Data Governance Framework for Statistical Interoperability (DAFI)

DRAFT

modernstats by HLG-MOS

Executive Summary

DRAFT

Contents

DRAFT

Acronyms – to be completed

FAIR – Findable, Accessible, Interoperable, Reusable

GSBPM – Generic Statistical Business Process Model

ISO – International Organization for Standardization

SDMX – Statistical Data and Metadata Exchange

DRAFT

1. Introduction

1.1. Background

The primary aim of statistical organisations is to produce high-quality information to reflect the society phenomena accurately, completely, and timely. Statistical information refers to describe different aspects such as demography, society, economy, and environment, among others. It is used as input in the design, monitoring, and evaluation of public policies as well as in the making of a wide set of other decisions. 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 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 programs in the statistical organisations or independent organisations in a 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. While statistics derived from surveys and censuses could offer an accurate and comprehensive portrayal of society, its economy and its environment through their systematic data collection approach and rigorous survey methodologies, 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 like 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 provides a reference framework that contains the core elements to implement a governance program focused on achieving data interoperability and thus helping statistical organisations improve their data management.

1.2. Problem Statement

The term interoperability refers to the ability of systems or services to exchange and make use of information without prior communication for each specific case. 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 the potential that can be delivered by integrating 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.

It is possible to exchange, integrate, link and query data and metadata sets for different purposes, link administrative registers and other sources to generate derived statistics with specific goals like national accounts, integrate statistical indicators from various sources to disseminate new aggregates, link open data sources for design, monitoring and evaluation of public politics or with research goals in public and private sectors, among others.

Regarding the exchange and integration of statistical information, the number corresponding to a statistical indicator represents the measurement of a concept that has been well-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 can 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 integrate statistical information from different sources, we need to ensure that the data obtained from those different sources pertains to the same concepts. This is the first condition in statistics to have interoperability.

Once we are sure that the same concepts are used, 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, can we 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, while if it is about economic variables, the disaggregation by sex may not be necessary.

The information that can help us to provide a semantic context of the statistics, as in the last examples described, is called structural metadata and 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 known as 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 made 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.

DRAFT

DRAFT

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 these possible problems that arise when they are not present:

- 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 facts. 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 conceptualizations. Our interoperability can be reduced or impeded by this cause.
- 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.

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

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 an Organisational Unit in charge of a certain statistical program 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 an Organisational Division in charge of several statistical programs establishes interoperability for the statistics produced in their programs, then at least statistics produced by the unit 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 program 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

Interoperability has been defined by González & Orell (2018) as a characteristic of good quality data, and it relates to broader concepts of value knowledge creation, collaboration, and fitness-for-purpose. This definition is more oriented to interoperability in data, so a broader definition of interoperability may be "the capacity of systems or services to exchange and make use of the statistical information without prior communication" which will be further explained in Section 2.

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 going to be 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.

DRAFT

DRAFT

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.

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.

An organisation can achieve interoperability only if it is in control of its information. **Data governance** is defined by DAMA (DAMA, 2017) as the exercise of authority and control (planning, monitoring, and enforcement) over the management of data assets. 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. DAMA (DAMA, 2017) defines **data management** 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. 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 the CES Task Force on Data Stewardship (2023), data stewardship is considered an “approach to data governance that formalizes 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 from the points of view addressed by different pieces of information.

The target audience of this document includes 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.

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), these types of information are not in the scope of this document.

[To be completed after other chapters finish] The rest of this document consists as follows: Section 2 delves into the concept of interoperability in more detail….

DRAFT

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 organizing 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 organize and format those descriptions. Similarly, if some process is interoperable, one needs to know the schema used to organize 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.

