



**EOSCFAIR**  
Executive Board Working Group



**EOSCArchitecture**  
Executive Board Working Group

## **EOSC Interoperability Framework (v1.0)**

**3 May 2020**

### **Draft for community consultation**

Oscar Corcho, Universidad Politécnica de Madrid, [0000-0002-9260-0753](https://orcid.org/0000-0002-9260-0753)

Magnus Eriksson, Swedish Research Council

Krzysztof Kurowski, Poznań Supercomputing and Networking Center IBCH PAS

Milan Ojsteršek, University of Maribor, [0000-0003-1743-8300](https://orcid.org/0000-0003-1743-8300)

Christine Choirat, Swiss Data Science Center, ETH Zürich and EPFL

Mark van de Sanden, SURFsara

Frederik Coppens, VIB-UGent Center for Plant Systems Biology, [0000-0001-6565-5145](https://orcid.org/0000-0001-6565-5145)

# Executive summary

This document has been developed by the Interoperability Task Force of the EOSC FAIR Working Group, with participation from the Architecture WG.

Achieving interoperability within EOSC is essential in order for the federation of services that will compose EOSC to provide added value for service users. In the context of the FAIR principles<sup>1</sup>, interoperability is discussed in relation to the fact that “research data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing”. Our view on interoperability does not only consider data but also the many other research artefacts that may be used in the context of research activity, such as software code, scientific workflows, laboratory protocols, open hardware designs, etc.

This document identifies the general principles that should drive the creation of the EOSC Interoperability Framework, and organises them into the four layers that are commonly considered in other interoperability frameworks (e.g., the European Interoperability Framework<sup>2</sup>): technical, semantic, organisational and legal interoperability.

For each of these layers, a catalogue of problems and needs, as well as challenges and high-level recommendations have been proposed, which should be considered in the development of the EOSC IF. Such requirements and recommendations have been developed after an extensive review of related literature as well as by running interviews with stakeholders from ERICs, ESFRI projects, service providers and research communities. Some examples of such requirements are: “every semantic artefact<sup>3</sup> that is being maintained in EOSC must have sufficient associated documentation, with clear examples of usage and conceptual diagrams”, or “Coarse-grained and fine-grained dataset (and other research object) search tools need to be made available”, etc.

The document finally contains a proposal for how the management of FAIR Digital Objects should be done in the context of EOSC, so as to provide support to the identified requirements, with the following core properties:

- Basic components of this framework are semantic artefacts, with many of them being common across communities and services
- Semantic artefacts should be available in repositories
- The descriptions of these elements should follow agreed metadata frameworks and elements, appropriate at a generic cross-disciplinary level as well as specific community-based interoperability frameworks.

---

<sup>1</sup> Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3:160018 doi: 10.1038/sdata.2016.18 (2016)

<sup>2</sup> [https://ec.europa.eu/isa2/eif\\_en](https://ec.europa.eu/isa2/eif_en)

<sup>3</sup> Semantic artefact is defined in this document as a machine-actionable and -readable formalisation of a conceptualisation enabling sharing and reuse by humans and machines. These artefacts may have a broad range of formalisation, from loose set of terms, taxonomies, thesauri to higher-order logics

# Table of contents

[Executive summary](#)

[Table of contents](#)

## [1. Introduction](#)

### [1.1 Context and definitions](#)

#### [1.1.1 The European Open Science Cloud \(EOSC\)](#)

#### [1.1.2 FAIR principles and the role of Interoperability](#)

#### [1.1.3 The European Interoperability Framework as a Starting Point for the EOSC Interoperability Framework](#)

#### [1.1.4 Definitions of relevant terms used in this document](#)

### [1.2 Purpose and scope](#)

### [1.3 How to read this document](#)

## [2 Interoperability layers](#)

### [2.1 Technical interoperability](#)

### [2.2 Semantic interoperability](#)

### [2.3 Organisational interoperability](#)

### [2.4 Legal interoperability \(not in v1\)](#)

## [3. Minimum Requirements and Recommendations for the EOSC Interoperability Framework](#)

### [3.1 Technical Interoperability](#)

#### [3.1.1 Problems and needs](#)

#### [3.1.2 Recommendations](#)

### [3.2 Semantic Interoperability](#)

#### [3.2.1 Problems and needs](#)

#### [3.2.2 Recommendations](#)

### [3.3 Organisational interoperability](#)

#### [3.3.1 Problems and needs](#)

#### [3.3.2 Recommendations](#)

### [3.4 Legal interoperability \(not in v1\)](#)

### [3.5 Some general recommendations from the European Interoperability Framework](#)

## [4. Towards an EOSC Interoperability Framework: Model and Components](#)

### [4.1 Model overview](#)

#### [4.1.1 Principles for FAIR digital objects](#)

#### [4.1.2 The European Interoperability Framework and the EOSC IF](#)

#### [4.1.3 Interlinking digital objects](#)

### [4.2 Basic components](#)

[4.2.1 Common semantic artefacts](#)

[4.2.2 Metadata frameworks and elements](#)

[4.2.2.1 The metadata framework core as a part of the foundation for semantic interoperability](#)

[4.2.3 Common resources for semantic artefacts, including examples](#)

[4.2.3.1 Common Semantic artefacts](#)

[4.2.3.2 Conceptual metadata standards & Data type registry models](#)

[Annex I. Interviews with stakeholders](#)

# 1. Introduction

## 1.1 Context and definitions

This section provides some context and general definitions related to this document.

### 1.1.1 The European Open Science Cloud (EOSC)

The European Open Science Cloud (EOSC)<sup>4</sup> is a European Commission initiative aiming at developing an infrastructure providing its users with services promoting Open Science practices.

EOSC aims to support three objectives: (1) to increase the value of scientific data assets by making them easily available to a larger number of researchers, across disciplines (interdisciplinarity) and borders (EU added value) and (2) to reduce the costs of scientific data management, while (3) ensuring adequate protection of information/personal data according to applicable EU rules.

### 1.1.2 FAIR principles and the role of Interoperability

In the context of the FAIR principles<sup>5</sup>, interoperability is discussed in relation to the fact that “research data usually need to be integrated with other data [...] in addition, the data need to interoperate with applications or workflows for analysis, storage, and processing”. The following principles are proposed:

- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (Meta)data use vocabularies that follow FAIR principles.
- I3. (Meta)data include qualified references to other (meta)data.

As discussed in the “Turning FAIR into Reality” report<sup>6</sup>, the role of interoperability frameworks is to “define community practices for data sharing, data formats, metadata standards, tools and infrastructure, recognising the objectives and cultures of different research communities”. And the report also stresses the fact that such frameworks need

---

<sup>4</sup> <https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>

<sup>5</sup> Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3:160018 doi: 10.1038/sdata.2016.18 (2016)

<sup>6</sup> Turning FAIR into Reality. Final Report and Action Plan from the European Commission Expert Group on FAIR Data. 2018. [https://ec.europa.eu/info/sites/info/files/turning\\_fair\\_into\\_reality\\_1.pdf](https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf)

to support FAIR across traditional discipline boundaries and in the context of high priority interdisciplinary research areas.

Achieving interoperability within EOSC is essential in order for the federation of services that will compose EOSC to provide added value for service users, no matter which scientific disciplines they work on. The services within the EOSC will provide value by provisioning digital objects (which refer to the aforementioned research artefacts and whose definition is provided in Section 1.1.4). In order to realize the value of the services, the digital objects exchanged need to be efficiently consumed by other EOSC services and user systems.

In order for the user systems to consume the digital objects provisioned by the EOSC services they must understand how to read and interpret them, what restrictions there are to use the object and what processes are involved in their production and consumption. And this should be independent from the specific scientific discipline where the digital objects were created or are being consumed.

Therefore, it needs to be possible for software/machines to deduce or obtain these characteristics from the information provided by the digital object itself through its metadata. The EOSC interoperability framework aims to provide a set of recommendations on the components that need to be provided in the ecosystem and on the principles guiding digital object producers and/or consumers on their use. This in order for the framework to set a foundation for an efficient machine-enabled exchange of digital objects within EOSC and between EOSC and the outside world. A final aspect to consider in this context is that there will be different degrees of interoperability that will be achievable, especially in interdisciplinary settings.

### 1.1.3 The European Interoperability Framework as a Starting Point for the EOSC Interoperability Framework

The structure of the EOSC IF is inspired by earlier work done for the European Interoperability Framework (EIF)<sup>7</sup>, as well as in the context of other domain-specific interoperability frameworks (e.g., the Shift2Rail Interoperability Framework<sup>8</sup>).

The EIF, promoted and maintained by the ISA<sup>2</sup> programme, targets public administrations in Europe, so that they can design and deliver public services in an interoperable manner, contributing to the development of a single digital market by

---

<sup>7</sup> New European interoperability framework. Promoting seamless services and data flows for European public administrations. Directorate-General for Informatics (European Commission). 2017. DOI: 10.2799/78681

<sup>8</sup> <https://shift2rail.org/research-development/ip4/>

fostering cross-border and cross-sectoral interoperability for the delivery of such European public services.

