Developing and implementing the semantic interoperability recommendations of the EOSC Interoperability Framework

Deliverable of EOSC-A TF Semantic Interoperability

Authorship Community:

Wolmar Nyberg Åkerström[1], Uppsala University (0000-0002-3890-6620),
Kurt Baumann[2], Switch (0000-0003-0627-8110),
Oscar Corcho[1], UPM (0000-0002-9260-0753),
Romain David, ERINHA (0000-0003-4073-7456),
Yann Le Franc[2], e-Science Data Factory (0000-0003-4631-418X),
Bénédicte Madon, Universidad de Sevilla (0000-0001-8608-3895),
Barbara Magagna[2], GO FAIR Foundation (0000-0003-2195-3997),
Andras Micsik[2], SZTAKI (0000-0001-9859-9186),
Marco Molinaro[2], INAF (0000-0003-3055-6002),
Milan Ojsteršek[2], University of Maribor (0000-0003-1743-8300),
Silvio Peroni[2], University of Bologna (0000-0003-0530-4305),
Andrea Scharnhorst[2], DANS-KNAW (0000-0001-8879-8798),
Lars Vogt[2], TIB (0000-0002-8280-0487),
Heinrich Widmann[2], DKRZ (0000-0001-9871-2687)

1. Co-Chair, EOSC Task Force on Semantic Interoperability
2. Member, EOSC Task Force on Semantic Interoperability

This work is based on the collective efforts of the EOSC Task Force on Semantic Interoperability and all members of the task force and its collaborators during the current mandate (2021–2023) are acknowledged as contributors in the repository where this deliverable is made available.

All authors have reviewed the manuscript and approved the submission.

Contents

Executive Summary

This **draft report** was submitted by the EOSC Semantic Interoperability Task Force to the EOSC Association's Quality Review Committee's (QRC) assessment on 18 January 2024.

This document expands on and provides nuance to some of the concepts defined in the EOSC Interoperability Framework (EOSC-IF) and its reference architecture. It accounts for a deep-dive into the landscape of semantic interoperability implementations and a wide range of interoperability scenarios focused around the *Semantic Interoperability Specification*, some subtypes of Semantic Business Objects, as well as the *Semantic Artefact Catalogue* and *Mapping Repository*. A small set of new concepts of relevance to this work and to EOSC at large have also been added.

The *introduction* provides context to the creation of this report, the *basic concepts* section provides and overview of the related components of the EOSC-IF, and the following four sections summarise explorations that frame the concluding set of recommendation to the EOSC community at large.

The explorations that frame the recomemndations are titled as follows:

- The Semantic Interoperability Specification: Implementation profiles for communities
- The Semantic Artefact Catalogue: Twelve maturity dimensions
- The Mapping Repository: Mappings, crosswalks and common (meta)data elements
- Implementation examples: Common use cases and real-world case studies

The recommendations themselves are organised under the following five broad categories:

1. Align emerging adaptations and implementations to the semantic view of the EOSC-IF reference architecture.
2. Identify and consolidate different approaches to representing and exchanging (meta)data with the FDO model described in the EOSC-IF.
3. Extend the EOSC-IF to include a research process perspective that can support convergence on solutions for common use cases.
4. Extend the set of semantic objects to include artefacts such as mappings and crosswalks.
5. Recognise the semantic artefact catalogue as a critical part of the long-term viability of any research data infrastructure.

EOSC Association AISBL
Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

Introduction

Semantic interoperability is a crucial aspect of the pluriform concept of interoperability, and therefore an integral part of a wider EOSC Interoperability Framework (Baumann et al., 2021). In short, semantic interoperability 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" (European Commission (DG RTD) et al., 2021). And in the context of information exchanged across distributed networks, it emphasises the importance of machine-executable operations.

Part of the European ambitions for EOSC is to create an European Research Data Commons[1] where data are findable, accessible, interoperable and reusable (FAIR) and at the same time as open as possible and as closed as necessary (EOSC Association, 2023). Semantic artefacts are broadly spoken machine-readable and interpretable Knowledge Organisation Systems and should play a central role in regulating the variety of metadata which describe a FAIR Digital Object (David et al., 2023). This includes, in particular, the need for metadata schemes to encompass bibliographic documentation and description of data including machine-readable keywords (Findability), regulate the access to data by data providers (Accessibility), support possible (automated) connections between data from various domains (Interoperability) and contain information on licences and provenance of the data (Re-Usability).

The types and kinds of data, and the meaning they have, highly depend on the context of the research area and the use case (Borgman, 2015). Hence, semantic interoperability instantaneously negotiates between rigour of knowledge representations created in certain domains and the generic elements in those representations which can be shared across domains. It does so in a landscape which is constantly changing. In this landscape, not only the domain specific representations change as knowledge progresses, also the technological possibilities to document, communicate about and collaboratively work on those representations change.

Beyond the negotiation between generic and specific, there is a second dimension in the discussion of semantic interoperability, and this concerns the target audience for making data FAIR. At first glance this might sound odd, as ultimately FAIR implementation is obliged to serve research communities. But, next to research communities there are also those communities which provide research infrastructures and although intertwined both groups are not identical. So, while the ultimate goal of FAIR is *FAIR for human actions*, the FAIR movement addressed the question of what is needed for *FAIR for machines* from the very beginning.

This deliverable clearly focuses on FAIR for machines. But, in doing so, we keep in mind that the purpose of such a machine-centric approach is to enable better FAIR for humans. In other words, we gauge machine actions towards their contribution to enhance semantics for human consumption. We do so by complementing an overview of current technological innovations with a methodological, survey-based approach to evaluate the implementation of FAIR principles in the daily work of communities (research and support).

