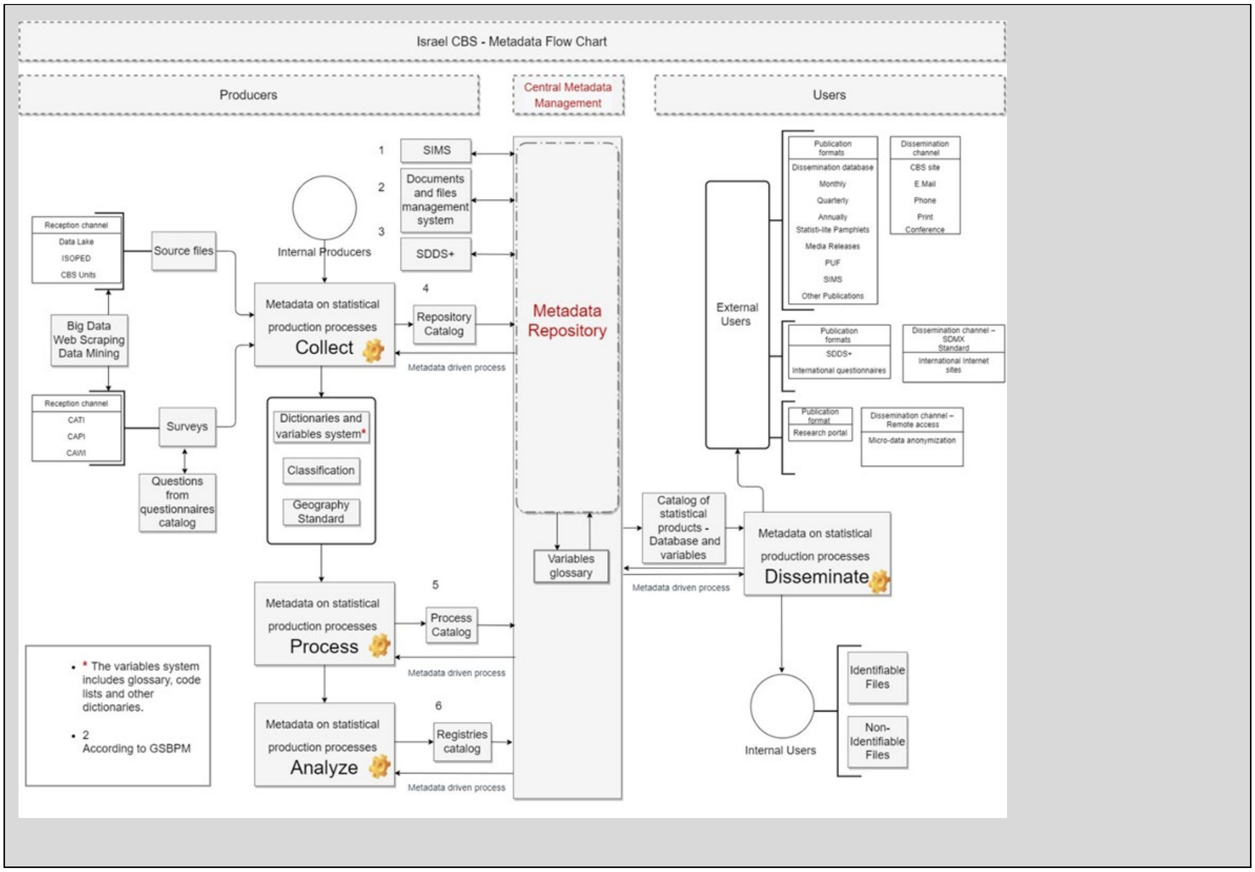
An example of the initial implementation of this process is the current implementation of the use of metadata standards (SIMS) across the organisation.

Permission & rights management

At the CBS we follow the principles of "Security by design" and the "Need-to-know" security principle.
The security-by-design policy ensures that systems and all their components are created from the very on-set with security in mind. It is about taking a proactive approach and integrating security from the very start.
The Need-to-know principle states that a user shall only have access to the information that their job function requires, regardless of their security clearance level or other approvals.
Also, we follow the Five Safes framework for helping make decisions about making effective use of data which is confidential or sensitive.
The Five Safes proposes that data management decisions be considered as solving problems in five 'dimensions': projects, people, settings, data, and outputs. The combination of the controls leads to 'safe use'.

- Safe projects - Is this use of the data appropriate?
- Safe people - Can the users be trusted to use it in an appropriate manner?
- Safe settings - Does the access facility limit unauthorized use?
- Safe data - Is there a disclosure risk in the data itself?
- Safe outputs - Are the statistical results non-disclosive?

Quality –
The quality assurance framework is an important part in the Governance Framework of the CBS.
It is constituted of two parts:
The first part describes the organisation's quality assurance management protocols - the appointment of the commissioner of statistical quality and his duties, the definition of quality trustees in the CBS departments, and their roles in the management and examination of Quality Indicators that will be generated in the CBS' data file management and processing system.
The second part is the methodological part which defines the development, execution, and examination of Leading Quality Control Indicators that will monitor the quality of sample-based and administrative data files in the CBS.
These two parts, together with the "statistics work regulations" of the CBS, which is based on the European Statistics Code of Practice, constitute the quality section of the CBS' Governance framework.



Israel CBS - Metadata Flow Chart
Producers
Central Metadata Management
Users
1
SIMS
2
Documents and files management system
3
SDDS+
4
Repository Catalog
Internal Producers
Reception channel
Data Lake
ISOPED
CBS Units
Source files
Big Data
Web Scraping
Data Mining
Reception channel
CATI
CAPI
CAWI
Surveys
Questions from questionnaires catalog
Metadata on statistical production processes
Collect
Metadata driven process
Dictionaries and variables system*
Classification
Geography Standard
Metadata Repository
Variables glossary
5
Metadata on statistical production processes
Process
Process Catalog
Metadata driven process
6
Metadata on statistical production processes
Analyze
Registries catalog
Metadata driven process
• * The variables system includes glossary, code lists and other dictionaries.
• 2
According to GSBPM
External Users
Publication formats
Dissemination database
Monthly
Quarterly
Annually
Statisti-lite Pamphlets
Media Releases
PUF
SIMS
Other Publications
Dissemination channel
CBS site
E.Mail
Phone
Print
Conference
Publication formats
SDDS+
International questionnaires
Dissemination channel – SDMX Standard
International Internet sites
Publication format
Research portal
Dissemination channel – Remote access
Micro-data anonymization
Catalog of statistical products - Database and variables
Metadata on statistical production processes
Disseminate
Metadata driven process
Internal Users
Identifiable Files
Non-Identifiable Files

4. Recommendations

This section contains a set of recommendations that include activities that help achieve a good level of statistical interoperability. There are recommendations to use some techniques - some specific, others more generic or "architectural," as well as the suggestion of some "organizational" changes that can support interoperability. Following most of these recommendations would allow the NSO to achieve a sort of "interoperability by design."