Therefore, the core target of the EIF (public administrations at all levels, including the national interoperability frameworks, and interactions between administrations - A2A -, administrations and citizens - A2C - and administrations and businesses -A2B-) is somewhat different to the target of the EOSC IF, which is mostly focused on individual researchers, research performing organisations, funding organisations and research infrastructures. However, they share many common underlying principles and core objectives. Indeed, using the EIF terminology, the EOSC IF may be seen as an example of a Domain-specific Interoperability Framework, which in turn focuses on multiple scientific domains.

For that reason, the EOSC IF is structured in a similar manner to EIF. More specifically, the EIF identifies four layers of interoperability (technical, semantic, organisational and legal), as shown in Figure 1, which have been also considered in the development of the EOSC IF, as described in this document.

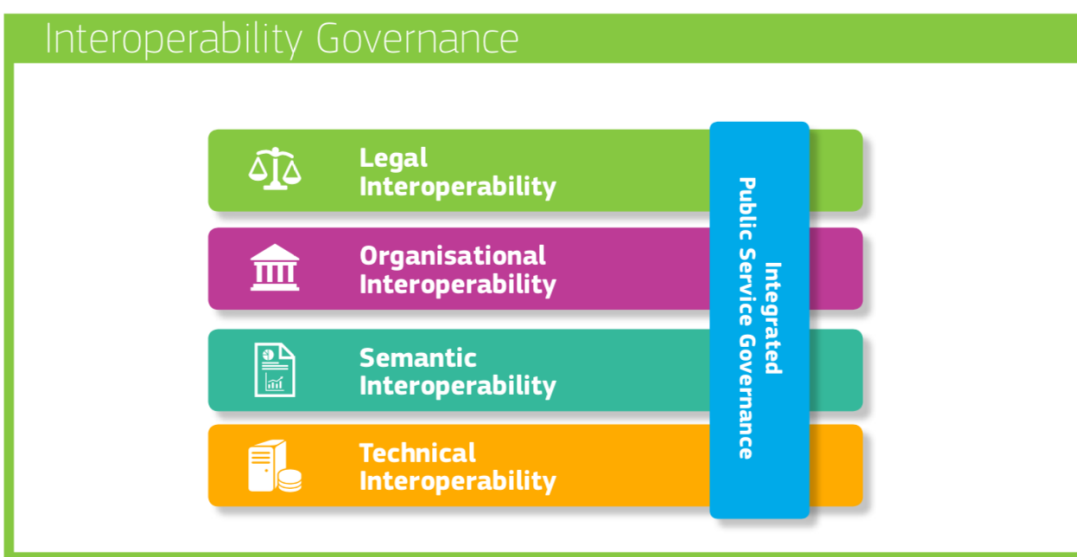


Figure 1. The European Interoperability Framework four levels of interoperability

#### 1.1.4 Definitions of relevant terms used in this document

In this document, we use the term **Digital Object** to refer to the kind of objects that allow binding all critical information about any entity. The information that we are interested in in the context of the EOSC IF includes research data, software, scientific workflows, hardware designs, protocols, provenance logs, publications, presentations, etc., as well as all their metadata (for the complete object and for its constituents).

Examples of Digital Objects that have been proposed in the past are Research Objects<sup>9</sup> and some of its implementations (e.g., RO-Crate<sup>10</sup>, the BagIt specification<sup>11</sup>). Another potential definition of Digital Object is the one provided by the RDA Data Foundation & Terminology (DFT) Core Terms and Model<sup>12</sup>, which states that “a Digital Object is represented by a bitstream, is referenced and identified by a persistent identifier and has properties that are described by metadata”.

We also use the term **metadata** widely. For this, we have decided to choose the ISO11179 definition of metadata, which defines it as "descriptive data about an object". That is, metadata is a kind of data: data becomes metadata when the descriptive relationship is revealed between the data (now metadata) and the target object(s). And metadata that is the same for more than one object is metadata for a class of objects...” (ISO/IEC CD 11179-1). This definition also aligns well with the definition used in the paper on the FAIR Principles, which states that the term “data” is used to refer to all types of digital resources (not just data in the restricted sense, but also, for example, software, workflows, hardware designs, etc.) and metadata is any description of a resource that can serve the purpose of enabling findability and/or reusability and/or interpretation and/or assessment of that resource. In this context, data and metadata may be published together or as different inter-related entities (with their own identifiers), and different blocks of metadata may be associated to the same digital object (as described further in Section 4).

Finally, different definitions around **interoperability** are available in the state of the art. We summarise some of those that we are taking in the context of this document here:

- **Interoperability.** A characteristic of an Information Technology (IT) system, whose interfaces are completely understood, to work with other IT systems, at present or in the future, in either implementation or access, without any restrictions or with a controlled access (source: Interoperability - Wikipedia).
- **Syntactic interoperability.** If two or more systems use common data formats and communication protocols and are capable of communicating with each other using open standards (source: Interoperability - Wikipedia)
- **Semantic Interoperability.** A stronger type of data exchange than typical Interoperability because it includes some knowledge of the meaning of the data content, system structure and operation, usage constraints, and the underlying

---

<sup>9</sup> Khalid Belhajjame, Jun Zhao, Daniel Garijo, Matthew Gamble, Kristina Hettne, Raul Palma, Eleni Mina, Oscar Corcho, José Manuel Gómez-Pérez, Sean Bechhofer, Graham Klyne, Carole Goble (2015) Using a suite of ontologies for preserving workflow-centric research objects, Web Semantics: Science, Services and Agents on the World Wide Web, <https://doi.org/10.1016/j.websem.2015.01.003>

<sup>10</sup> <https://researchobject.github.io/ro-crate/>

<sup>11</sup> <https://tools.ietf.org/html/draft-kunze-bagit-17>

<sup>12</sup> <http://hdl.handle.net/11304/5d760a3e-991d-11e5-9bb4-2b0aad496318>



assumptions. (source: RDA - [https://smw-rda.esc.rzg.mpg.de/index.php/Semantic\\_Interoperability](https://smw-rda.esc.rzg.mpg.de/index.php/Semantic_Interoperability))

And a few additional general definitions are provided next:

- **Cloud** is the on-demand availability of computer system resources, especially data storage and computing power, without direct active management by the user. The term is generally used to describe e-infrastructures available to many users over the Internet. (source: Cloud computing - Wikipedia)
- **Service-oriented architecture (SOA)** is a style of software design where services are provided to the other components by application components, through a communication protocol over a network. The basic principles of service-oriented architecture are independence of vendors, products and technologies (source: Service-oriented architecture - Wikipedia).

## 1.2 Purpose and scope

The EOSC IF is meant to be a generic framework that can be used by all the entities participating in the development and deployment of EOSC, providing a common understanding of the requirements, challenges and recommendations that they should take into account, as well as a general set of principles on how these recommendations may be addressed. The EOSC IF does not propose any specific recommendation on how these recommendations should be actually implemented, although it provides a non-exhaustive list of illustrative examples of how some of them are being addressed.

The different providers of EOSC-related services are also a relevant target for this document, since it provides some general recommendations for achieving interoperability across these services (e.g., interoperability in authentication and authorisation, interoperability in the exchange of data, interoperability for ensuring the findability of resources), enabling multidisciplinary and multi organisational collaborations.

## 1.3 How to read this document

This document is organised in three main sections:

- Section 2 provides a general overview of the interoperability layers already identified in the European Interoperability Framework, and the types of challenges that are being addressed in each of them.
- Section 3 provides a summary of the main problems, needs, challenges, and recommendations at each layer, based on the analysis done on existing literature, plus the results of an extensive set of interviews run with researchers

from different research communities, some of them involved in ESFRI projects and ERICs, as well as service providers.

- Finally, section 4 makes a proposal on how interoperability may be addressed by adopting FAIR digital objects, and the main elements behind this proposal.

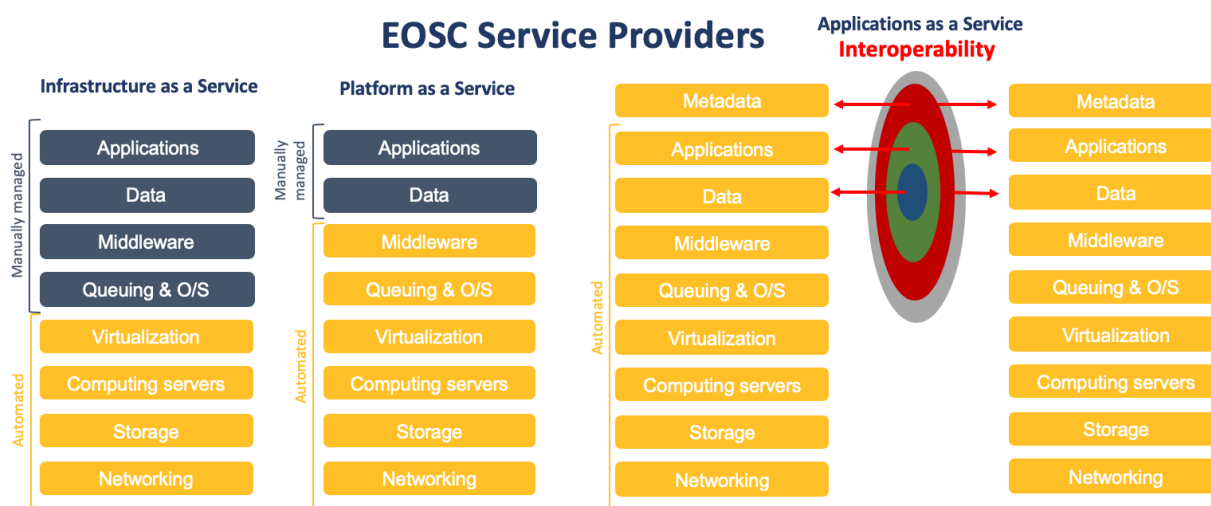
Annex I contains further information related to the interviews that have been performed as a first step towards the creation of this document.