[1] EOSC Association's website, https://www.eosc.eu

EOSC Association AISBL
Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

Authorship and approach

From these challenges, it is logical to seek for good practices and shareable methods to ensure semantic interoperability rather than to try to establish a once-and-for-all schema or way of working. The quest for 'good practices' in a changing landscape is best supported by bringing experts together who represent with their experiences, a certain (group of) domain(s) and who are involved in on-going innovation. So, the quality of this deliverable relies to a large extent on the composition of the group. Being aware of this from the beginning, the task force decided to operate with an open membership (open to all who were experts and willing to devote time). The scientific background of the about-50 expert members covers all scientific fields[2] with a clear dominance of Computer Sciences, Information Science and Information Technology skills, not unsurprisingly for experts working in research management and in particular in research infrastructures. The group of experts is linked to a variety of past and on-going European funded projects. The connection to target domains for implementing new principles of semantic interoperability is represented by active roles of the experts in European Research Infrastructures covering again all Thematic Areas described in the ESFRI Roadmap[3], such as eLTER, DARIAH, ELIXIR.

The TF's coordination supported all processes of self-organisation which led to the emergence of various sub-groups, and encouraged broad dissemination of results both in science-policy, infrastructural and scientific settings (as documented in the appendix).

[2] Using the Frascati classification for scientific fields (Natural sciences and mathematics, Engineering and technology, Medical and Health sciences, Agricultural sciences, Social sciences and Humanities), https://www.oecd.org/innovation/frascati-manual-2015-9789264239012-en.htm

[3] Data, computing and digital research infrastructures; energy; health & food; physical sciences & engineering; environment; social & cultural innovation, https://roadmap2021.esfri.eu/landscape-analysis/section-1/

EOSC Association AISBL
Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

Basic concepts:
Building on the EOSC Interoperability Framework

This section defines concepts that are used throughout this document and are foundational to the understanding of the recommendations and activities of the task force. It provides an overview of the relevant terms and the definitions of concepts related to semantic interoperability and the approaches used to inform the task force activities.

As this document expands on the semantic interoperability recommendations of the EOSC Interoperability Framework[4] (EOSC-IF), it is natural to use this framework as a starting point in the further discussion. The EOSC-IF itself mirrors some of the concepts proposed by the European Interoperability Framework (EIF) for public administrations in Europe and the corresponding European Interoperability Reference Architecture[5] (EIRA). Other frames of reference in the context of this document are the glossaries of the Strategic Research & Innovation Agenda (SRIA)[6] and the EOSC Partnership's Monitoring and Evaluation Framework[7] (MF) of the EOSC Association.

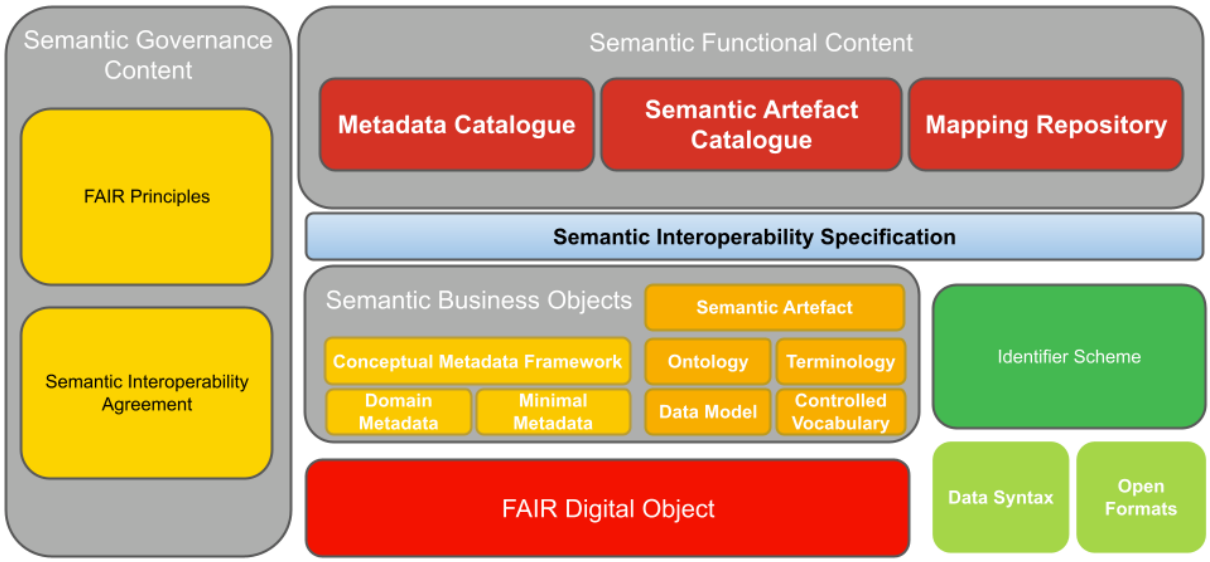

Figure 1: EOSC-IF Semantic view

The EOSC-IF extends the EIRA view with a special focus on FAIR Digital Objects (Figure 1). While a FAIR Digital Object encapsulates metadata and other semantic descriptions in the object itself, these definitions rely on so-called Semantic Business Objects that provide the necessary semantic foundations and meanings.