4.1. Develop Interoperability Strategy and Monitor Implementation

A strategy is a set of choices and decisions that together chart a high-level course of action to achieve the goals [1] and offers direction over the longer term for the organisation. The interoperability intertwines with other data management functions such as data security and quality, thus often forms a pillar of a broad data strategy of the organisation or nation-wide initiatives[7].

The interoperability strategy includes components such as a vision, business case, guiding principles, process, roles, structure, and standards. The interoperability strategy depends heavily on the context of individual organisation as it needs to consider priorities and resources of the organisation. For example, the scope of legal mandates of the organisation may range from data in subject areas (e.g., social, and economic) to geospatial data and all state data; consequently, defining the scope of the strategy.

A key principle is that the development of interoperability strategy should be inclusive and involve all stakeholders, including NSO staff, data users, and other government agencies where needed. The following are some of the key steps that are typically involved in the development:

1. Define the vision and goals for interoperability: What does the NSO want to achieve through interoperability? What does success look like in the future?
2. Establish the value proposition: What are the benefits that it hopes to achieve?
3. Identify the stakeholders. Who are the people and parts of the organisation that will be affected by interoperability? Who needs to be involved in the governance process?
4. Establish a governance structure. This should include a clear definition of roles and responsibilities. This also includes a governance process for making decisions. This is important to ensure that the governance process is transparent and accountable, and that the needs of all stakeholders are considered. The governance process should also be flexible and adaptable: this is important because technology is constantly changing, and NSOs need to be able to adapt their governance processes accordingly.

[7] For example, In UK the Open Standards Board works with the Cabinet Office and is accountable for transparent selection and implementation of open standards (see https://www.gov.uk/government/groups/open-standards-board), managing also the interoperability issues among public bodies and private companies; In Australia a framework describing also governance process were implemented in the "Australian Government Technical Interoperability Framework" available at https://www.unapcict.org/sites/default/files/2019-01/Australian%20Government%20Technical%20Interoperability%20Framework.pdf

5. Develop and implement interoperability standards. These standards should define how different systems and technologies will communicate with each other.

6. Monitor and evaluate interoperability initiatives. This is important to ensure that interoperability is being achieved in a way that meets the needs of the NSO and its stakeholders. Implementing an interoperability strategy can be time-consuming; hence, organisations might choose to initiate a trial through a small-scale project which allows for learning and adaptation, identifying successful aspects and areas for improvement. Consequently, modifications can be refined based on learned insights.

Metrics available to evaluate Interoperability level.

Establishing a set of clear metrics is important given that the journey toward interoperability involves multiple stakeholders and can span an extended period. These metrics serve as measurable indicators to assess progress and effectiveness in a concrete manner, providing a tangible means to evaluate how far the organisation has come. Here some suggestions of metrics that can be used to evaluate the interoperability "level" inside statistical organisations:

- Percentage of statistical processes that use common data standards and definitions - extent to which NSOs are using common standards to collect, process, and disseminate data.

- Percentage of statistical processes that use the same software tools in specific sub-processes - generalised software as opposed to ad-hoc software (e.g., starting from standard tools for data exchange like SDMX).

- Percentage of statistical processes that use the standard metadata system – level of use of standard metadata system.

- Percentage of statistical processes whose data are disseminated as Linked Open Data (LOD) - extent to which the "statistical" data are semantically integrated with other data.

- Percentage of statistical processes that can share data with each other seamlessly - extent to which different processes are able to share data with each other without the need for manual intervention.

- Level of automation of statistical processes - extent to which interoperability enables automation of statistical processes. Efficient interoperability should reduce manual interventions and streamline workflows.

- Stakeholder and user satisfaction with the interoperability of NSOs' statistical processes - collected through surveys or other feedback mechanisms to assess how satisfied users are with the ability to access and use NSO data from multiple sources.

Other assessment tools that can be useful to measure interoperability levels are maturity models. These are "tools that set out criteria and steps that help organizations measure their ability and continuous improvement in particular fields or disciplines" [9]. Maturity models define levels to characterise the state of specific fields or areas. Box 5 shows a few examples of interoperability maturity models.

Box 4.1. Examples of Interoperability Maturity Models (IMMs)

- European Commission's ISA programme IMM: focused on measuring how a public administration interacts with external entities to organize the efficient provisioning of its public services to other public administrations, businesses and or citizens. The model distinguishes three domains of interoperability: Service delivery, service consumption and service management and uses a five-stage model to indicate the interoperability maturity of the public service.

- DOE's Office of Scientific and Technical Information (OSTI) IMM: Written for stakeholders in technology integration domains, identifies interoperability criteria grouped into six main categories: Configuration and evolution, Safety and security, Operation and Performance, Organisational, Informational and Technical. In addition, as several criteria focus more on the culture changes and collaboration activities required to help drive interoperability improvements in an ecosystem or community of stakeholders, an additional "Community" category was formed. The maturity levels in the IMM are based on the Capability Maturity Model Integration (CMMI).

- National Archives of Australia Data IMM: can be applied to all data produced by an agency that has the potential to be integrated, exchanged, or shared. The DIMM helps to measure an agency information and data governance across five key themes: business, security, legal, semantic, and technical. Each theme is split into categories and each category has 5 steps that describe the common data interoperability behaviours, events, and processes for the corresponding level of maturity.

- GPSDD-UNSD Joined- Up Data Maturity Assessment: designed to be used by official statisticians and professionals of the sustainable development sector. The Maturity Assessment has three components: interoperability layers (organisational, human, data, and technological), each one with its dimensions and maturity levels.

Interoperability Maturity Models present relevant concepts and help to specify a strategic vision for interoperability. They focus on the relationship between interoperability and other specific areas that could be improved based on organisational objectives and can be used to identify the current level of data interoperability maturity, to identify gaps between the actual and desired interoperability level or to the planification of improvements to reach the maturity levels needed by an organisation. These maturity models besides considering semantic, structural, syntactic and system or technological interoperability facets, they consider other aspects like the legal, organisational, and human capabilities that can aid to achieve statistical interoperability.