## 2 Interoperability layers

Here we will focus on the four layers identified by the European Interoperability Framework: technical interoperability, semantic interoperability, organisational interoperability and legal interoperability. Each of these will be described in its own subsection below.

### 2.1 Technical interoperability

Technical interoperability is commonly defined as the “ability of different information technology systems and software applications to communicate and exchange data”. This definition may be also completed by adding the “ability to accept data from each other and perform a given task in an appropriate and satisfactory manner without the need for extra operator intervention”. That is, it is sometimes completed with an aspect focused on the complete automation of such data exchange.

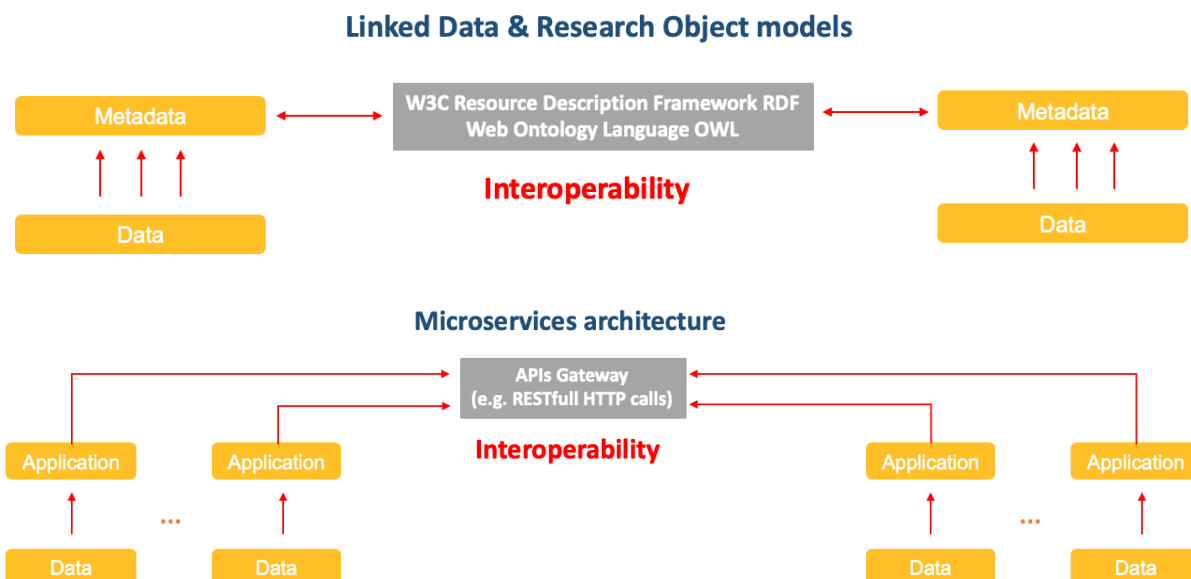


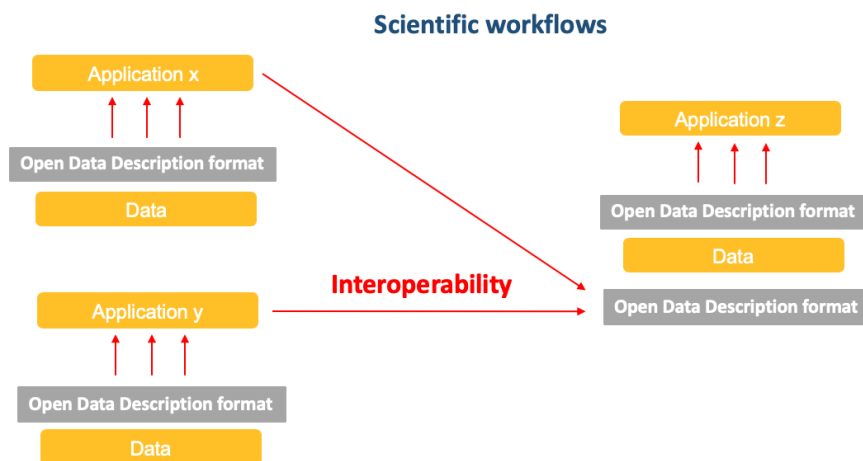
In the context of this document, we are referring not only to the exchange of data (across scientific experiments, organisations or even communities), but also of other research artefacts that are commonly used in research (software, workflows, protocols,

hardware designs, etc.). According to the EIF, technical interoperability covers “the applications and infrastructures linking systems and services, including interface specifications, interconnection services, data integration services, data presentation and exchange, and secure communication protocols”.

In the context of our interviews the aspects related to technical interoperability have arisen in many occasions, not only across communities, but also in the context of the same scientific communities, where for example different systems that are used for the generation of data or for its consumption are not compatible with each other, or where different user identification methods exist for researchers that need to make use of different types of systems. Best practices have also been identified in this context, as a result of our interviews. For example, in the context of radio astronomy, many efforts have been done in the past on the creation of the Virtual Observatory (<http://www.ivoa.net/>), not only as a technical platform for sharing and exchanging data, but also as a set of specifications and standards for the definition of data sources that can be used by researchers, with a clear governance model.

Some examples of technical interoperability aspects and different models for describing data and metadata that have been used in the state of the art are summarised in the following figures:





All of these aspects will be addressed in section 3.1.

## 2.2 Semantic interoperability

Semantic interoperability can be defined as “the ability of computer systems to transmit data with unambiguous, shared meaning. Semantic interoperability is a requirement to enable machine computable logic, inferencing, knowledge discovery, and data federation between information systems”.<sup>13</sup>

That is, semantic interoperability is achieved when the information transferred has, in its communicated form, all of the meaning required for the receiving system to interpret it correctly, even when the algorithms used by the receiving system are unknown to the sending system. Syntactic interoperability (which is commonly associated with technical interoperability) is sometimes identified as a prerequisite to semantic interoperability. It ensures that the precise format and meaning of exchanged data and information is preserved and understood throughout exchanges between parties, in other words ‘what is sent is what is understood’.

In the context of our interviews, aspects related to semantic interoperability have also arisen in many occasions, mainly related to the need to have common metadata formats across communities and services so that the interpretation of the data is made easier, as well as shared semantic artefacts (ontologies, thesauri) across the communities, which allow homogenising the interpretation and treatment of the exchanged data, and all of its associated resources. Best practices have been also identified, for instance in the case of CESSDA or in many cases in Life Sciences (e.g.

<sup>13</sup> FAIRsFAIR deliverable D2.1 Report on FAIR requirements for persistence and interoperability 2019. <https://zenodo.org/record/3557381>

Genomics), where community-based repositories of semantic artefacts are being maintained, with a clear governance process.

## 2.3 Organisational interoperability

According to the EIF, organisational interoperability refers to the way in which organisations align their business processes, responsibilities and expectations to achieve commonly agreed and mutually beneficial goals. This type of interoperability is also focused on meeting the requirements of the user community by making services available, easily identifiable, accessible and user-focused.

Considering the overall agreed goal of Open Science that underlies all the activities at EOSC, this level of interoperability should be focused on the documentation, integration or alignment of the processes of different organisations providing services in EOSC, so as to ensure that researchers can reach their Open Science goals.

In the context of our interviews, this is the aspect that has been less discussed, possibly because most research communities are already accounting for the need to align to the overall goals for Open Science that EOSC is looking for. It seemed that most of the interviewees understood the current impediments in their communities (additional work required to register their artefacts as Open Science-enabled ones and provide sufficient metadata, lack of recognition for this additional work, both from institutions and colleagues, lack of commonly agreed principles across funding agencies and organisations with respect to the Open Science approach, etc.).

## 2.4 Legal interoperability (not in v1)

<<to be completed in v2 of this document>>

Legal interoperability covers the broader environment of laws, policies, procedures and cooperation agreements needed to allow the seamless exchange of information between different organisations, regions and countries.

# 3. Minimum Requirements and Recommendations for the EOSC Interoperability Framework

This section presents some of the usual problems and needs that are being faced by the user communities targeted by EOSC, as well as by those aiming at providing services for EOSC. These problems and needs are structured according to the interoperability layers that have

been used in previous sections (technical, semantic, organisational and legal), and can be understood as requirements for the EOSC IF.

They have been compiled through a literature review of common types of requirements reported (including key documents such as the RDA FAIR data maturity guidelines<sup>14</sup> or the aforementioned FAIRsFAIR report on FAIR requirements for persistence and interoperability 2019<sup>15</sup>), as well as through the series of interviews that have been run by the Interoperability task force members of the EOSC IF Working Group.