This document expands on and provides nuance to some of the concepts defined in the EOSC-IF and its reference architecture. It accounts for a deep-dive into the landscape of

[4] European Commission (DG RTD) et al., 2021, https://doi.org/10.2777/620649
[5] European Commission (DG DIGIT), 2017, https://joinup.ec.europa.eu/asset/eia/description
[6] EOSC Association, 2023, https://eosc.eu/sria-mar
[7] EOSC Association, 2022, https://eosc.eu/monitoring-reporting/

EOSC Association AISBL
Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

semantic interoperability implementations and a wide range of interoperability scenarios focused around the *Semantic Interoperability Specification,* some subtypes of *Semantic Business Objects*, as well as the *Semantic Artefact Catalogue* and *Mapping Repository*. A small set of new concepts of relevance to this work and to EOSC at large have also been added as illustrated in the diagram below (Figure 2).

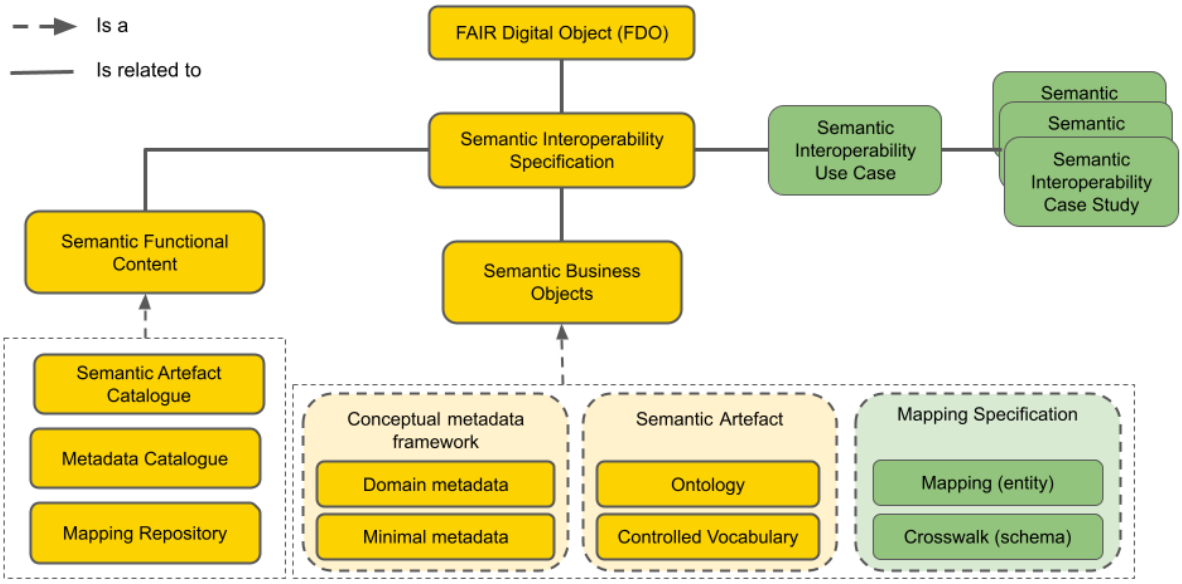

Figure 2: An overview of the subset of EOSC-IF terms reused in this deliverable together with newly introduced terms (in green) and how they are related.

Crosswalk (schema) [8] [9] [10]

Crosswalks establish relationships between elements in different models to achieve interoperability and effective data exchange, particularly when dealing with data from various domains. They involve translation, converting (meta)data from one schema to another, and promoting cross-domain (meta)data discovery. Actionable crosswalks are typically presented in tables that align and map data across different schemas.

FAIR Digital Object [FDO] [11] [12] [13] [14] [15]

A FDO is an information entity composed by a persistent identifier (PID) such as a DOI resolving to a PID Record that gives the object a type along with a mechanism to retrieve its

[8] Riley, 2004
[9] ISO/IEC 11179, Information technology, Metadata registry (MDR)
[10] Schema crosswalk, https://en.wikipedia.org/wiki/Schema_crosswalk
[11] European Commission (DG RTD), 2018
[12] European Commission (DG RTD) et al., 2021
[13] Soiland-Reyes et al., 2023
[14] nfdi4phys, FAIR Digital Object, https://nfdi4phys.de/fdo/
[15] FAIR Digital Object Framework, https://www.go-fair.org/today/fair-digital-framework/

EOSC Association AISBL
Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

bit sequences, metadata and references to possible operations according to the FAIR principles.

Editor's note: The EOSC-IF goes into some length to describe the Digital Object concept (section 1.1.4) and then expand on the role of FDO:s in EOSC (section 4.1). At the same time the activities around the FDO Framework have made progress towards a formal specification. Accurately summarising this body of work is beyond the scope of this document and a recent assessment is accounted for in (Soiland-Reyes et al., 2023).

Mapping (entity)[16]

Mapping is an explicit relationship between concepts, terms or even data elements from two different semantic artefacts, metadata schemas, information systems or databases. The relationship can have different levels of complexity ranging from a simple one-to-one equivalent relation to complex data transformations.

Ontology[17]

In information science, an ontology encompasses a representation, formal naming, and definitions of the categories, properties, and relations between the concepts, data, or entities that pertain to one, many, or all domains of discourse. More simply, an ontology is a way of showing the properties of a subject area and how they are related, by defining a set of terms and relational expressions that represent the entities in that subject area.

Semantic Artefacts[18]

A Semantic Artefact is 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.

Editor's note: The term was coined to avoid the ambiguity associated with the term ontology, as described in the introduction to the D2.2 FAIR Semantics: First recommendations report (Le Franc et al., 2020) of the FAIRsFAIR project.

Semantic Artefacts Catalogue[19]

A Semantic Artefacts Catalogue is a dedicated web-based system that fosters the availability, discoverability, long-term preservation and maintenance of semantic artefacts.