4.2. Expand Use of Standards

Adoption of open standards is an important step in realising interoperability, establishing good governance practices, and achieving transparency of data and the processes used to generate them within statistical offices. But open standards are more than just accessible and usable by anyone.

They are open for any stakeholder to join the development process (unrestricted), the ones that join are a representative sample of the stakeholder community (balanced), the steps by which a standard is developed are easily inspected (transparent), the rules of the process apply fairly to all (fair), and decision making is determined by consensus (consensus). Such standards are called open, and they are the best candidates for statistical offices from which to choose.

Introduce Open Standards

Open standard refers to a standard that is openly accessible and usable by anyone. Compliance with open standards can significantly enhance interoperability within National Statistical Offices (NSOs) in several keyways:

1. Consistency and compatibility: Open standards provide a common framework for data representation and exchange. By adhering to these standards, NSOs ensure that their data formats and structures are consistent and compatible with those of other systems, enabling data integration and exchange between different entities.

2. Facilitated data sharing: Open standards create the basis for data sharing among NSOs and external stakeholders. When data is formatted and documented according to open standards, it becomes easier for different systems and organisations to share and access data, fostering improved collaboration and information exchange.

3. Reduced integration efforts: Open standards streamline the process of integrating data from disparate sources by providing a well-defined set of rules and protocols. NSOs can use these standards to minimize the efforts required for data integration, allowing for more efficient and cost-effective interoperability between systems and platforms.

4. Enhanced accessibility and transparency: Compliance with open standards promotes data accessibility and transparency, as it ensures that data is accessible and comprehensible to a wider audience. This accessibility fosters greater transparency in data sharing and dissemination, enabling stakeholders and the public to access, analyse, and utilize statistical information.

5. Long-term sustainability: By aligning with these standards, NSOs can ensure the long-term sustainability of their data management systems, as they remain compatible with evolving technological advancements and changing data requirements.

6. Promotion of innovation: Open standards encourage innovation and the development of new tools and technologies. NSOs can leverage open standards to foster a culture of innovation, enabling the integration of new technologies and methodologies for data collection, processing, and dissemination.

Overall, compliance with open standards in NSOs plays a vital role in fostering a more interconnected and efficient data ecosystem, promoting collaboration, transparency, and innovation within the statistical community. In Box 4.2, we find an example of use of open standards as enablers for interoperability by Statistics Canada.

Box 4.2. Enablers for interoperability: Statistics Canada "Enterprise Information and Data Management" (EIDM) project

Under the EIDM project, data management, metadata management, standards and governance were brought together in a unified vision that would enable interoperability.

As part of this 4-year project, DDI, SDMX, DCAT and associated DCAT application profiles are standards that were officially adopted by StatCan governance bodies. Policy instruments were updated to reflect the mandatory use of these standards and the enterprise tools that are based on these standards.

The project led to the implementation of the following tools and standards:

- Colectica to manage metadata on instruments and conceptual/referential metadata, e.g., variables, universes, and studies, using the DDI standard.
- .Stat/Istat/Fusion suite of tools to manage metadata on data structures and data exchanges, using the SDMX standard.
- Ariā, a classification management tool based on GSIM and Neuchâtel Model, with the possibility of mapping to SKOS and XKOS
- OpenLink Virtuoso, an Open-source RDF-based linked data platform, used as an integrated repository for metadata. Metadata from Colectica, SDMX, Ariā and CKAN will be converted to StatDCAT-AP or SKOS/XKOS, both of which are RDF formats. This repository also serves as the enterprise-wide data catalogue. A GUI-based Exploration Tool provides a window into the data catalogue. This brings StatCan a step closer to semantic interoperability.

All standards-based tools chosen use REST APIs and many can provide their information in CSV format, which are key to structural and syntactic interoperability.
Finally, a FAIR assessment tool was created to assist program areas in measuring their alignment in relation to FAIR. Use of the enterprise tools, approaches, frameworks, and standards enhances FAIRness. All new projects within the organization require Architecture governance approval to advance to the various stages of delivery. This ensures that all new projects align with approved standards, fundamental principles, and enterprise approaches.

Introducing data standards and common exchange models

Achieving interoperability is a multiple-step process that will take time to implement. There are several steps that can be taken both simultaneously and iteratively to introduce the standards that will lead to interoperability:

1. Establish requirements: In trying to determine which interoperability aspects one would like to address, first consider the needs of the users: introducing standards to drive interoperability must come from the business. What are the pain points that interoperability will help address? This will require a thorough analysis of the current state. There are likely to be multiple pain points, but not all needs can be addressed at the same time. Once the current state is understood, map out the state "to be."

2. Determine a core logical model: If there is a requirement to have better metadata, developing a core logical model will be key to the success of this initiative, as it will guide which standard to use to address the metadata gap.

3. Research and choose a standard that best meets the requirements: Examine which standards will help address the interoperability issues into play. Use the GSIM-based model in Section 3 and examine the associated standards from Annex 3.

   If you have metadata needs that are related to describing microdata, consider the use of DDI (Data Documentation Initiative); if your metadata needs are related to better describing aggregate statistics, consider the use of SDMX (Statistical Data and Metadata eXchange).

Sometimes there are multiple standards that can meet the requirements. Decisions on which standard to use can be based on the following: availability of open standards, availability of open-source applications, wide use of a standard, fit within the current ecosystem, etc.

Consider which ones would be the easiest to implement and start with one or a few of those. There are likely already processes and systems in place that can be used or expanded upon.

4. Investigate approach through experiments: Investigate your proposed approach through small and focussed experiments. By doing so, it will be easier to learn through them and to iterate.

   Ensure that you involve business partners through all steps of the experiments; a user-centric approach will ultimately lead to the success of the implementation.

   Explore how this new initiative would fit within an existing ecosystem.

   If applicable, consider how it would relate to existing metadata. This will require mapping the legacy metadata to the standard being considered. You may choose to migrate existing metadata or simply link it within the ecosystem.

   Note that re-using existing applications based on the chosen standard will likely yield quicker results than building new applications.

5. Validate results: It is important to validate the tools, the standards, and the business processes. If the experiment doesn't yield the expected results, reconsider the standard, tool or business processes used. You may find that the business requirements were not properly expressed.

6. Move towards deployment: Once an approach has been determined as optimal, ensure that the proper governance tools are in place to make mandatory the use of the standard and its related tools.

   Complete all steps necessary to move the standard, tools, and business processes into active use.

   If your new initiative is part of an existing ecosystem, it will be necessary to integrate it within the ecosystem. If your interoperability initiative is related to improved metadata, migrate or link to existing metadata.