Besides, this section compiles a set of recommendations organised by these levels.

## 3.1 Technical Interoperability

### 3.1.1 Problems and needs

At the level of technical interoperability, some of the usual problems identified by the communities that have been consulted and by ongoing work on other working groups are the following:

- When trying to work with infrastructures or services across communities, **authentication and authorisation often needs to be performed separately for each community/service**. Even though there are technical means and industry-based standards (e.g., SAML, OpenID) to overcome this, authentication often involves transferring personal information between identity provider and service provider, and authorisation is very hard to harmonise based on centrally-maintained user attributes.
- Research data may be made **available in multiple general-purpose formats** (CSV, Excel, database dumps, JSON, XML, shapefiles, etc.) **or community-based models** (Darwin Core, VOTable and VOResource, FITS, NetCDF), which are usually hard to align when reusing datasets across communities. In the case of general-purpose formats, semantic interoperability problems also appear because of the lack of agreement in attributes or column headers, the absence of headers or adequate documentation, etc.
- Coarse-grained or fine-grained research data from other communities may be difficult to find, given the **lack of knowledge about how to query their repositories**.

---

<sup>14</sup> [https://docs.google.com/document/d/1pDGGL3-BbBJu18KlfZUI3AizKLHXGXdli\\_mPtpEWmeg/edit#](https://docs.google.com/document/d/1pDGGL3-BbBJu18KlfZUI3AizKLHXGXdli_mPtpEWmeg/edit#)

<sup>15</sup> <https://zenodo.org/record/3557381>

- Multiple service providers for different types of PIDs exist (e.g., IUPAC International Chemical Identifier<sup>16</sup>, DOI<sup>17</sup>, PURL<sup>18</sup>, Life Science Identifiers<sup>19</sup>, handle<sup>20</sup>, IVOA<sup>21</sup>, RRID<sup>22</sup>). As a result different sets of policies are enforced to varying degrees, and sometimes the identifiers are not resolvable (e.g., IUPAC InChi-KEY is a reverse identifier: given the chemical, the identifier can be generated, but not in the opposite direction).

As a result of this analysis, these are some of the needs that can be identified at the level of technical interoperability

- There is a need for **support for the process of authenticating to and obtaining the rights to use** the services offered by EOSC in a way that is as unobtrusive as possible [Reference: Architecture WG Authentication and Authorization Infrastructure (AAI) principles] and that is independent of any single community.
- There is a need for EOSC to provide a **trust (and sustainability) framework** across scientific communities, collaborations and infrastructures. For the user this means that what works today will work tomorrow, only better [Reference: Architecture WG AAI principles].
- There is a need for simpler tools that allow dealing seamlessly with data **available in multiple generic or community-based formats**.
- When searching for research data (or other research objects) that may be reusable across communities, such **data may need to be discovered at different levels of granularity**: high level / coarse-grained (e.g., look for data about DNA sequences or land-use) or low level / fine-grained (inside data collections, e.g., look for a specific DNA sequence or land-use in Hamburg).
- There is a need to have a **common and well-understood PID policy** across communities [Reference: PID Technical Architecture charter].

### 3.1.2 Recommendations

Some of the recommendations that can be made in this respect are:

- Use open specifications, where available, to ensure technical interoperability when establishing EOSC services.

---

<sup>16</sup> <https://www.inchi-trust.org/>

<sup>17</sup> <https://www.doi.org/>

<sup>18</sup> <https://sites.google.com/site/persistenturls/>

<sup>19</sup> <https://fairsharing.org/bsg-s001184/>

<sup>20</sup> <http://handle.net/>

<sup>21</sup> <http://www.ivoa.net/documents/IVOAIdentifiers/index.html>

<sup>22</sup> <https://scicrunch.org/resources>

- Define a common security and privacy framework and establish processes for EOSC services to ensure secure and trustworthy data exchange between all involved parties.
- There should be an AAI process for EOSC that is common across communities, easy to implement by resource providers and easy to understand by users.
- The Service-Level Agreements for all EOSC resource providers should be easy to understand by users from different communities.
- EOSC must enable easy access to data sources available in different formats, either generic or community-based, to facilitate overcoming their heterogeneity and allow integrating data across communities, and to tools enabling the usage of these data.
- Coarse-grained and fine-grained dataset (and other research object) search tools need to be made available. There will be a range of general-purpose and domain-specific/specialised search tools, exploiting general-purpose and domain-specific metadata.
- There should be a clear EOSC PID policy, accommodating any appropriate PID usage, recognising that established practises are at different levels of maturity for different resources and new PID types may emerge.

## 3.2 Semantic Interoperability

### 3.2.1 Problems and needs

At the level of semantic interoperability, some of the usual problems that are identified by the communities that have been consulted are the following:

- There is a generalised **lack of common explicit definitions** about the terms that are used by user communities. This is especially a problem in the case of trying to share resources across communities.
- Not only term definitions are usually lacking, but also **common semantic artefacts across communities** (e.g., general ontologies that can be shared). And in case that they exist, these artefacts may not be sufficiently well documented.
- The previous problem is exacerbated by the fact that there is a generalised **lack of common reference repositories** or registries of semantic artefacts (e.g., ontology catalogues). Only some communities are actively maintaining such



resources (e.g., Schema.org<sup>23</sup>, BioPortal<sup>24</sup>, Agroportal<sup>25</sup>, CESSDA's Thesaurus Manager System)

- Data collections are usually poorly documented, in terms of the metadata that is made available for them. Besides, there is **no common metadata schema across communities**, what results in different ones being used in different communities (e.g., DCAT, DDI4, DataCite, DarwinCore, RDA Metadata Directory<sup>26</sup>, FAIRSharing<sup>27</sup>)
- Depending on the discipline, there is a **lack or over-abundance of metadata models** that allow the description, functional preservation and ultimately re-use of the data stored.
- In some communities, there is **lack of expertise and skills related to semantics**, what influences negatively in the availability and use of common definitions, semantic artefacts, reference repositories, etc. This aspect is sometimes known as the “human interoperability” problem.

As a result of this analysis, these are some of the needs that can be identified at the level of semantic interoperability

- Need for **principled approaches and tools for ontology and metadata schema** creation, maintenance, governance and use. Different communities are using different tools and representation models for their semantic artefacts. It is not uncommon to see UML models being used as standardised models for such representation, lacking sometimes the needed formality to describe terms and their relationships.
- Need for **harmonisation across disciplines**. It should be possible for a user of one community to add metadata to existing items (data and semantic artefacts) according to their own research discipline practices (e.g., a social scientist can add DDI-based metadata for a dataset coming from an environmental scientist). Allow a researcher from a discipline to transform metadata (or data) from one discipline's format/annotations to another.

---

<sup>23</sup> Three communities are relevant in this context: **Libraries** (<https://bib.schema.org/>) – they have produced several classes and properties from library and information science; **Archives** (<https://www.w3.org/community/architypes/>) - their proposal for additional classes can be found on [https://www.w3.org/community/architypes/wiki/Alternative\\_1\\_model\\_proposal](https://www.w3.org/community/architypes/wiki/Alternative_1_model_proposal); **Health and medicine** (<https://bioschemas.org/>) - Bioschemas aims to improve the findability of data in the life sciences, some types and properties are available on <https://bioschemas.org/types/> and another link is <https://www.w3.org/community/schemed/>

<sup>24</sup> Bioportal - <https://bioportal.bioontology.org/>

<sup>25</sup> Agroportal - <http://agroportal.lirmm.fr/>

<sup>26</sup> RDA Metadata standard directory - <http://rd-alliance.github.io/metadata-directory/standards/>

<sup>27</sup> Fairsharing.org - <https://fairsharing.org/standards/>

- Need to **harmonise the same type of data** (e.g., observational data in environmental sciences as being done in the I-ADOPT RDA WG, a consistent coding for geographical locations where a sample was obtained, etc.).
- Need for **federated access over existing research data repositories** (both inside a discipline and across disciplines). How to support discovery of data on the basis of a high-level description, and possibly also on more details like concepts related to observations and variables?

### 3.2.2 Recommendations

Some of the recommendations that can be done in this respect are:

- All communities should be generating **clear and precise definitions for the concepts** that they use, as well as their metadata and data schemas. These definitions should be publicly available, referenced by a persistent identifier and shared in EOSC. Furthermore, a classification for research disciplines (e.g., DFG's subject area classification) should be also explicitly created and shared.
- Every semantic artefact that is being maintained in EOSC must have **sufficient associated documentation**, with clear examples of usage and conceptual diagrams. Furthermore, any semantic artefact should also be FAIR.
- Semantic artefacts should be **preferably open** (e.g., like in W3C).
- EOSC should provide support for the **maintenance of a repository of semantic artefacts, and a governance framework** for such a repository. For example, SKOS thesauri may be maintained using services similar to the CESSDA Vocabulary Service.
- There should be **extensibility options** to allow for disciplinary metadata that is typical for some research communities, allowing users/researchers to add annotations according to the established practices of their communities, if relevant (e.g., a social scientist adding DDI-based metadata on a GIS dataset that has only geographical-oriented metadata), with sufficient provenance information on the annotations, and with versioning support.
- A simple vocabulary should be proposed for allowing **discovery over existing federated research data and metadata** (extension of DCAT-AP, DDI 4 Core, or DataCite core schema). There should be some alignment among them, and maybe this should be layered/prioritised
- Not only data should be considered in this context, but also this should be extensible to other types of resources used in Science, such as **software, methods, scientific workflows, laboratory protocols, hardware designs**, etc.
- There should be clear protocols and building blocks for the **federation/harvesting of semantic artefacts catalogues**.