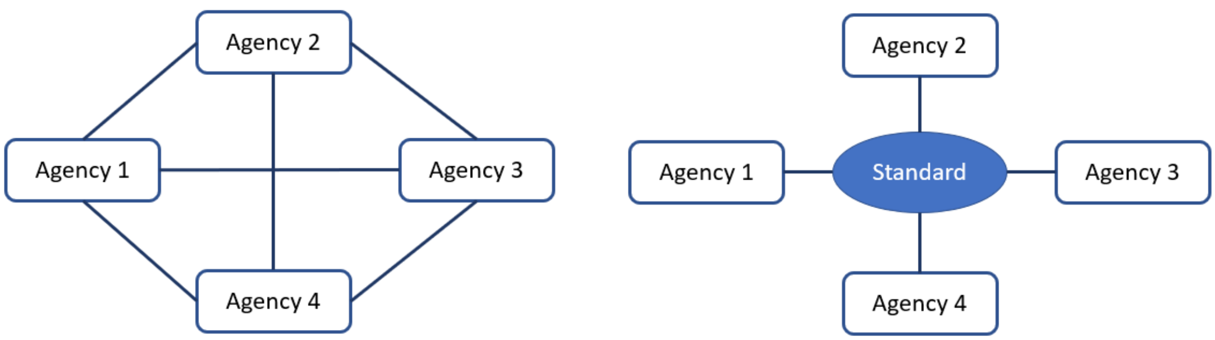
DRAFT

catalogs, 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 [1])*



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 specialized – 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.

DRAFT

To add Section 3 examples of standards for interoperability, see Section 3.3.

Relationship with open data and FAIR

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 focus 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 [2]. 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.

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, first data 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 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[2] and interoperability foster the flow of data between participants of national data systems and enable cross-border data collaboration.

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 the purpose of 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
…	…	…	…	…	…	…

Example 2: in CSV

```
Country, Country code, Region, Population 2000, Population 2022, Average annual population growth, Currency unit
```

Example 3: in JSON

```
[
 {
  "Country": "United Kingdom",
  "Country code": “GB”,
```

[2] 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/

“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”	"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 [3]. 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 authorization, 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.

DRAFT

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 codelist 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.

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

DRAFT

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 GSBPM. The non-Production Activity Areas of 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., codelist, 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" [4]. 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, codelist, 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" [4]. 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.

DRAFT

3. **Build Phase**: “this phase builds and tests the production solution to the point where it is ready for use in the ‘live’ environment” [4]. While many design decisions are made during Design Phase, there are several choices made at the implementation stage that could impact the interoperability. For example, specific data collection system might use different data format 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” [4]. With statistical organisations increasingly involved with sources that are not under their direct control (e.g., administrative data, big data from web), ensuring interoperability becomes even more challenging. Without proper documentation of the data and mapping (e.g., between codelists used by different sources), the risk of introducing non-interoperable elements during this phase significant increases.

5. **Process Phase:** this phase “describes the processing of input data and their preparation for analysis” [4]. In this phase, various processes are applied to data and lack of availability of data provenance information would lead to the 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” [4]. 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” [4]. 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) [5] 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 the organisational roles, the legal and business policies that influence interoperability, the standards and technologies that facilitates it and the steps of the governance process to support interoperability.

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. In order 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.

DRAFT

- *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. [3]
- 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. [4]

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

[4] See ONS Statistical Quality Improvement Strategy at https://www.ons.gov.uk/methodology/methodologytopicsandstatisticalconcepts/qualityinofficialstatistics/onsstatisticalqualityimprovementstrategy

- Information Technology Steering Committee: is responsible for guiding the overall IT strategy within the NSO. It oversees the implementation of IT systems and infrastructure to support data management and interoperability, including decisions on the adoption of standardized 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. [5]
- 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 3.3 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 its specific section is exactly 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 standardization of data and metadata and its re-use within an organization and across organizations.