[16] Broeder et al., 2021
[17] Ontology (information science), https://en.wikipedia.org/wiki/Ontology_(information_science)
[18] Le Franc et al., 2020)
[19] Corcho et al., 2023a

EOSC Association AISBL
Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

Semantic Interoperability[20] [21]

Semantic Interoperability is about enabling meaning-preserving exchange of information across machines and humans. For this purpose, the format and meaning of the exchanged data and information must be precisely specified and described (meaning preservation).

Editor's note: In the context of machine-to-machine operations, it's the ability of computer systems to exchange and interpret data with a common understanding of the meaning of that data. In other words, it's about ensuring that data from one system can be correctly understood and used by another, even if those systems were developed independently or used different data formats and structures.

Semantic Interoperability Case Study

Description of a real-life situation where semantic artefacts are used in practice. These can be broad, covering several semantic artefacts and actors to achieve semantic interoperability.

Semantic interoperability Use Case

A usage scenario for the semantic artefacts abstracted from a case study such that it can be generalised to other disciplines, systems or regions. Keeping track of this will support comparisons across case studies. It describes sequences of interactions between actors and semantic artefacts to achieve semantic interoperability.

Semantic Interoperability Specification

Semantic Interoperability Specifications refer to a comprehensive set of rules, standards, and specifications aimed at ensuring accurate data and information exchange among diverse systems, devices, and organisations, both for human and computer systems. These specifications encompass: Data standards, formats, structures and schemas; terminology and semantics; mappings, crosswalks and transformation standards for metadata; validation and compliance mechanism and documentation comprising guidelines to assist developers and users.

[20] Sheth & Larson, 1990
[21] European Commission (DG RTD) et al., 2021

EOSC Association AISBL
Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

The Semantic Interoperability Specification: *Implementation profiles for communities*

The EOSC IF was based, among other sources, on a survey conducted to assess the knowledge and practices related to interoperability from different stakeholders. Taking into account new developments we want to extend this information source with a new survey to document the semantic interoperability specification for each community of practice formalising the survey concept and the possible answers into a knowledge model as formerly already used to collect FAIR Implementation Profiles (FIPs) (Magagna, Schultes, et al., 2022). This is done by extending the FIP Ontology[22] with concepts needed to address specifically semantic interoperability aspects. Figure 3 shows how these terms (in red) fit into the previous overview. Some of the new terms can be interpreted as synonyms for terms in Figure 2. How they relate to each other and definitions for all these concepts are provided below.

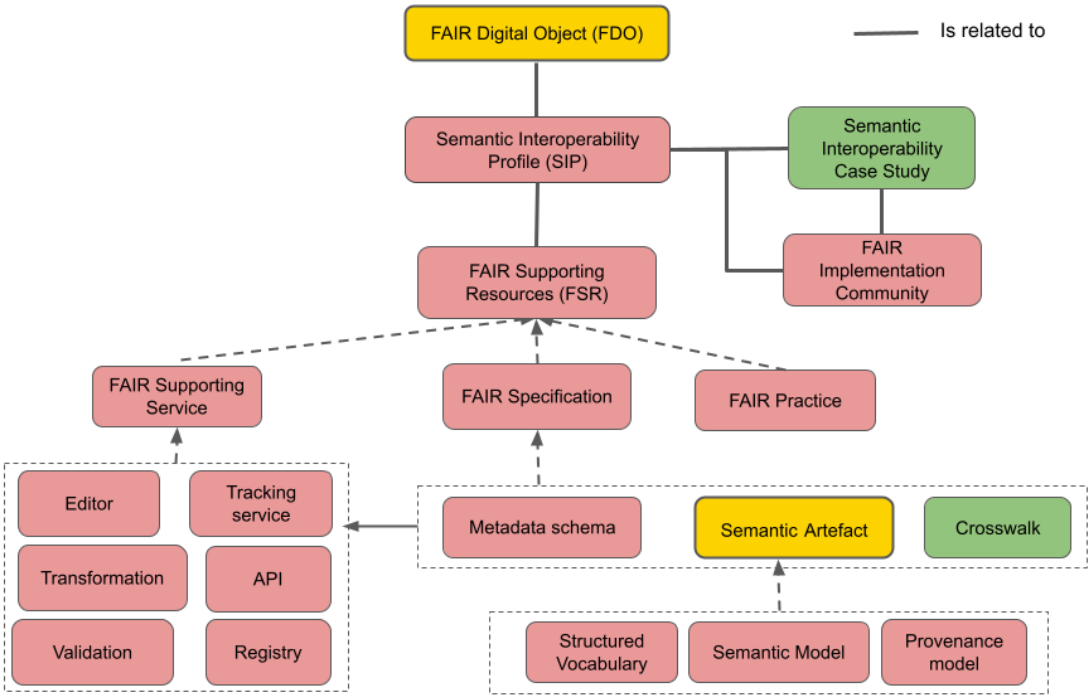

Figure 3: Overview of SIP specific terms (in red), next to reused terms from Fig. 2

Method for collecting Semantic Interoperability Profiles (SIPs)

To collect supporting resources to solve semantic interoperability problems (FAIR Supporting Resources, in short FSRs) we designed a questionnaire for data management experts involved in cluster projects, EOSC projects, or research infrastructures.

[22] FIP Ontology, https://w3id.org/fair/fip/

EOSC Association AISBL
Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

The questionnaire is implemented using an instance[23] of the Data Stewardship Wizard (DSW)[24] to model the questions. The DSW provides a human-readable interface, helping users to understand its aims and functionalities intuitively and interact with it. In addition, it allows for a machine-readable output for a better comparability of the results. The FIP Ontology is used as the semantic foundation of the knowledge model. The ontology basically defines how communities can declare which resources they use to address each of the FAIR Principles.