   Note that the last two steps may be part of their own individual projects to ensure that initiatives stay small and focussed.

7. Onboard new users: The final step is to onboard new users. This will require a change management plan, where communications, training and user guides are key components.

Use semantic web techniques (semantic interoperability)

The Semantic Web is an extension of the World Wide Web that allows machines to understand the meaning of data. This is done by adding metadata to data, which describes what the data is about and how it can be used.

Semantic Web technologies enable people to create data stores on the Web, build vocabularies, and write rules for handling data. This is empowered by technologies such as RDF, SPARQL, OWL, LOD and SKOS, which are standards from the World Wide Web Consortium (W3C).

The semantic web can also help to improve the quality of official statistics by enabling data to be more easily validated, integrated, and analysed. By providing a common framework for data exchange, the semantic web can help to reduce the risk of errors and inconsistencies in data and enable more accurate and reliable statistical analysis.

The Resource Description Framework (RDF) is a framework for expressing information about resources. Resources can be anything, including documents, people, physical objects, and abstract concepts.

RDF is intended for situations in which information on the Web needs to be processed by applications, rather than being only displayed to people. RDF provides a common framework for expressing this information so it can be exchanged between applications without loss of meaning. Since it is a common framework, application designers can leverage the availability of common RDF parsers and processing tools. The ability to exchange information between different applications means that the information may be made available to applications other than those for which it was originally created.

Many of the standards in place that are listed in the annex of chapter 3 and applicable to statistical data or metadata are both RDF-based and open: DCAT (Data Catalog Vocabulary) and associated application profiles DCAT-AP, StatDCAT-AP and GeoDCAT-AP; XKOS to name a few.

Linked Open Data (LOD) is a basic component of the Semantic Web techniques as they provide a standardized, linked, and machine-readable framework for representing and exchanging statistical data. LOD can improve statistical interoperability in several ways:

- **Publish statistical data as LOD.** This will make it easier for machines to understand and use the data, and to link it to data from other sources.
- **Use LOD to create a central repository for statistical metadata.** This will make it easier for users to find and understand the data that is available.
- **Develop applications that use LOD to automatically discover and use statistical data.** This will make it easier for users to access and use the data, and to create new and innovative statistical products and services.

4.3. Foster Culture Change and Support Staff

While interoperability involves adhering to specifications and standards, the challenges often arise not from lack of standards, but from resistance to following the standards. Therefore, achieving interoperability requires shift of mindset and overcoming resistance across the organisation. Resistance can be attributed from legal or regulatory hurdles (e.g., need for data subject permission to share data), but some:

(i) Technical challenges barrier (e.g., need for learning new technologies, tools, and staff training,

(ii) Organisational barriers such as the need for workflow changes (silo breakdowns), and a culture shift towards collaboration. The fragmentation within the organisation is not necessarily due to lack of knowledge or skills to implement technical standards for data interoperability, but rather lack of time and resources reflecting how difficult it is to change course in the way day-to-day operations are carried out when staff must deal with a continuous demand for new data products while keeping key legacy systems running.[8]

People who are affected by the adoption of standards, especially those who must alter their established practices, tend to resist change. This is why involving stakeholders from the outset is crucial. It helps them view the requirements not as impositions but as necessary steps to deliver greater value to the organisation.

The vision and business case set out in the interoperability strategy (section 4.1) therefore plays an important role in imbuing the common sense of direction.

Sponsorship by high-level management can be a valuable for overcoming resistance to the introduction of interoperability providing **credibility and legitimacy,** building trust among the staff and stakeholders, and **advocacy** building support for the initiative and to overcome resistance.

[8] See Francesca Perucci, "Data Interoperability: Lessons from UN Statisticians" at https://unite.un.org/blog/data-interoperability-lessons-from-un-statisticians

References

[1] DAMA International; DAMA-DMBOK. Data Management Body of Knowledge. 2nd Edition; Technics Publications; USA; 2017 (https://technicspub.com/dmbok/).

[2] CES Task Force on Data Stewardship (2023) Data stewardship and the role of national statistical offices in the new data ecosystem (https://unece.org/sites/default/files/2023-04/CES_02_Data_stewardship_for_consultation.pdf; accessed July 2023).

[3] National Academies of Sciences, Engineering, and Medicine 2022. Transparency in Statistical Information for the National Center for Science and Engineering Statistics and All Federal Statistical Agencies. Washington, DC: The National Academies Press. https://doi.org/10.17226/26360.

[4] The Open Data Institute (2013). What makes data open? https://theodi.org/insights/guides/what-makes-data-open/

[5] World Bank (2021). World Development Report 2021: Data for better lives. https://www.worldbank.org/en/publication/wdr2021

[6] Kécia Souza, Larissa Barbosa, Rita Suzana Pitangueira; Interoperability Types Classifications: A Tertiary Study; ACM Digital Library; USA; 2021; https://doi.org/10.1145/3466933.3466952

[7] GSBPM v5.1 (2019) The Generic Statistical Business Process Model https://unece.org/statistics/documents/2019/01/standards/gsbpm-v51

[8] ECE/CES (2022). Classification of Statistical Activities (CSA) 2.0 and explanatory notes. https://unece.org/sites/default/files/2022-05/ECE_CES_2022_8-2205369E.pdf

[9] González Morales Luis, Orell Tom; Introducing the Joined-Up Data Maturity Assessment; UNSD-GPSDD (2020) https://www.data4sdgs.org/sites/default/files/file_uploads/Joined_Up_Data_Maturity_Assessment_draft5.pdf

Annex 1 - Standardised Vocabularies, Methods, Formats, Frameworks, Languages, Workflows and Data models