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 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 1 and Box 2 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 geographic information 2022 – 2046 stablishes a specific goal and a general action related to the

---

[5] See "Data Stewardship and the Role of National Statistical Offices in the New Data Ecosystem", UNECE CES 2023 at https://unece.org/sites/default/files/2023-04/CES_02b_Data_stewardship_summary_for_consultation.pdf

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

DRAFT

consolidation of the interoperability between statistical and geographic 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 in the following section for example from Italy).

Box 1: 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

a) 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;

[…]

e) 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 2: 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) promulgated in 2008 defines the minimum information infrastructure for each national information subsystem: demographic and social, economic, geographic 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. It is made up of catalogs, classifications, statistical and geographic records and methodologies. The use of a common Information Infrastructure, concepts, classifications and methods in the production process facilitates the integration or linkage of information from different statistical programmes.

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 NSIs:

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 NSI 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.

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, NSIs 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

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.

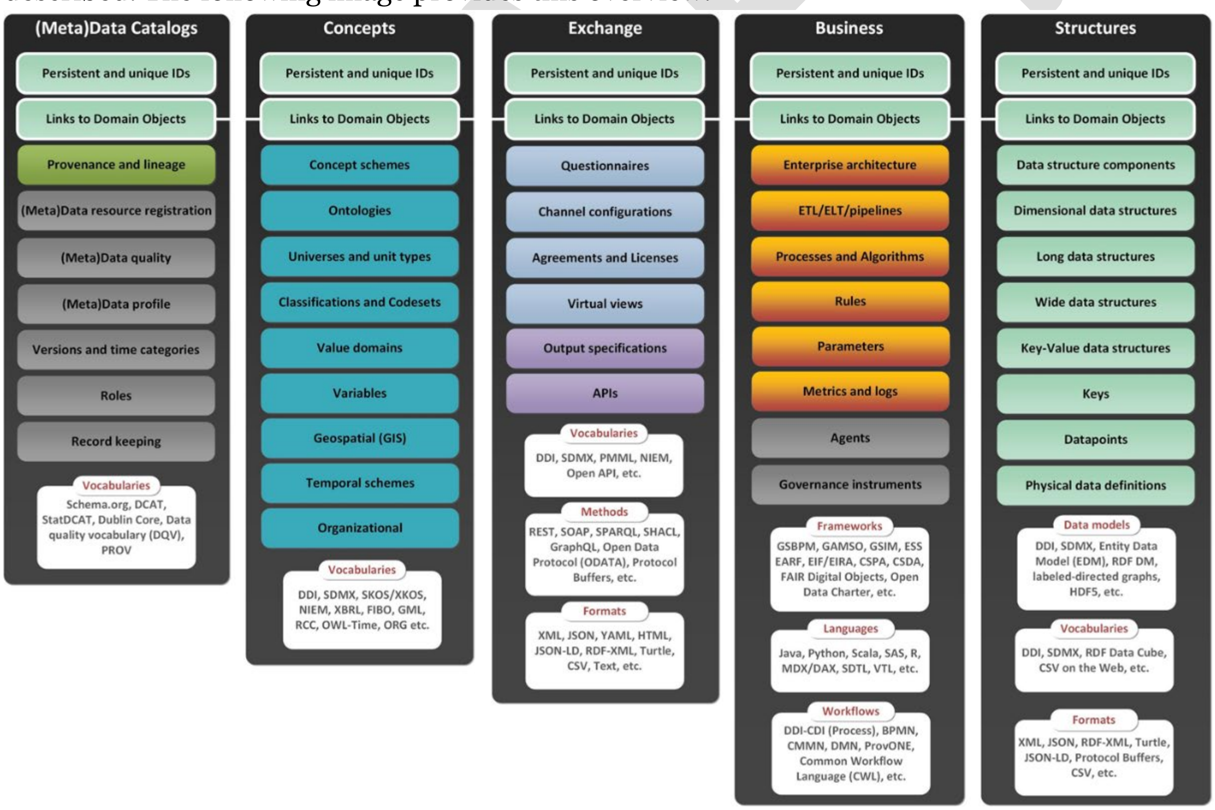
Examples of statistical data and metadata standards:

- 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 Channels*. 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 a number of 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 organized 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.