## 3.3 Organisational interoperability

### 3.3.1 Problems and needs

At the level of organisational interoperability, some of the usual problems that are identified by the communities that have been consulted, as well as taking into account some of the initial material generated by the Rules of Participation working group, are the following:

- There is not yet (although it is expected soon) a clearly-defined governance structure for EOSC that includes the **governance framework that will deal with interoperability across organisations and disciplines**, among many other aspects.
- There is not yet a **clear description of the “terms and conditions” and “acceptable use policies” that will rule the services provisioned by EOSC**, and most specifically in what respects to the management of interoperability aspects (e.g., how will metadata services be ruled, the governance of metadata schemas and other semantic resources, etc.).
- The current draft of the Rules of Participation does not enter into the details of how interoperability will be achieved across organisations and user communities in the context of EOSC.
- It is not always clear for users whether the infrastructures or services that they can use from other communities will be still running in the medium or long-term, because of lack of knowledge about their sustainability policies or because of long-term funding plans for the services.

As a result of this analysis, these are some of the needs that can be identified at the level of organisational interoperability:

- Need for a **clear governance framework** that includes clear instructions on how the other levels of interoperability will be handled across organisations and user communities (data formats, AAI services, metadata schemas, ontologies, etc.).
- Need for **documents explaining terms and conditions and acceptable use policies for services providing interoperability**. For instance, providing clear descriptions of the service-level agreements of those providing catalogues and registries of semantic artefacts, or providing systems to overcome semantic differences between different data sources, or alignments between models.
- Need for **interoperability certification mechanisms for service providers**, so that service users can set their own expectations about the support for interoperability of those services.

### 3.3.2 Recommendations

Some of the recommendations that can be done in this respect are:

- The **current set of rules of participation recommendations should be completed with aspects related to interoperability**. For instance, for data providers this may include asking explicitly that data is published according to specific data formats and/or vocabularies for a specific community.
- The same is applicable to **services**, which may be recommended to ingest or output data according to such standardised data formats and/or vocabularies, and with their corresponding metadata, with some level of quality.

## 3.4 Legal interoperability (not in v1)

## 3.5 Some general recommendations from the European Interoperability Framework

We have also included in this section some general recommendations extracted from the European Interoperability Framework<sup>28</sup>, which we consider are applicable to the EOSC IF with some adaptations. These have been included as part of the more specific recommendations in the previous sections, and are maintained in this separate section to facilitate tracing back to the original EIF proposals:

- Ensure that national interoperability frameworks and interoperability strategies are aligned with the EOSC IF and, if needed, tailor and extend them to address the national context and needs.
- Publish research outputs openly unless certain restrictions apply (*“as open as possible, as closed as necessary”*).
- Use open source software. If you implement your own software for data generation, presentation or analysis, publish it as open source with the most appropriate open source license, and make sure that it is well developed and documented.
- Give preference to open specifications, taking due account of the coverage of functional needs, maturity and market support and innovation.
- Secure the right to the protection of personal data, by respecting the applicable legal framework.
- Reuse and share solutions (e.g. software components, Application Programming Interfaces, standards), and cooperate in the development of joint solutions when implementing EOSC services.

---

<sup>28</sup> New European interoperability framework. Promoting seamless services and data flows for European public administrations. Directorate-General for Informatics (European Commission). 2017. DOI: 10.2799/78681

- Reuse and share information and data when implementing EOSC services, unless certain privacy or confidentiality restrictions apply.
- Ensure that all EOSC services are accessible to all research organizations, researchers, citizens, including persons with disabilities, the elderly and other disadvantaged groups. EOSC services should comply, as much as possible, with e-accessibility specifications that are widely recognised at European or international level.
- Ensure data portability, namely that data is easily transferable between systems and applications supporting the implementation and evolution of EOSC services without unjustified restrictions, if legally possible.
- Use multiple channels (physical and digital) to provide the EOSC services, to ensure that users can select the channel that best suits their needs.
- Put in place mechanisms to involve users in analysis, design, assessment and further development of EOSC services.
- As far as possible under the legislation in force, ask users of EOSC services once-only and relevant-only information, if possible, respecting regulations such as GDPR.
- Use information systems and technical architectures that cater for multilingualism when establishing an EOSC service. Decide on the level of multilingualism support based on the needs of the expected users.
- Formulate a long-term preservation policy for information related to EOSC services and especially for information that is exchanged across borders. To guarantee the long-term preservation of digital records and other kinds of information, formats should be chosen to ensure long-term accessibility, including preservation of associated digital signatures or seals.

## 4. Towards an EOSC Interoperability Framework: Model and Components

This section describes our proposal for the design and implementation of the EOSC IF. It discusses first the proposed model for the description of digital objects to be maintained and shared in EOSC, and then proceeds with a further description of the basic components of such a digital object model.

It is important to mention that this document does not provide a concrete recommendation on how such digital objects should be implemented, as this is out of the scope of this document, but only general guidelines to be followed by potential implementations.

### 4.1 Model overview

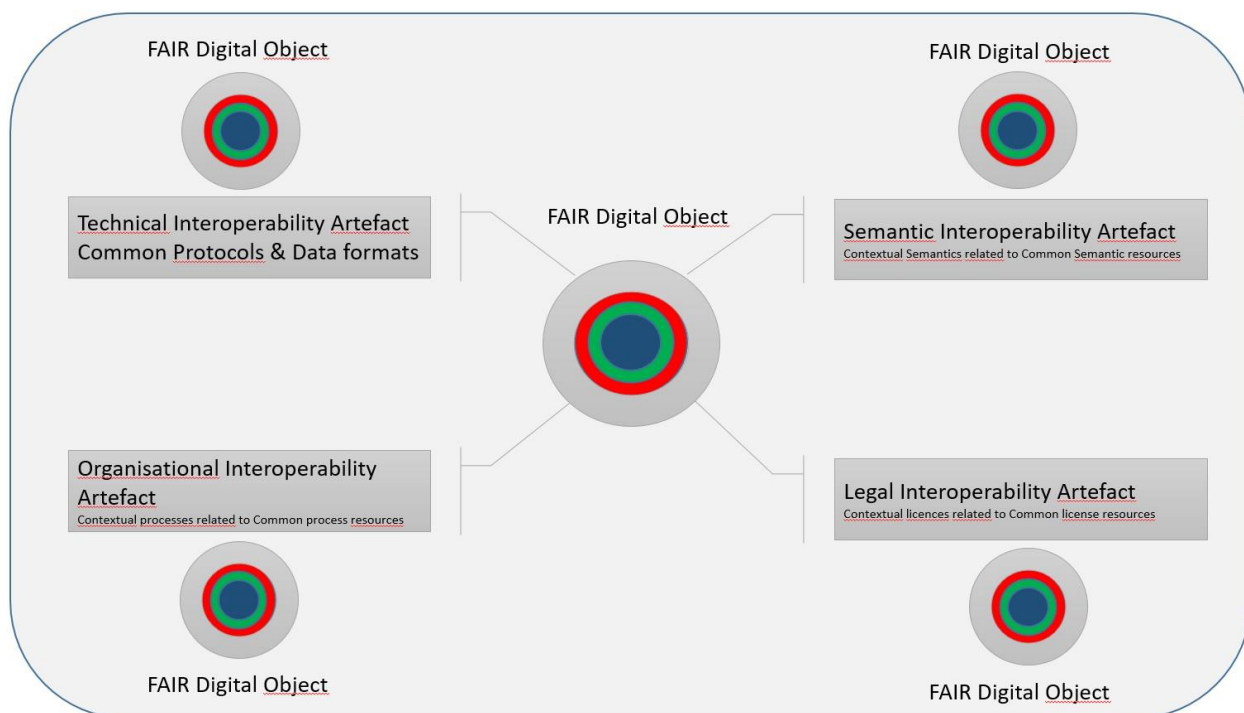
In order for EOSC service providers and service consumers software to form a consensus on how the digital objects are to be read, interpreted and used they need to be able to use an agreed upon set of references to common resources describing these different aspects.

#### 4.1.1 Principles for FAIR digital objects

At the core of the EOSC interoperability framework we find the concept of FAIR digital object, which is described in the EC report “Turning FAIR into reality” as “the atomic entity for a FAIR ecosystem”<sup>29</sup>. The FAIR digital object metadata section will be essential in providing the elements needed to achieve different degrees of interoperability within the FAIR ecosystem. Below we visualise the links between the metadata needed for interoperability.

---

<sup>29</sup> <https://doi.org/10.2777/1524>, p39



The PID links held in the metadata section of the FAIR digital object resolves into FAIR digital objects themselves in order to provide value in the ecosystem and provides the metadata needed to provision:

- Technical interoperability.
- Semantic interoperability.
- Organizational interoperability.
- Legal interoperability.