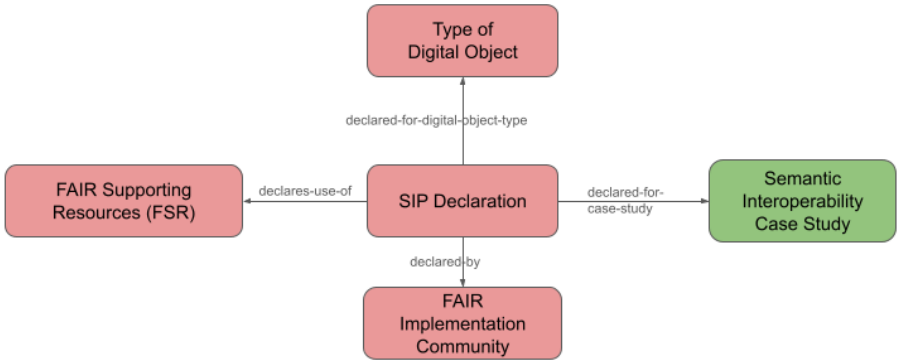

Figure 4: The FIP ontology adapted for the SIP Declaration

The Semantic Interoperability Profile is a slight adaptation of this approach (see Fig. 4) and defines the list of resources used to address a specific semantic interoperability case study as defined by a community. These FSRs are described as nanopublications to provide a globally unique, persistent and resolvable identifier and machine-readable representation based on a specific metadata schema. A nanopublication is the smallest unit of publishable information, with associated provenance, expressed as a knowledge graph that is formal and machine-interpretable[25]. FAIR Connect[26] is used as the search engine for all FSRs and SIPs. The survey is an ongoing effort that will provide results beyond the publication of this deliverable. After a critical number of contributions by communities the SIP outcomes will become a valuable knowledge base for successful implementations. Convergence in the reuse of these implementations should be reconsidered for generic semantic interoperability use cases and provide new recommendations to be endorsed.

The key messages of the SIP method are:

- The extended FIP Ontology is used to collect information on current interoperability resources in the form of a **Semantic Interoperability Profile (SIP)** chosen by a **FAIR Implementation Community**.
- **Semantic Interoperability Case Studies** document real-life situations to solve semantic interoperability for a set of **FAIR Digital Objects**. The FAIR Supporting

[23] Semantic Interoperability Profiles (SIPs), https://sip-wizard.ds-wizard.org/

[24] Data Stewardship Wizard (DSW), https://ds-wizard.org/

[25] Nanopublications, https://nanopub.net/

[26] FAIR Connect, https://fairconnect.pro/

EOSC Association AISBL
Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

Resources represent the artefacts used to implement semantic interoperability and are described in a **Semantic Interoperability Profile (SIP)**.

- Case studies and SIPs are maintained and endorsed by **FAIR Implementation Communities**.
- SIPs list the applied **FAIR Supporting Resources (FSR)**, further grouped into three categories: specifications, services and practices.
- **FAIR Specifications** like crosswalks, semantic artefacts require **FAIR Supporting Services** in order to include terms from controlled vocabularies, transform, edit, validate, exchange or register them. Metadata schemas require editors that allow the inclusion of concepts from semantic artefacts.
- All FSRs are documented via nanopublications retrievable via FAIR Connect, whereas the Semantic Interoperability Case Studies are documents referred to via Zenodo DOIs.

Relations to the terms used in the EOSC-IF:

- The Semantic Interoperability Profile is a Semantic Interoperability Specification for a specific FAIR Implementation Community,
- The Semantic Business Objects and the Semantic Functional Content are addressed in the SIP as FAIR Supporting Resources,
- Catalogues and Repositories are addressed in the SIP as Registries,
- Semantic Artefacts are addressed in the SIP as controlled vocabulary, semantic model and provenance model, as these are defined as specific FAIR Enabling Resources for FAIR Principles I2, I3 and R1.2 in the FIP Ontology,
- Conceptual metadata framework is addressed in the SIP as a metadata schema.

Crosswalk: A specification consisting of a set of rules that define how (meta)data elements or attributes from one schema can be aligned and mapped to (meta)data elements or attributes in another schema that share the same constraints and thus share the same semantic role.
(Source: https://w3id.org/fair/fip/terms/Crosswalk)

FAIR Implementation Community

A FAIR Implementation Community (FIC) is a self-identified collection of people and/or organisations with the aim to implement the FAIR Principles.
(Source: https://w3id.org/fair/fip/terms/FAIR-Implementation-Community)

Editor's note: We reuse this concept here in the context of Semantic Interoperability as a community will also have to find agreements on which services to use for dealing with specifications like metadata schemas and semantic artefacts.

FAIR Practice

An adopted use of a specific protocol, tool, procedure or workflow to support FAIRification within a community.
(Source: https://w3id.org/fair/fip/terms/FAIR-Practice)

EOSC Association AISBL
Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

Editor's note: FAIR practices in the context of FAIR supporting resources involve adopting standardised approaches, documentation and strategies aligned to the FAIR principles. Such practice iteratively facilitates the creation, management and dissemination of resources enhancing overall FAIRness of data (David et al., 2020), and digital assets (everything that is created, stored digitally, and identifiable/discoverable with provided values). Furthermore, it allows better data sharing, supports reusability and has an impact on research.