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

DRAFT

Box 3: 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 Catalog

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 through the use of statistical models and standards existing at an international level (GSBPM, GSIM, etc.), has clearly emerged in the context of the National Data Catalogue and 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.

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 goal is not to be a catalogue for purely documentary purposes, but to provide 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.

It is clear how crucial the aspects of governance and shared rules are before the development of the system.

3.4. Governance Process

A key principle is that the governance process for interoperability should be inclusive and involve all stakeholders, including NSO staff, data users, and other government agencies. 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 the technology is constantly changing, and NSOs need to be able to adapt their governance processes accordingly.

The following are some of the key steps that are typically involved in the governance process for interoperability in NSOs:

1. Define the vision and goals for interoperability. What does the NSO want to achieve through interoperability? What are the benefits that it hopes to achieve?
2. Identify the stakeholders. Who are the people and organisations that will be affected by interoperability? Who needs to be involved in the governance process?
3. Establish a governance structure. This should include a clear definition of roles and responsibilities, as well as a process for making decisions.
4. Develop and implement interoperability standards. These standards should define how different systems and technologies will communicate with each other.
5. Monitor and evaluate interoperability initiatives. This is important to ensure that interoperability is being achieved in a way that is meeting the needs of the NSO and its stakeholders.

By implementing a sound governance process for interoperability, NSOs can improve the quality, efficiency, and effectiveness of their data management processes.

Often the governance includes the participation in national initiatives: here a couple of examples:

- 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

In the following box the governance framework of the Israel Central Bureau of Statistics is described and summarised.

Box 4: Israel Governance Framework
The 'data governance framework' includes three main components:

DRAFT

- Principles for maintaining connectivity, privacy, quality and trust, that are intended to create a common understanding, alignment and coherence of the organizational 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 catalogs, 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 for the production of 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 –

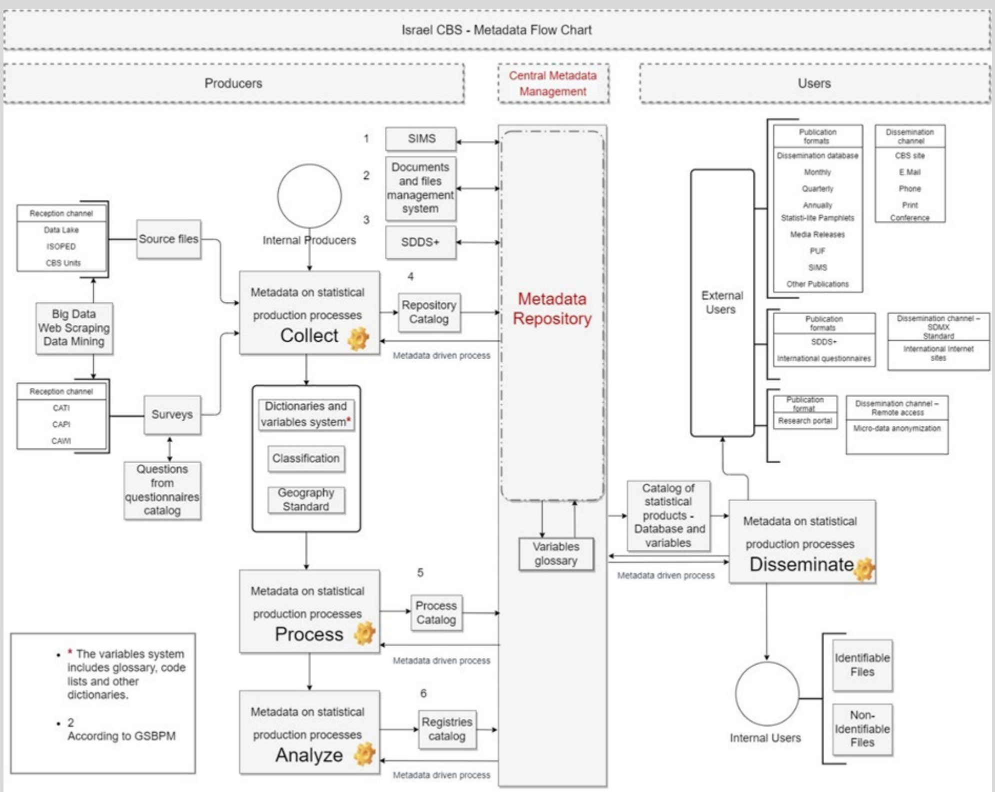
The quality assurance framework is an important part in the Governance Framework of the CBS.