This perspective and the need for a PID infrastructure, see PID policy for EOSC<sup>30</sup>, supporting it, is described in the paper “Digital Objects as Drivers towards Convergence in Data Infrastructures”<sup>31</sup>.

The framework based on FAIR digital objects with PID links to common artefacts addresses problems expressed by the community during the interviews, see section 3 above. The general lack of common explicit definitions that describe the data to be exchanged in a machine readable way can then be met by linking, with a persistent identifier, to a common semantic artefact that is shared within the EOSC.

<sup>30</sup> <https://doi.org/10.5281/zenodo.3574203>

<sup>31</sup> <http://doi.org/10.23728/b2share.b605d85809ca45679b110719b6c6cb11>

#### 4.1.2 The European Interoperability Framework and the EOSC IF

As discussed throughout this document, the aspects of interoperability described above maps to the European Interoperability Framework's four layers of interoperability - EIF and are realized by resolving the references provided by the FAIR digital object.



European Interoperability Framework<sup>32</sup>

The PID links need to point to common FAIR resources, such as machine readable licenses, semantic artefacts, technical and organisational artefacts, that all resolve into metadata that can be commonly used by the EOSC service providers and service consumers to reach interoperability.

When conducting interviews we learned that “human interoperability” sometimes was an overlooked perspective that related to the common use of resources for interoperability such as metadata standards, terminologies/ontologies, licenses among others. Although the services provide machine readable representations of the different artefacts the people setting up mappings to, or using, metadata standards, concepts, licenses etc. often have different grounds for interpreting them and how the work should be done.

We would therefore like to lift the perspective of human interoperability and the common FAIR resources needed to build the skills and competence needed to set a common ground for shared FAIR resource usage.

The EIF also point to the lack of skills/competence needed to enable interoperability as “a barrier to implementing interoperability policies”<sup>33</sup>, common FAIR resources to build skills and competence can contribute to remedy this.

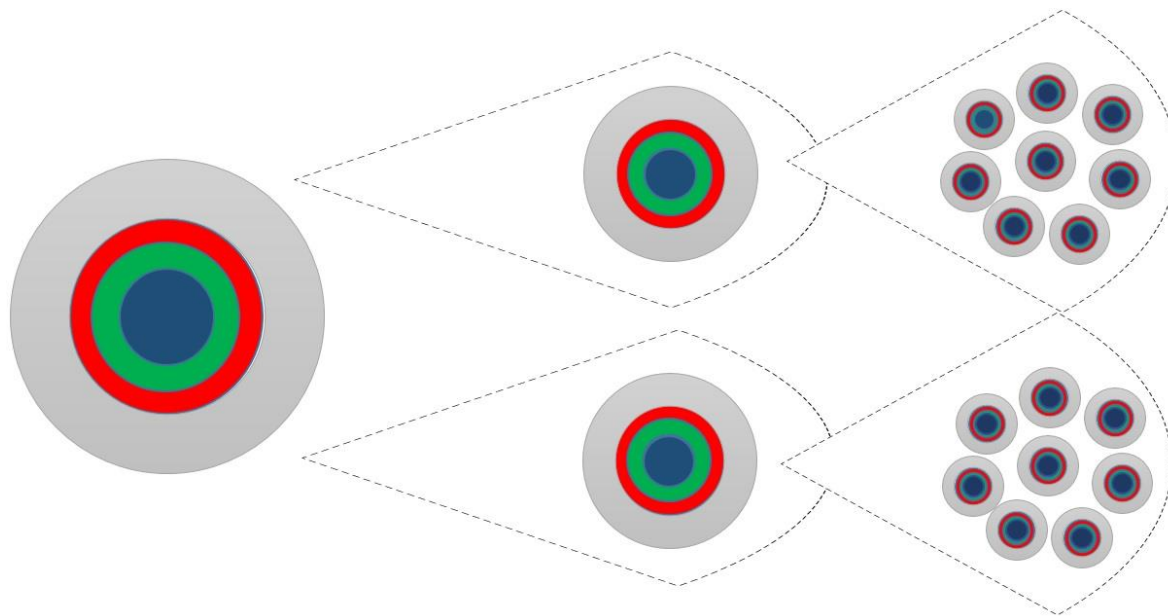
<sup>32</sup> <http://doi.org/10.2799/78681>

<sup>33</sup> <http://doi.org/10.2799/78681>, p23



### 4.1.3 Interlinking digital objects

The FAIR digital object that acts as “the atomic entity for a FAIR ecosystem”<sup>34</sup> can exist on several levels of granularity. One example can be a variable acting as a FAIR digital object in its own right with references to the FAIR resources needed for interoperability but also being a part of a dataset that is a FAIR digital object on a less granular level.



If the EOSC service provides a digital object, the interoperability PID links will thus be provisioned with the object regardless of its granularity.

## 4.2 Basic components

### 4.2.1 Common semantic artefacts

In order to reach semantic interoperability the FAIR Digital Objects, and the metadata elements describing them, need to express and interpret semantic content in a clear and machine-readable way. PID links to common concepts within semantic artefacts is a way of enabling this. Regardless of the data or metadata elements designations and how they are structured common semantic PID links can provide the semantic mapping needed.

---

<sup>34</sup> <http://doi.org/10.2777/1524>, p39

Today there are two large groups of semantic artefacts

- Formally represented: Ontologies and thesauri
- Less formal: UML models, database models, XML schemas etc.

Less formal semantic artefacts can, treated as FAIR digital objects, provide PID links to more formal semantic artefacts to increase the level of interoperability. A UML model containing a class with the name “Person” can in this scenario provide a machine readable link to a concept defining the meaning of “Person” more precisely than what is possible to deduct from the class references in the UML model .

#### 4.2.2 Metadata frameworks and elements

There are many different domain specific metadata standards in use and since the domains differ the metadata standards also differ.

Different domain specific metadata standards sometimes map to conceptual metadata standards/frameworks, or selections thereof. The frameworks/conceptual metadata standards then describes how the metadata should be described and what metadata concepts to be used but not what to describe since that differs between domains. This is used as a way to establish commonalities for increased interoperability, and a common language for data semantics, acknowledging differences between domains.

Examples of this are:

- Data documentation initiative, that is used within CESSDA, that maps to GSIM.
- HL7FHIR, that is used for exchanging health data, that maps to ISO11179<sup>35</sup>

The examples of frameworks/conceptual metadata standards/data type registry model presented below provide a mechanism for interpreting and expressing semantic content about an object and its representation(-s).

In the Generic Statistical Information Model (GSIM)<sup>36</sup> case, it could be a variable (e.g., diagnosis) that inherits its meaning from a Concept (e.g., a cancer diagnosis), that have two representations (e.g., two different codelists for categorising different kinds of cancer diseases). These representations in turn also inherit their meaning from a concept and this also applies to the provided diagnosis codes. The (meta)data elements designation can then differ but the semantic interoperability can be evaluated by

---

<sup>35</sup> ISO/IEC JTC1 SC32 WG2,ISO/IEC 11179

<sup>36</sup> Generic Statistical Information model (GSIM):Specification, United Nations Economic Commission for Europe (UNECE)

comparing the referenced concepts on each level of granularity, that is on the variable level, the representation level and the level of the code used in the codelist.

In the ISO11179<sup>37</sup> standard a similar mechanism is constructed using the data element and the data element concept, where the relation between them provides a mechanism for semantic mapping.

Analogous to the example above the data element concept (e.g., cancer diagnosis) may be related to two data elements (e.g., different cancer variables) using two different codelists for representation. Different designations and representations can, in a similar way as in the earlier example, be semantically compared by comparing the linked concepts on each level of granularity.

The mechanism as presented above is also enabled by the RDA Data Type Registries Model that was presented in the RDA WG output<sup>38</sup>. In the work continued by the RDA WG more detailed models are further describing this in the context of data type registries see “Documentation of Data Representations - A proposed scheme for documenting data structures and vocabularies for machine applications”<sup>39</sup>. This work maps well to ISO11179 and GSIM.

Domain specific and community driven metadata standards that are not mapping to a framework/conceptual metadata standard/data type registry model today can progress towards improved interoperability by mapping to one. Implementation of a semantic mapping mechanism and linking to common concepts will support progress towards higher levels of interoperability.

#### 4.2.2.1 The metadata framework core as a part of the foundation for semantic interoperability

We propose that the core mechanism provided in the frameworks/conceptual metadata standards/data type registry model described above sets the foundation also for the semantic interoperability section of the EOSC Interoperability framework.

We then get a way to manage the relation between semantic artefacts on different levels of formality and, more important, a foundation for mapping FAIR digital objects and its different metadata elements to common semantic artefacts independent of how the metadata elements are structured into different standards.