FAIR Representation Service

A transformation service that converts non-FAIR data into a FAIR representation using machine-readable knowledge representation languages.
(Source: https://w3id.org/fair/fip/terms/FAIR-representation-service)

Editor's note: FAIR representation services can be editors (an editor is a service that provides user-friendly interface for easy editing of metadata, vocabularies and crosswalks), transformation-/validation processes, e.g. a validation service is a system that verifies the accuracy, completeness, or compliance of data, code or processes against predefined criteria or standards, APIs, and registries, e.g. the registry used can be the same one as for the digital objects themselves.

FAIR Specification

A precise description of features, requirements, constraints and recommendations for a specific implementation of a component, system or service supporting the implementation of the FAIR principles.
(Source: https://w3id.org/fair/fip/terms/FAIR-Specification)

Editor's note: Examples of specifications are metadata schema, semantic artefact, and crosswalk.

FAIR Supporting Resources (FSR)

Any resource that supports FAIR Data Stewardship. FSRs are represented as FAIR Digital Objects (using the nanopublication framework) with Globally Unique, Persistent, Resolvable Identifiers (GUPRI) that resolve to machine-readable metadata about the resource. (Source: https://w3id.org/fair/fip/terms/FAIR-Supporting-Resource)

FAIR Supporting Service

Any online accessible software system or component that supports the implementation of the FAIR Principles. (Source: https://w3id.org/fair/fip/terms/FAIR-Supporting-Service)

Editor

A service that provides a user-friendly interface for easy editing of metadata, vocabularies or crosswalks. (Source: https://w3id.org/fair/fip/terms/Editor)

Editor's note: The editor can be understood as an online available tool (service).

EOSC Association AISBL
Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

Knowledge representation language

A language specification that enables knowledge to be processed by machines.
(Source: https://w3id.org/fair/fip/terms/Knowledge-representation-language)

Metadata Schema

A specification that specifies the structured representation of metadata describing attributes of data or other digital objects in terms of semantics, syntax and optionality.
(Source: https://w3id.org/fair/fip/terms/Metadata-schema)

Registry (-> Catalog; -> Repository)

A service that indexes metadata and data and provides search over that index.
(Source: https://w3id.org/fair/fip/terms/Registry)

Editor's notes: Depending on the type of the digital objects and how they are organised there can be:

- Registries for metadata records,
- Registries for semantic artefacts (-> Semantic Artefact Catalogue),
- Registries story for crosswalks and for term mappings,
- Registries for data.

Repositories additionally provide storage and preservation services for digital objects.

Provenance model

A specification that specifies metadata describing the origin and lineage of data or other digital objects.
(Source: https://w3id.org/fair/fip/terms/Provenance-model)

Provenance Tracking Service

A system that systematically captures, stores and manages detailed information about the origin, history, and lifecycle of digital objects creating metadata based on a provenance model.
(Source: https://w3id.org/fair/fip/terms/Provenance-Tracking-Service)

Semantic Model (-> Ontology)

A specification that defines qualified relations between entities describing data or other digital objects according to the Linked Data principles. This can include semantic data models and ontologies.
(Source: https://w3id.org/fair/fip/terms/Semantic-model)

EOSC Association AISBL
Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

Structured Vocabulary (-> Controlled Vocabulary)

A specification for a controlled list of uniquely identified and unambiguous concepts with their definitions represented using web standards.
(Source: https://w3id.org/fair/fip/terms/Structured-vocabulary)

Editor's notes: Special types of vocabularies include thesauri, taxonomies, controlled vocabularies, ontologies.

Semantic Interoperability Profile (SIP)

A Semantic Interoperability Profile (SIP) is a list of FAIR Supporting Resources chosen by a community to support semantic interoperability of (meta)data. The SIP is derived from the definition of FIP, the FAIR Interoperability Profile: (Source: FAIR Implementation Profile (FIP) Ontology, Semantic interoperability Profile,.
(Source: https://w3id.org/fair/fip/latest/Semantic-Interoperability-Profile)

SIP Declaration

The expression of a community on how it addresses a SIP question.
(Source: https://w3id.org/fair/fip/latest/SIP-Declaration)

Web Application Programming Interface (Web-API)

An Application Programming Interface (API) for the World Wide Web that allows different software applications to communicate with each other.
(Source: https://w3id.org/fair/fip/terms/Web-API)

Validation Service

A system that automatically verifies the accuracy, completeness, or compliance of data, code or processes against predefined criteria or standards.
(Source: https://w3id.org/fair/fip/terms/Validation-Service)

EOSC Association AISBL
Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

The Semantic Artefact Catalogue:
Twelve maturity dimensions

Among the ongoing effort of the EOSC Task Force on Semantic Interoperability, there has been a particular focus on the need for identifying dimensions to assess the *maturity* of semantic artefact catalogues. A catalogue of this kind is a critical resource, acting as a "keyholder" and guarantor for semantic interoperability implementations, since semantic artefacts are the "keys" that permit semantic interoperability of systems.

Understanding the maturity of these catalogues is a crucial aspect to consider when envisioning how to enable and improve the long-term preservation of semantic artefacts. Indeed, a maturity model for assessing such catalogues would provide recommendations for governance. It would be based on defined workflows for preserving and maintaining semantic artefacts and should help assess/address interoperability challenges towards the vision for a Europe-wide shared data infrastructure based on a FAIR ecosystem of data and services.

The goal of this section is, thus, to propose dimensions and features that can be used to assess the maturity of semantic artefact catalogues. By gathering various definitions concerning catalogues that store and serve semantic artefacts (either at the metadata or data level or both), and then analysing the current literature on the topic, we have defined a maturity model to measure, compare and evaluate available semantic artefact catalogues. This maturity model (Corcho et al., 2023b) is composed of several *dimensions* in which catalogues could be compliant and/or improved.