(Meta)Data catalogues

Vocabularies

- Schema.org: is a reference website that publishes documentation and guidelines for using structured data mark-up on webpages (called microdata). It is a part of the semantic web project.
- DCAT (Data Catalog Vocabulary): is an RDF vocabulary designed to facilitate interoperability between data catalogues published on the Web. Several application profiles were created and are in use:
    - DCAT-AP: is the DCAT Application Profile for data portals in Europe (DCAT-AP). It is a specification based on the Data Catalogue Vocabulary (DCAT) developed by W3C. This application profile is a specification for metadata records to meet the specific application needs of data portals in Europe while providing semantic interoperability with other applications based on reuse of established controlled vocabularies (e.g., EuroVoc) and mappings to existing metadata vocabularies (e.g., Dublin Core, SDMX, INSPIRE...).
    - StatDCAT-AP: is an extension of the DCAT Application Profile for Data Portals in Europe (DCAT-AP) to enhance interoperability between descriptions of statistical data sets within the statistical domain and between statistical data and open data portals.
    - GeoDCAT-AP: is a geospatial extension for the DCAT application profile for data portals in Europe.
- Dublin Core: also known as the Dublin Core Metadata Element Set (DCMES), is a set of fifteen main metadata items for describing digital or physical resources. Dublin Core has been formally standardized internationally as ISO 15836 and as IETF RFC 5013.
- Data Quality Vocabulary (DQV): is a (meta)data model implemented as an RDF vocabulary, which extends the DCAT with properties and classes suitable for expressing the quality of datasets.
- PROV is a specification that provides a vocabulary to interchange provenance information. Users can do so by marking up their web page or by making available provenance information expressed as linked data.
- DDI-RDF Discovery Vocabulary (Disco): defines an RDF Schema vocabulary that enables discovery of research and survey data on the Web. It is based on DDI XML formats of DDI Codebook and DDI Lifecycle.

Concepts

Vocabularies

- DDI (Data Documentation Initiative) is a free international standard for describing the data produced by surveys and other observational methods in different sciences. DDI can document and manage different stages in the research data lifecycle.
- SDMX (Statistical Data and Metadata eXchange) is an international initiative that aims at standardising and modernising the mechanisms and processes for the exchange of statistical data and metadata among international organisations and their member countries.
- SKOS/XKOS: SKOS (Simple Knowledge Organization System) is an area of work developing specifications and standards to support the use of knowledge organization systems (KOS) such as thesauri, classification schemes. XKOS leverages the SKOS for managing statistical classifications and concept management systems.
- NIEM (National Information Exchange Model) is a common vocabulary that enables efficient information exchange across diverse private and public organisations.
- XBRL (eXtensible Business Reporting Language) is an open international standard for digital business reporting, managed by a global not for profit consortium. It provides a language in which reporting terms (mostly financial) can be authoritatively defined.
- FIBO (Financial Industry Business Ontology) is the industry standard resource for the definitions of business concepts in the financial services industry. This dictionary enables you to detect the terminology defined by the FIBO Vocabulary
- GML (Geography Markup Language) is an OpenGIS Implementation Specification designed to store and transport geographic information. GML is a profile (encoding) of XML.
- RCC (Region Connection Calculus) is a method used in AI of representing and reasoning about space. It is based on the idea of dividing space into regions and representing the relationships between regions using a set of calculus rules.
- OWL-Time is an ontology of temporal concepts, for describing the temporal properties of resources in the world or described in Web pages. The ontology provides a vocabulary for expressing facts about relations among instants and intervals and information about durations and temporal position.
- ORG is a core ontology for organisational structures, aimed at supporting linked data publishing of organisational information across several domains.

Exchange

Vocabularies

- DDI (Data Documentation Initiative) is a free international standard for describing the data produced by surveys and other observational methods in different sciences. DDI can document and manage different stages in the research data lifecycle.

- SDMX (Statistical Data and Metadata eXchange) is an international initiative that aims at standardising and modernising the mechanisms and processes for the exchange of statistical data and metadata among international organisations and their member countries.
- PMML (Predictive Model Markup Language) is an XML-based predictive model interchange format. It provides a way to describe and exchange predictive models produced by data mining and machine learning algorithms.
- NIEM (National Information Exchange Model) is a common vocabulary that enables efficient information exchange across diverse private and public organizations.
- The OpenAPI Specification is a specification language for HTTP APIs that provides a standardized means to define your API to others.

Methods

- REST (Representational state transfer) is a software architectural style that was created to guide the design and development of the architecture for the World Wide Web.
- SOAP (Simple Object Access Protocol) is a messaging protocol specification for exchanging structured information in the implementation of web services in computer networks.
- SPARQL (SPARQL Protocol and RDF Query Language), is the standard query language and protocol for Linked Open Data on the web or for RDF triplestores.
- SHACL (Shapes Constraint Language) is a W3C standard language for describing Resource Description Framework (RDF) graphs.
- GraphQL is an open-source data query and manipulation language for APIs and a query runtime engine.
- ODATA (Open Data Protocol) is an open protocol (ISO standard) that allows the creation and consumption of queryable and interoperable REST APIs in a simple and standard way.
- Protocol Buffers is a free and open-source cross-platform data format language-neutral, platform-neutral extensible mechanism for serializing structured data.

Formats

- XML (Extensible Markup Language) is a markup language for storing, transmitting, and reconstructing arbitrary data. It defines a set of rules for encoding documents in a format that is both human-readable and machine-readable.
- JSON (JavaScript Object Notation) is a lightweight data-interchange format, easy both for humans and for machines to parse and generate. JSON is a text format completely language independent but uses conventions that are familiar to programmers of the C-family of languages (C, C++, Java, Python …). These properties make JSON an ideal data-interchange language.
- YAML (YAML Ain't Markup Language) is a human-readable data-serialization language for all programming languages. It is commonly used for configuration files and in applications where data is being stored or transmitted.

- HTML (HyperText Markup Language) is the standard markup language for documents designed to be displayed in a web browser.

- JSON-LD (JSON for Linking Data) is a lightweight Linked Data format. It is based on the already successful JSON format and provides a way to help JSON data interoperate at Web-scale. JSON-LD is an ideal data format for REST Web services and unstructured databases such as Apache CouchDB and MongoDB.

- RDF/XML (Resource Description Format/eXtensible Markup Language) is a syntax, defined by the W3C, to express an RDF graph as an XML document. RDF/XML is sometimes simply called RDF because it was historically the first W3C standard RDF serialization format.

- Turtle (Terse RDF Triple Language) is a syntax and file format for expressing data in the Resource Description Framework data model. Turtle syntax is like that of SPARQL, an RDF query language.

- CSV (Comma-Separated Values) file is a delimited text file that uses a comma to separate values. Each line of the file is a data record. Each record consists of one or more fields, separated by commas.

- Text file (sometimes spelled textfile; an old alternative name is flatfile) is a kind of computer file that is structured as a sequence of lines of electronic text. A text file exists stored as data within a computer file system.

Frameworks

- GSBPM (Generic Statistical Business Process Model) is a model that describes statistics production in a general and process-oriented way. It is used as a common basis for work with statistics production in different ways, such as quality, efficiency, standardisation, and process-orientation.

- GAMSO (Generic Activity Model for Statistical Organisations) describes and defines the activities that take place within a typical organisation that produces official statistics. It extends and complements the GSBPM by adding additional activities needed to support statistical production.