It is constituted of two parts:

The first part describes the organization'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.



4. Recommendations

DRAFT

References – to be rearranged

From section 1

González Morales Luis, Orell Tom; **Data Interoperability: A Practitioner's Guide to Joining Up Data in The Development Sector**; UNSD; 2018 (https://unstats.un.org/wiki/display/InteropGuide/Home)

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

European Union, New European Interoperability Framework; 2017 (https://ec.europa.eu/isa2/sites/default/files/eif_brochure_final.pdf)

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)

From section 2

[1] 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.

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

[3] 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

[4] GSBPM v5.1 (2019)

[5] 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

DRAFT

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 web-pages (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 catalogs 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 on the basis of 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.

DRAFT

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 organizations.
- 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 organizational structures, aimed at supporting linked data publishing of organizational 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.

DRAFT

- 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.

DRAFT

- 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 called simply 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 similar to 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 a number of 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

DRAFT

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.

<u>Structures</u>

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.

DRAFT

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 called simply 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 similar to 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.

DRAFT

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 catalogs

- 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 catalog 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

DRAFT

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

DRAFT

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 organizational elements and between organizations; re-use and standardization of data over time, space, and applications; harmonization and standardization of data within an organization and across organizations; 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 organization 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 organizational roles and responsibilities associated with the administered item registration process are identified and suggested.

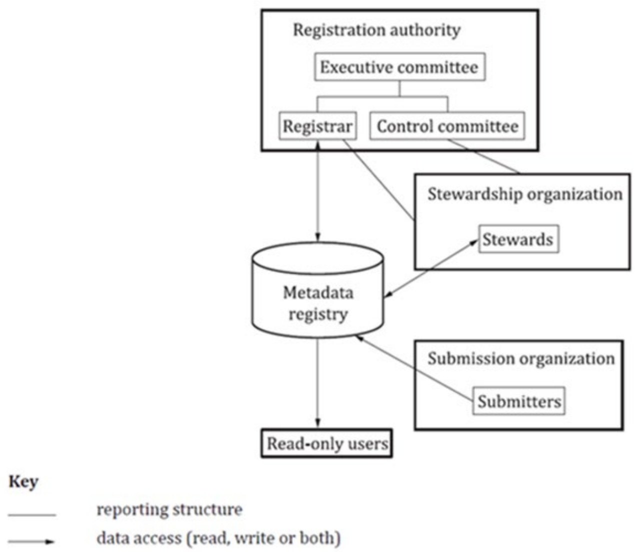
There are three types of registration acting bodies (RAB) in the framework of the ISO/IEC 11179:

- registration authorities
- submitting organizations
- stewardship organizations

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 organizational roles are related within the context of a metadata registry.

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

DRAFT



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).	In order 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.

DRAFT

	- 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, in particular, 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;

DRAFT

	- 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;

DRAFT

	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;

	h) ensuring that data registration procedures and formats are followed within their assigned functional area; i) recommending submitters to the registration authority.
Submitting organizations (SuO)	
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.	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.
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.	**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 registrar; 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.
Others	
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.	
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.	

DRAFT