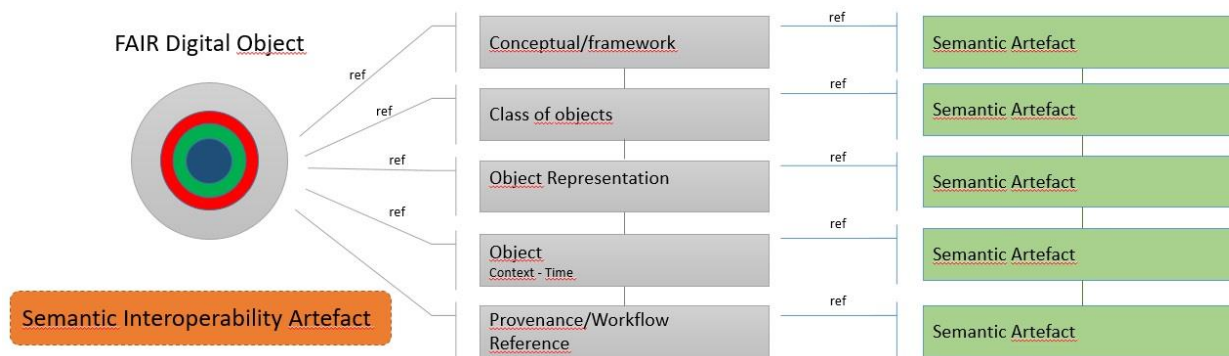
---

<sup>37</sup> ISO/IEC JTC1 SC32 WG2,ISO/IEC 11179

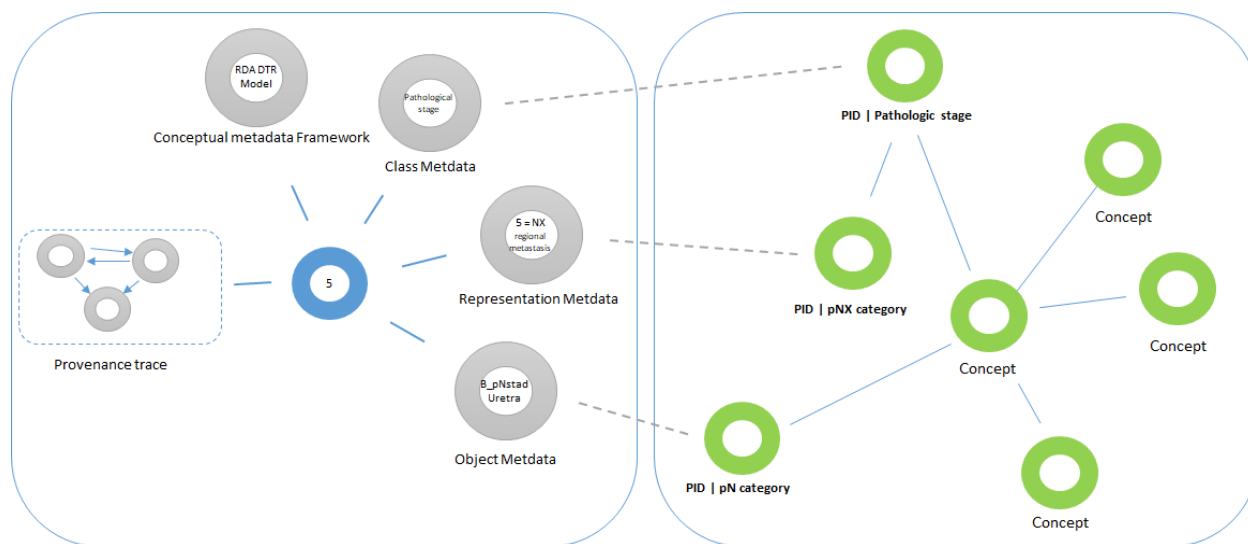
<sup>38</sup> <http://doi.org/10.15497/A5BCD108-ECC4-41BE-91A7-20112FF77458>

<sup>39</sup> <https://github.com/usgin/usginspecs/raw/gh-pages/DataTypeModelDraft.pdf>

These common metadata PID links should be included in the FAIR digital object metadata section and it's semantic interoperability artefact. A link to the conceptual framework/metadata standard used should also be included in order to be able to establish a common interpretation of the object to be exchanged.



In the figure above the commonalities of the frameworks/conceptual metadata standards/data type registry models have been extracted into a simplified set of metadata elements as an example and the references to common semantic artefacts is highlighted. The Provenance/Workflow reference is an important addition that provides input to the context being evaluated and also, of course, is essential for evaluating trust and reusability.



Example Bladder Cancer health data

The elements of the simplified model are described below and examples provided above:

### Metadata links to Conceptual metadata standards/frameworks

- The first level of linkage is to the conceptual framework used in order to be able to evaluate the commonalities of the metadata and semantic artefacts in the context intended. On this initial level this can be a conceptual framework used within a scientific domain or across domains. What classes of objects, types of representation etc that is used to describe the digital object and what Semantic Artefacts providing the semantic meaning often differ between domains. The conceptual metadata standard thus provides the semantic artefact, concepts and context within which we will interpret the metadata elements used to describe the digital object.

### **Metadata references to semantics describing classes of digital objects**

- In search for enabling (machine readable) evaluation of the commonality between digital objects needed for interoperability we start by providing a metadata reference to the concept(-s) defining the common characteristics for the class, that is the collection of digital objects, that are to be exchanged.

### **Metadata references describing the representation(-s) describing the class**

- A class of objects can be represented in different ways and the representation used will be valuable metadata when establishing commonalities.

### **Metadata references to semantics describing the object**

- The metadata references to the concept(-s) describing the actual digital object to be exchanged need to contain additional contextual concepts e.g defining temporal and spatial aspects.

### **Metadata references to semantics describing the provenance trace**

- The object, containing temporal, spatial and other contextual aspects have provenance metadata describing how it was created, from what and by whom. This provenance metadata is essential in order to evaluate the reusability of the object and should be described in a machine readable way. This will enable a programmatic evaluation of trust.

### 4.2.3 Common resources for semantic artefacts, including examples

In order to establish semantic commonalities we then need services/resources providing common semantic artefacts to be referenced. The types of services needed would be:

- Repositories providing metadata standards/frameworks where the semantic artefacts, and concepts within, are assigned linkable identifiers.
- Repositories providing more formal semantic artefacts like ontologies, terminologies and vocabularies where the semantic artefacts, and concepts within, are uniquely identifiable and linkable.
- Applications that enable easy and effective reuse of semantic artefacts provided by the repositories mentioned above. Since the common use of semantic artefacts will make the analysis of commonalities easier and more effective.
- Mapping services where similarities between semantic artefacts and concepts within these can be managed, found and reused.
- A PID infrastructure which is economically and technically effective to use when assigning identifiers at the concept level and where references to interoperability artefacts can be managed.

#### 4.2.3.1 Common Semantic artefacts

**CESSDA common vocabulary service.** This service has been implemented during 2019 and has not yet been populated with semantic artefacts from the social science domain besides DDI and one CESSDA vocabulary. The repository service is provided by the CESSDA ERIC. Implementation of concept level references by identifiers have not been implemented at this time but is planned.<sup>40</sup>

**SnomedCT.** The healthcare terminology SnomedCT<sup>41</sup> is in use within the health data domain. Concepts have URIs for reference but the common resource, the runtime terminology platform visualised by the SnomedCT browser<sup>42</sup>, have people as the target group not machines. A terminology service for use directly by external stakeholders is not provided by SnomedCT, instead they refer to open source tools and third party repository implementations who uses a Snomed API for reading terminology content like changes etc<sup>43</sup>. Change requests to the terminology can be posted to the Snomed content request service. The input is processed by Snomed and the members' National release centers.<sup>44</sup>

---

<sup>40</sup> See Interoperability taskforce interview with CESSDA CTO

<sup>41</sup> <http://www.snomed.org/>

<sup>42</sup> <https://browser.ihtsdotools.org/?>

<sup>43</sup> <https://confluence.ihtsdotools.org/display/TOOLS/Architecture+Blueprint>

<sup>44</sup> <http://www.snomed.org/snomed-ct/change-or-add>

**ELIXIR.** ELIXIR has domain focussed community activities addressing the delivery and evaluation of community standards and has tooling specifically targeted at interoperability use cases. ELIXIR's Recommended Interoperability Resources<sup>45</sup> resources include vocabulary services. FAIRsharing<sup>46</sup> is the recommended resource on data and metadata standards, inter-related to databases and data policies. ELIXIR has listed recommended Deposition Databases for different datatypes in life sciences<sup>47</sup>. The deposition databases require metadata for submission of data, this is the main driver for researchers to adhere to metadata standards. An important refinement is that data are FAIR and resources are FAIR capable - the two are not equivalent as not all datasets in a FAIR capable resource are necessarily FAIR. As ELIXIR provides core data resources this distinction is important<sup>48</sup>.

#### 4.2.3.2 Conceptual metadata standards & Data type registry models

The different types of semantic artefacts are both used separately and together (when the digital object described is data). Examples of how frameworks/conceptual metadata standards use less formal models with metadata elements that then references Semantic Artefacts and concepts within these are given below. Enhanced machine readability of conceptual metadata standards and data type registry models would be beneficial in order to meet the FAIR requirements. This also includes machine readable references to metadata elements within these.

**ISO11179**<sup>49</sup>. This standard "...addresses the semantics of data (both terminological and computational), the representation of data, and the registration of the descriptions of that data". The standard uses a basic model with two parts (the data element concept and the data element). The data element concept provides meaning and can be represented by the data element. Both the data element concept and the data element describe a recursive, object class with properties. The data element adds representation.