The application of the dimensions was tested by analysing a collection of 26 semantic artefact catalogues (Busse et al., 2023), aiming, on the one hand, at completing the maturity model by adding additional features (or sub-criteria) for each of the dimensions identified and, on the other hand, at showing how existing catalogues comply with such dimensions and sub-criteria. These features serve as a specification of different levels of compliance of a catalogue against the related dimension and provide a categorical view to assess the maturity of semantic artefact catalogues.

The target audience of our work and the maturity model presented here refer to at least the following users:

- semantic artefact providers and users, to let them know about potential catalogues to share and find resources;
- developers of semantic artefact catalogues, to let them know about criteria relevant for users and that can be used to enhance their catalogues.

The twelve maturity dimensions identified are summarised as follows – and are described with more details, with their related features (or sub-criteria) by Corcho et al. (2023a).

Metadata: The identification of the minimal set of metadata to describe the catalogue and its semantic artefacts. A huge importance is also given to using metadata standards and schemas (e.g., MOD[27], DCAT or Schema.org), adopting machine-readable formats, the documentation associated, and the licences used to release the metadata.

[27] https://github.com/FAIR-IMPACT/MOD

EOSC Association AISBL
Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

Openness: The concept of being open from different perspectives. On the one hand, it concerns technical openness, referring to the metadata handled in the catalogue, the software used to run the catalogue, and the services and protocols used to access the metadata. On the other hand, openness also refers to the social attitude of enabling anyone interested in depositing and helping govern the catalogue.

Quality: The possibility of having mechanisms to check part of the quality of the metadata provided (like O'FAIRe[28] and FOOPs[29]) and, thus, the catalogue itself. In particular, if processes and workflow are in place for peer reviewing new entities and curating the catalogue.

Availability: It refers to the availability of the metadata and whether there are methods in place for guaranteeing privacy and access only to certain data due to legal or other contextual issues.

Statistics: The availability of statistics referred to the catalogue (number of semantic artefacts handled, number of users, etc.) in time to measure the usage of the catalogue and its growth.

PID: The use of persistent identifiers (PIDs) that refer to the metadata of the various semantic artefacts described in the catalogue and the semantic artefacts themselves.

Governance: The rules that define the governance of the catalogue and its goals and purpose which should allow community input and responsibility for the integrity of the metadata.

Community: The mechanism in place to involve the community in the catalogue, identifying and reaching target users' expectations and attracting stakeholders from diverse lived experiences and viewpoints.

Sustainability: The models in place to sustain services financially and preserve the catalogue in the long run.

Technology: The tools that the catalogue should provide to enable users to have a better experience in exploring the data, such as REST APIs, Web search interfaces, SPARQL endpoints, etc.

Transparency: The processes behind the catalogue, from the elections of new members of the various governing boards, curators, etc., to the clarity in exposing fees for the services offered by the catalogue and its revenue model.

Assessment: The presence of some practice in place for assessing the catalogue against all these dimensions, e.g. by adopting self-assessment exercises and/or by asking third parties to run an independent assessment of the catalogue.

Each of these dimensions could be assessed in different ways and have different relative importance depending on the context and the type of object designated (as appropriate for the community concerned). The maturity model is a work in process and would be used in combination with other recommendations, such as the document produced by the ESOC Task Force on FAIR Metrics and Data Quality[30] (Lacagnina et al., 2023; Wilkinson et al., 2022) and the forthcoming outputs of the EOSC Task Force on PID Policy and Implementation[31].

[28] https://github.com/agroportal/fairness
[29] https://foops.linkeddata.es/FAIR_validator.html
[30] https://eosc.eu/advisory-groups/fair-metrics-and-data-quality
[31] https://www.eosc.eu/advisory-groups/pid-policy-implementation

EOSC Association AISBL
Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

The Mapping Repository:
Mappings, crosswalks and common (meta)data elements

Achieving semantic interoperability among the diversity of data resources available to build EOSC requires the possibility to share a common semantic description of the metadata and the data. However, each resource uses its own metadata schema, its own semantic artefact to convey meaning and its own data model. To ensure a semblance of common semantic description, similar or closely related concepts, terms, metadata fields, data elements from the different sources should be linked together. These links, here called mappings, can be used to improve the quality of search results across heterogeneous and distributed resources or to integrate different data resources together using a common data model.

These mappings can be done at the level of metadata e.g between individual classes of ontologies (also called semantic alignment, semantic matching), concepts from controlled vocabularies, metadata fields from two metadata schemas or at the level of the data. For metadata, relations should describe the semantic or/and logical relation between two concepts such as the concept "high_type_cloud_area_fraction" within the Climate and Forecast Standard Names vocabulary, hosted on the NERC Vocabulary service[32] is sameAs the "high_type_cloud_area_fraction" defined in the Marine Metadata Interoperability platform[33]; or the concept "limb" (UBERON:0002101) defined in the UBERON ontology is the sameAs with the "free limb" (FMA:24875) concept, defined in the Foundational Model of Anatomy ontology[34]. For the data level, this relation should make explicit the transformation to convert the value to another value e.g. temperature in °C to temperature in Kelvin or height in inches to height in cm.

The creation of sets of mappings or crosswalks provide the transformation of a source metadata schema into a target metadata schema (e.g. DataCite to DCAT, …). Mappings are used in various contexts and are often represented as a correspondence table between the elements. This representation does not provide any explicit information regarding the exact relation and does not offer the possibility to associate additional information such as provenance of the mappings. A large number of existing mappings are presented as tables e.g. Codemeta mappings[35], Datacite to Dublin Core mappings[36], DCAT to schema.org mappings[37]. Creating these mappings is time and resource consuming. Therefore they should be shared with others.