- GSIM (Generic Statistical Information Model) is a reference framework of internationally agreed definitions, attributes and relationships that describe the pieces of information that are used in the production of official statistics (information objects).

- ESS EARF (European Statistical System Enterprise Architecture Reference Framework) is a set of documents containing several key artefacts, which can be used at various stages in projects as well as in the overall governance of the realisation of Eurostat Vision 2020.

- EIRA (European Interoperability Reference Architecture) is a four-view reference architecture for delivering interoperable digital public services across borders and sectors. It defines the required capabilities for promoting interoperability as a set of architecture building blocks (ABBs).

- CSPA (Common Statistical Production Architecture) is a reference architecture for the statistical industry, which has been developed and peer reviewed by the international statistical community. CSPA provides a framework, including principles, processes, and guidelines, to help reduce the cost of developing and maintaining processes and systems.

- CSDA (Common Statistical Data Architecture) is a Data Architecture developed by UNECE, focused on capabilities related to data and metadata, which can be seen as "data management resources," rather than on the structure and organization of data assets.
- FAIR (Find, Access, Interoperate, and Reuse) Digital Objects provide a framework to develop cross-disciplinary capabilities, deal with the increasing data volumes, build tools that help to increase trust in data, create mechanisms to efficiently operate in the scientific domain, and promote data interoperability.
- International Open Data Charter is a set of principles and best practices for the release of governmental open data, formally adopted by many governments.

Business

Languages

- Java is a high-level, class-based, object-oriented programming language that is designed to have as few implementation dependencies as possible.
- Python is an interpreted, object-oriented, high-level programming language with dynamic semantics.
- Scala combines object-oriented and functional programming in one concise, high-level programming language.
- SAS is a statistical software suite developed by SAS Institute for data management and statistical analysis.
- R is a free and open-source extensible language and environment for statistical computing and graphics.
- MDX (MultiDimensional eXpressions) is a query language is a query language used to create calculations and aggregations / DAX (Data Analysis eXpressions) is a formula language used to create calculations and aggregations. The languages were developed by Microsoft.
- SDTL (Structured Data Transformation Language) is an independent intermediate language for representing data transformation commands. SDTL, developed by DDI Alliance, consists of JSON schemas for common operations.
- VTL (Validation and Transformation Language) is a standard language for defining validation and transformation rules (set of operators, their syntax, and semantics) for any kind of statistical data.

Workflows

- DDI-CDI (Cross Domain Integration) is a specification aimed at helping implementers integrate data across domain and institutional boundaries. DDI-CDI focuses on a uniform approach to describing a range of needed data formats: traditional wide/rectangular data, long [event] data, multi-dimensional data, and NoSQL/key-value data.

- BPMN (Business Process Model and Notation) is a standard set of diagramming conventions for describing business processes. It visually depicts a detailed sequence of business activities and information flows needed to complete a process.

- CMMN (Case Management Model and Notation) is a graphical notation used for capturing work methods that are based on the handling of cases requiring various activities that may be performed in an unpredictable order in response to evolving situations.

- DMN (Decision Model and Notation) is a modelling language and notation for the precise specification of business decisions and business rules. DMN is easily readable by the different types of people involved in decision management.

- ProvONE is defined as an extension of the W3C recommended standard PROV, aiming to capture the most relevant information concerning scientific workflow computational processes.

- CWL (Common Workflow Language) is an open standard for describing how to run command line tools and connect them to create workflows. Tools and workflows described using CWL are portable across a variety of platforms.

Structures

Data models

- DDI (Data Documentation Initiative) is a free international standard for describing the data produced by surveys and other observational methods in different sciences. DDI can document and manage different stages in the research data lifecycle.

- SDMX (Statistical Data and Metadata eXchange) is an international initiative that aims at standardising and modernising the mechanisms and processes for the exchange of statistical data and metadata among international organisations and their member countries.

- EDM (Entity Data Model) is a set of concepts that describe the structure of data, regardless of its stored form. The EDM borrows from the Entity-Relationship Model described by Peter Chen in 1976, but it extends its traditional uses.

- RDF DM (Data Model) is a standard model for data interchange on the Web. RDF has features that facilitate data merging, and it specifically supports the evolution of schemas over time.

- A Labelled Directed Graph is, as the name suggests, a Directed Graph whose arrows have labels on them. A Directed graph (or digraph) is a graph that is made up of a set of vertices connected by directed edges, called arcs.

- HDF5 (Hierarchical Data Format version 5), is an open-source file format that supports large, complex, heterogeneous data. HDF5 uses a "file directory" like structure that allows you to organize data within the file in many different structured ways, as you might do with files on your computer.

Vocabularies

- DDI (Data Documentation Initiative) is a free international standard for describing the data produced by surveys and other observational methods in different sciences. DDI can document and manage different stages in the research data lifecycle.
- SDMX (Statistical Data and Metadata eXchange) is an international initiative that aims at standardising and modernising the mechanisms and processes for the exchange of statistical data and metadata among international organisations and their member countries.
- RDF Data Cube provides a means to publish multi-dimensional data, such as statistics, on the web in such a way that it can be linked to related data sets and concepts using the W3C RDF (Resource Description Framework) standard.
- CSVW (CSV on the Web) is a standard method for publishing and sharing data held within CSV files.

Formats

- XML (Extensible Markup Language) is a markup language for storing, transmitting, and reconstructing arbitrary data. It defines a set of rules for encoding documents in a format that is both human-readable and machine-readable.
- JSON (JavaScript Object Notation) is a lightweight data-interchange format, easy both for humans and for machines to parse and generate. JSON is a text format completely language independent but uses conventions that are familiar to programmers of the C-family of languages (C, C++, Java, Python …). These properties make JSON an ideal data-interchange language.
- RDF/XML (Resource Description Format/eXtensible Markup Language) is a syntax, defined by the W3C, to express an RDF graph as an XML document. RDF/XML is sometimes simply called RDF because it was historically the first W3C standard RDF serialization format.
- Turtle (Terse RDF Triple Language) is a syntax and file format for expressing data in the Resource Description Framework data model. Turtle syntax is like that of SPARQL, an RDF query language.
- JSON-LD (JSON for Linking Data) is a lightweight Linked Data format. It is based on the already successful JSON format and provides a way to help JSON data interoperate at Web-scale. JSON-LD is an ideal data format for REST Web services and unstructured databases such as Apache CouchDB and MongoDB.
- Protocol Buffers is a free and open-source cross-platform data format language-neutral, platform-neutral extensible mechanism for serializing structured data.
- CSV (Comma-Separated Values) file is a delimited text file that uses a comma to separate values. Each line of the file is a data record. Each record consists of one or more fields, separated by commas.