**Generic Statistical Information Model (GSIM)**<sup>50</sup>. It "...provides a set of standardized, consistently described information objects, which are the inputs and outputs in the design and production of statistics. Each information object is defined and its attributes and relationships are specified."

---

<sup>45</sup> <https://elixir-europe.org/platforms/interoperability/rirs>

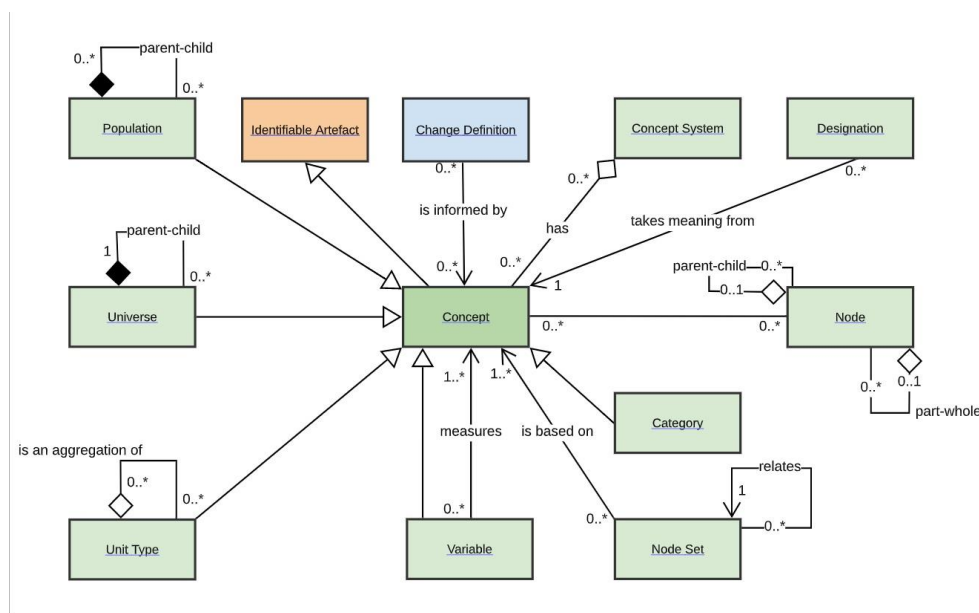
<sup>46</sup> <https://fairsharing.org>

<sup>47</sup> <https://elixir-europe.org/platforms/data/elixir-deposition-databases>

<sup>48</sup> <https://elixir-europe.org/events/webinar-fairness-CDR>

<sup>49</sup> ISO/IEC JTC1 SC32 WG2, ISO/IEC 11179

<sup>50</sup> Generic Statistical Information model (GSIM): Specification, United Nations Economic Commission for Europe (UNECE)



Information objects like populations, variables, nodesets etc all inherit their meaning from one information object, the Concept, that can exist in different semantic artefacts/concept systems providing meaning in different contexts.

The GSIM framework/conceptual metadata standard thus shares several traits with ISO11179 when it comes to the separation of meaning from representation. Data element instances or instances of information objects modeled in above standards can both reference more formal terminological concepts within semantic artefacts providing meaning.

**RDA Data Type Registries (DTR) Output.** The RDA DTR output group<sup>51</sup> was endorsed by RDA and has been accepted as an ICT technical specification. The group's work has been paused while ISO is considering DTR as a standard<sup>52</sup> but an updated draft is available from 2016-01-20 where the proposed model is presented in a set of UML diagrams. The model is based on existing standards "ISO19110, ISO19115, ISO11179, OGC10-090r3 (NetCDF common data model) and the RDA data Type registry prototype (WG output, March 2015)."<sup>53</sup> Since this model includes input from ISO11179 the core is similar and also maps well to GSIM that also has its origins in ISO11179.

**Galaxy**<sup>54</sup> is an open, web-based platform for accessible, reproducible, and transparent computational research, that allows to string tools together into workflows. This requires

<sup>51</sup> <http://doi.org/10.15497/A5BCD108-ECC4-41BE-91A7-20112FF77458>

<sup>52</sup> <https://www.rd-alliance.org/groups/data-type-registries-wg.html>

<sup>53</sup> <https://github.com/usgin/usginspecs/raw/gh-pages/DataTypeModelDraft.pdf>

<sup>54</sup> <https://galaxyproject.org>



the interoperability of data outputs and inputs between tools. Therefore Galaxy uses defined data types<sup>55</sup>, which are annotated using EDAM ontology<sup>56</sup>, which contains bioinformatics operations, types of data, data identifiers, data formats, and topics. The same ontology is e.g. used in bio.tools<sup>57</sup>, the ELIXIR tools and services registry, thus building an interconnected network of interoperable services.

---

<sup>55</sup> <https://galaxyproject.org/learn/datatypes/>

<sup>56</sup> <http://edamontology.org/page>

<sup>57</sup> <https://bio.tools>

## Annex I. Interviews with stakeholders

During the process of creating this document we performed a set of interviews to different stakeholders (researchers from different disciplines) in order to gain a better understanding of their views related to interoperability. As a result, many of the examples used throughout this document are based on examples provided by the interviewees.

The interview process was done during Q4 2019. The following disciplines were covered: Astrophysics, Vulcanology, Marine Sciences, Social Science, Language Resources and Technologies, and Biobanks.

The interview template was as follows:

According to the definition provided in the FAIR data principles (<https://doi.org/10.1038/sdata.2016.18>), Interoperability is focused on making sure that the data can be integrated with other data, and can be used with applications or workflows for analysis, storage, and processing. Furthermore, the following principles are identified (for data and its corresponding metadata):

- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (Meta)data use vocabularies that follow FAIR principles
- I3. (Meta)data include qualified references to other (meta)data

Q0.1: Do you agree with this definition and principles?

Q0.2: What would you add, if any, based on your understanding of interoperability in your research area?

Q0.3: Is there any other type of resource that should be considered in the context of addressing interoperability (e.g., software, methods, protocols)?

There are classifications of interoperability that focus on different levels: technical, semantic, legal, organisational.

Q1.1: Do you understand and agree with these levels?

Q1.2: Do they happen in your research area?

Q1.3: Would you add any other level, or propose changes to this classification?

### **Interoperability**

Q2.1 Do you or your organisation encounter issues using/integrating data/services from different sources? If so, describe the issues and how you tackle these interoperability issues?

Q2.2 Are there any best practices that you would recommend checking?

Q2.3 Do you have training or use external consulting services regarding any aspects of interoperability?

### **Technical interoperability**

Q3.1 Is technical interoperability relevant for your research area/project/services?

Q3.2 If relevant, how do you address technical interoperability in your research area/project/service? Can you provide some examples?

Q3.3 Are the principles/techniques applied in your area/project applicable to other areas or projects/services? Which principles/techniques? To which types of areas/projects/services?

Q3.4 What are the next steps in technical interoperability that should be addressed in the short and medium-term in your area?

### **Semantic interoperability in metadata**

Q4.1 What metadata standards are recommended in your community? (from [FAIRsFAIR survey](#))

Q4.2 Are metadata standards published in a FAIR manner? Which ones?

Q4.3 Do metadata standards reuse other existing metadata standards (generic, such as Dublin Core, or domain specific)?

Q4.4 How do the metadata standards used utilise or relate to semantic resources/concept systems such as ontologies, terminologies, vocabularies?

Q4.5 Are researchers adding such metadata normally? Are they helped by librarians?

Q4.6 Do they normally fill in all the metadata items or only a subset of them (e.g., Dublin-Core like)?

Q4.7 In your experience, are the metadata standards available well suited for your community? If not, please elaborate (from [FAIRsFAIR survey](#))

Q4.8 Do any of your metadata standards have the potential to be reused/used by another community?

### **Semantic interoperability in data**

Q5.1 Are there any good practices in your community on how to best publish data in a usable/reusable manner?

Q5.2 Is data published using any standards (e.g., W3C standards such as RDF, or as Linked Data)?

Q5.3 Do you use semantic resources (ontologies/thesauri/terminologies/vocabularies) in your community to achieve semantic interoperability? If yes, which ones?

Q5.4 Are semantic resources published in a FAIR manner? Which ones?

Q5.5 Do such semantic resources reuse other existing resources (generic or domain-specific)?

Q5.6 Do most researchers know how to use (and effectively use) such semantic resources?

Q5.7 In your experience, are the semantic resources available well suited for your community? If not, please elaborate

Q5.8 Do any of your semantic resources have the potential to be reused/used by another community?

### **Legal interoperability**

Q6.1 Are there any legal obstacles/barriers for the exchange of data in your community (e.g., data protection, copyright issues, etc.)?

Q6.2 Do researchers understand well those barriers and the actions needed to overcome/deal with them?

Q6.3 Is there any agent/mediator that provides legal support in a centralised or distributed manner, or is it done locally at each project/organisation?

**Organisational interoperability**

Q7.1 Do you have any policies or procedures defined in advance to encourage your community to work together and exchange information?

Q7.2 Do you have to obey any cooperation agreements with respect to interoperability?

Q7.3 Do you participate in training sessions or use external consulting services regarding interoperability?