Annex 2 - Applications that use standards.

This annex includes a non-exhaustive list of applications which use standards. This list does not in any way preclude endorsement of these tools; it is simply meant as a starting point in finding applications for some of the standards.

RDF-based metadata management open-source tools

- Based on SKOS, iQvoc supports vocabularies that are common to many knowledge organisation systems, such as thesauri, taxonomies, classification schemes and subject heading systems.
- Tematres is an open-source vocabulary server, web application to manage and exploit vocabularies, thesauri, taxonomies, and formal representations of knowledge based on the SKOS standard.
- VocBench is a web-based, multilingual, collaborative development platform for managing OWL ontologies, SKOS(/XL) thesauri, Ontolex-lemon lexicons, and generic RDF datasets.
- BluLab is a web based SKOS Editor developed by BluLab, Ohio. The web based SKOS editor allows users to create, curate, version, manage, and visualise SKOS resources.
- Vocabs editor is a web-based tool for collaborative work on controlled vocabularies development. The editor follows the SKOS data model for the main elements of a vocabulary. The Dublin core schema is used to capture the metadata (such as date created, date modified, creator, contributor, source and other) about each element. Each concept scheme as well as each individual concept can be downloaded in RDF/XML and Turtle format.

Open-source RDF-based linked data platforms

- OpenLink Virtuoso is a high-performance and scalable Multi-Model RDBMS, Data Integration Middleware, Linked Data Deployment, and HTTP Application Server Platform.
- Apache Jena is a Java framework for building Semantic Web and Linked Data applications.

Open-source data catalogues

- CKAN is an open-source DMS (data management system) for powering data hubs and data portals. CKAN makes it easy to publish, share and use data.
- GeoNetwork is a catalogue application to manage spatially referenced resources. It provides powerful metadata editing and search functions as well as an interactive web map viewer.

DDI-based tools

A list of DDI-based tools, which cover various versions of DDI (e.g., codebook, lifecycle), as well as a variety of functionalities from authoring and editing to data transformations and conversions, can be found at:

DDI Tools | Data Documentation Initiative (ddialliance.org)

SDMX-based tools

A range of SDMX tools, which allows structural metadata management, reference metadata editing, data management, reporting, dissemination, and other functionalities can be found at:

Tools | SDMX – Statistical Data and Metadata eXchange

Annex 3 - Roles and responsibilities from ISO/IEC 11179

The ISO/IEC 11179 is an international standard for representing, storing, and maintaining metadata in a controlled environment (a metadata registry). This standard, consisting of six parts, is focused on semantics, representation, and description of data. Its purpose is to promote standard description of data; common understanding of data across organisational elements and between organisations; re-use and standardization of data over time, space, and applications; harmonization and standardization of data within an organisation and across organisations; management of the components of descriptions of data; re-use of the components of descriptions of data.

ISO/IEC 11179 is a general description framework for data of any kind, in any organisation and for any purpose. ISO/IEC 11179 does not address other data management needs, such as data models, application specifications, programming code, program plans, business plans and business policies.

The 6th part of the standard provides registration guidelines, describing the procedure by which metadata items, or other registry items, required in various application areas can be assigned an internationally unique identifier and registered in a metadata registry maintained by one or more Registration Authorities. Part of the Annex B is specifically devoted to the roles associated with the metadata registry. A summary is provided below.

Roles associated with the metadata registry.

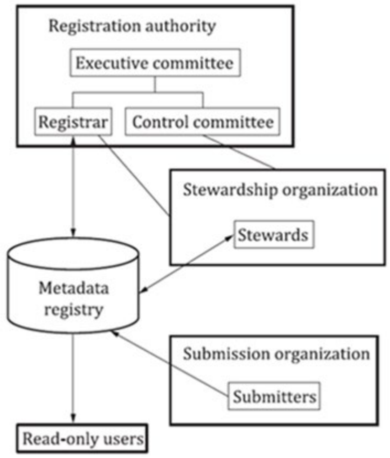
In the Annex B of the sixth part of the standard organisational roles and responsibilities associated with the administered item registration process are identified and suggested.

There are three types of registration acting bodies (RAB) in the framework of the ISO/IEC 11179:

- registration authorities
- submitting organisations
- stewardship organisations

Each type of registration acting body should meet the criteria, fulfil the roles, and assume the responsibilities. The Figure below provides a high-level view of how these organisational roles are related within the context of a metadata registry.

Organisational roles to the metadata registry and their relationships (Source: ISO/IEC 11179-6:2023 ed. 4)



Key
reporting structure
data access (read, write or both)

Role	Responsibilities
Registration authorities (RA)	
Metadata **registry registration authority** Organizational unit that establishes and publishes procedures for the operation of its metadata registry. A registration authority should receive and process proposals from submitting organizations for registration of administered items falling within its registration domain. A registration authority is responsible for maintaining the metadata register of administered items and issuing of international registration data identifiers (IRDIs).	To establish itself as a **registration authority**, an organization should complete the following. - Secure a Registration Authority Identifier (RAI), namely a unique internationally unique recognized organization code. - Prescribe, amend, and interpret the procedures to be followed for the registration of administered items in accordance with this document. - Determine any additional conditions specifically required by its domain of registration within its metadata registry. - Specify the format for each attribute and specify the media by which an item for administration should be submitted for registration. - Establish and publish the rules by which its metadata registry should be made available. The registration authority shall specify the allowable users, the accessible contents, the frequency of availability, and the language(s), media, and format in which the information is provided for the metadata registry. Regarding applications for registering items for administration, a registration authority should fulfil the following responsibilities.

	- Receive and process applications for the registration of items for administration from its submitting organizations. - Assign international registration data identifier values, and maintain a metadata register in accordance with its procedures. - Consult the appropriate stewardship organizations when requests affect the mandatory attributes of the administered items being registered. - Handle all aspects of the registration process in accordance with good business practice and take all reasonable precautions to safeguard the metadata register. - Review and facilitate the progression of the applications through the registration cycle. - Assign an appropriate registration status. - Notify submitting organizations of its decisions according to the procedure specified in it rules.
Registrar Organizational unit within the registration authority, expert in registration processes, responsible for facilitating the registration of administered items and making those administered items widely accessible and available to the community.	The **registrar** provides a single point-of-contact responsible for managing and maintaining information about data in the metadata register, under the authority of the registration authority. The registrar should be responsible for: a) monitoring and managing the metadata registry contents b) enforcing policies, procedures, and formats for populating and using the metadata registry; c) proposing procedures and standard formats for the metadata registry to the control committee for consideration; d) recording current registration status for administered items in the metadata register; e) ensuring access for authorized users to contents in the metadata registry; f) assisting in the progression of administered items through the registration status levels; g) assisting in the identification and resolution of duplicate or overlapping semantics of administered items in the metadata register; h) acting on direction from the registration authority; i) effecting registration of administered items in external metadata registers or dictionaries; j) enforcing data registration procedures for submitting administered items to the metadata registry, e.g.: - how to prepare, submit, and process submissions of administered items; - how the metadata registry is used to avoid duplicate administered items submissions to the metadata register;

	- how the metadata registry is used to effect harmonization of data across metadata registers of participating organizations; - how external metadata registers are used as a source of administered items for reuse in the metadata register; k) maintaining a separate document recording the appropriate contact information for all members of the control committee and the executive committee; l) adding new users or organizational entities that may become authorized to access the metadata register; m) maintaining other controlled word lists of the metadata registry.
Executive committee Organizational unit responsible for administering responsibilities and authority delegated by the registration authority.	The **executive committee** should be responsible for overall policy and business direction for the metadata registry, to include: a) establishing overall metadata registry policies; b) resolution of all business management issues pertaining to the metadata registry, e.g. copyrights, stewardship, executive committee membership, etc; c) ensuring the long-term success and performance of the metadata registry; d) establishing and updating the metadata registry charter and strategic plans; e) meeting periodically in face-to-face meetings, with additional meetings and/or teleconferences held as needed. The executive committee will normally fulfil its responsibilities via consensus building. Intractable issues may be resolved by an established procedure.
Control committee It provides technical direction and harmonization of administered items for the metadata register. The membership of the control committee may include registrars and stewards.	The **control committee** provides overall technical direction of, and resolution of technical issues associated with, the metadata registry, its contents, and its technical operations. The control committee should be responsible for: a) overall conduct of registration operations; b) promoting the reuse and sharing of data in the metadata register within and across functional areas, and among external interested parties to the enterprise; c) progressing administered items through "Qualified," "Standard," and "Preferred Standard" registration status levels; d) resolving semantical issues associated with registered administered items, e.g. overlap, duplication, etc; e) approving updates to Administered Items previously placed in the metadata register with the "Qualified," "Standard," or "Preferred Standard" registration status levels; f) proposing metadata registry policies to the executive committee for approval; g) approving authorized submitters, read-only users, and types of users, of the metadata registry; h) approving metadata registry content, procedures, and formats; i) submitting management-related recommendations and issues to the Executive Committee;

	j) acting on directions from the executive committee; k) meeting periodically in face-to-face meetings, with additional meetings and teleconferences held as needed. The control committee will normally fulfil its responsibilities via consensus building in accordance with an established procedure. Intractable issues may be resolved by an established procedure.
Stewardship organizations (StO)	
Stewardship organizations They are usually designated by an organizational unit to ensure consistency of related administered items managed by its submitting organizations. A stewardship organization is the organization, or part thereof, that is responsible for the integrity and accuracy of the attribute values of the administered item, e.g. the semantics of administered items maintained and controlled by a registration authority.	A **stewardship organization** should: - at the registration authority's request, advise on the semantics, name, and permissible values for the administered item's attribute values submitted for registration; - notify the registration authority of any amendments to the administered items assigned to the stewardship organization; - decide, in case of confusion and/or conflict, on the attribute values of the assigned Administered Items.
Steward Stewards should be responsible for the accuracy, reliability, and currency of descriptive metadata for administered items at a registration status level of "Qualified" or above within an assigned area. Stewards should be responsible for metadata within specific areas and may have responsibilities that cut across multiple areas (e.g. value domains such as date, time, location, codes for the countries of the world).	**Stewards** provide specific expert points of contact responsible for coordinating the identification, organization, and establishment of registered data for use throughout the enterprise within an assigned functional area. Stewards should be responsible for: a) coordinating the identification and documentation of administered items within their assigned functional area; b) ensuring that appropriate administered items in their assigned functional area are properly registered; c) coordinating with other stewards to attempt to prevent or resolve duplicated efforts in defining administered items; d) reviewing all administered items once they are in the "Recorded" status to identify and attempt to resolve conflicts among administered items with other stewards assigned functional areas; e) ensuring the quality of metadata attribute values for administered items they propose for the "Qualified" registration status level, reusing standardized data from external metadata registers where applicable; f) proposing "Standard" registration status level administered items in their assigned functional area; g) Proposing "Preferred Standard" registration status level administered items in their assigned functional area;

<table>
<tr><td></td><td>h) ensuring that data registration procedures and formats are followed within their assigned functional area;
i) recommending submitters to the registration authority.</td></tr>
<tr><td colspan="2">Submitting organizations (SuO)</td></tr>
<tr><td>Submitting organization
Any organization that submits items to a registration authority for entry into its metadata registry. Each registration authority may establish its own criteria for registration eligibility.</td><td>A submitting organization is responsible to:
- provide the information specified as required by the registration authority;
- provide any additional information relevant to the item submitted for registration;
- ensure that when an Administered Item has been registered, specification of the attribute values of the administered item is not changed without first advising the registration authority.</td></tr>
<tr><td>Submitter
Organizational unit approved by a process defined by the registration authority. A submitter is authorized to identify, and report administered items suitable for registration.</td><td>Submitters are organization elements that are familiar with or engaged in development and operational environments. Submitters maintain current administered items and are engaged to describe and submit new administered items following the registration requirements.
A submitter should be responsible for:
a) identifying himself to the register;
b) identifying and documenting administered items appropriate for registration in the metadata register;
c) submitting administered items to the metadata register;
d) ensuring the completeness of mandatory metadata attributes for administered items proposed for the “Recorded” registration status level.</td></tr>
<tr><td colspan="2">Others</td></tr>
<tr><td colspan="2">All others
A registration authority may establish guidelines on the use of their metadata registry by other users. The general goal should be to provide an open area that anyone may use to obtain and explore the metadata that is managed within the metadata registry.</td></tr>
<tr><td colspan="2">Read-only user
Organizational unit or individual that is approved to review the contents of the metadata register. A “read-only” user has access to the contents in the metadata register, but is not permitted to submit, alter, or delete contents.</td></tr>
</table>