To compensate for a huge gap in standardisation of the mapping representation, the biomedical community proposed a "Simple Standard for Sharing Ontology Mapping" (SSSOM). This model builds on the usual common practices of presenting the mapping in a table format but extends the model with contextual and provenance metadata. SSSOM mappings can be

[32] NERC Vocabulary service, https://vocab.nerc.ac.uk/collection/

[33] Marine Metadata Interoperability ORR, https://mmisw.org/ont#/

[34] Examples from SSSOM specification, https://mapping-commons.github.io/sssom/spec/

[35] The CodeMeta Project Crosswalks, https://codemeta.github.io/crosswalk/

[36] https://schema.datacite.org/meta/kernel-4.4/doc/DataCite_DublinCore_Mapping.pdf

[37] https://ec-jrc.github.io/dcat-ap-to-schema-org/

EOSC Association AISBL
Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

serialised as RDF. The SSSOM model allows capturing simple mappings i.e. mappings between two entities.

In some cases, these mappings can become more complex and may require linking one entity to two entities in the target metadata schema such as "full name" in the metadata schema 1 and "first name" & "last name" in the metadata schema 2. The two individual mappings linking the element of metadata schema 1 and each element of the metadata schema 2 could be encoded using SSSOM. However, in order to transform the data from one to the other, a notion of order is necessary to ensure that full name is the aggregate of first name and last name which is not supported by SSSOM. Therefore, it is necessary to ensure either the extension of SSSOM or the creation of another model to describe the different complex mappings.

The importance of the mappings has been considered in the EOSC-IF with the inclusion of a mapping repository as a Semantic Functional Components. Surprisingly, the mappings and crosswalks are not considered as part of the Semantic Business Object. This need for crosswalk has been also emphasised in the SRIA and the MAR and two Horizon Europe projects are tackling these issues: FAIR Impact which aims at providing both recommendations on how to make mappings and crosswalks FAIR and FAIRCORE4EOSC which develops a mapping repository as described in the EOSC-IF.

By creating mappings and crosswalks between the more widely used metadata schemas, it becomes possible to identify a set of concepts/terms that will be common to all these metadata schemas. Such work was initiated by the EOSC-IF team and a table of crosswalks between metadata schemas is presented by Ojsteršek (2021).

EOSC Association AISBL
Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

Implementation examples:
Common use cases and real-world case studies

In this section, we showcase how semantic interoperability can be understood and achieved within and between research infrastructures through common and domain specific scenarios. Case studies and use cases are different forms of scenarios, each describing how stakeholders and other actors/systems interact with an information system. And in the context of the EOSC and the EOSC-IF, a collection of these types of scenarios can help demonstrate the value and potential reuse of solutions across initiatives and engage stakeholders across diverse communities in requirements gathering.

Example: Discovery and access across heterogeneous resources

A common high-level goal that often appears as an example in discussions around EOSC is the ability to discover and access data from across domains to increase the amount of available data or to facilitate interdisciplinary studies. And discoverability is still one of the key challenges for open science: in many ways, we cannot cash the cheques written by this movement if we do not increase the visibility of research outputs. Many research data discovery services have thus emerged, and a review of the status quo and open challenges within "The Open Ecosystem of e-Infrastructures for Data Discovery" is discussed in Bardi et al. (2022).

A common approach relies on metadata harvesting and homogenisation, i.e., harvesting would gather and associate metadata with their corresponding semantic artefacts and harmonisation could rely on mappings and crosswalks to consolidate them into a single metadata catalogue (semantic space, database, index, etc) conforming to a coherent metadata framework and limited set of semantic artefacts. This approach is exemplified in the centralised architecture of the Discovery Service implemented in the current EOSC Portal Marketplace, and by the vast majority of data portals across Europe.

The same approach can be implemented in different ways and in the case of the EOSC Portal, metadata harvesting follows the EOSC Rules of Participation for Data Sources and the homogenisation is delegated to each Data Source ensure that they can provide metadata compliant with the OpenAIRE Guidelines.

Alternative approaches can also be considered to address the same high-level goal, and in this context an example could include a peer-to-peer model where discovery and access relies on federated search capabilities across external metadata catalogues without explicit harvesting as exemplified for the EOSC Platform in Keith & Broeder (2024).

Common use cases and case studies for interoperability in EOSC

In this context, a common use case is a scenario that describes an approach (such as the ones exemplified above) using terms and abstractions that would allow them to generalise well across implementations. Some concepts and useful abstractions are provided by the EOSC-IF reference architecture, while others can be referenced from widely recognised maturity models, standards and recommendations associated with specific components or the

EOSC Association AISBL
Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

relevant thematic contexts. Candidates for scenarios to generalise can be identified in a multitude of sources describing specific implementations or solutions adopted in thematic contexts or as part of design documents.

A case study that describes interoperability challenges and solutions can provide the context necessary to support recognition of gaps to be addressed; demonstrate the value of adoption; or to inform future developments. As a complement to the common use cases in this context, the case study is an experience report that describes real-world scenarios and can be used to bridge to the specific problem descriptions, implementations and semantic artefacts used in a community.

To support the synthesis of this report, members and contributors to the EOSC-A Semantic Interoperability Task Force contributed a selection of case studies (some of which are summarised in Annex I). A template was also proposed to support the identification of candidates for common use cases and how to align them with the components of the reference architecture of the EOSC-IF (Nyberg Åkerström et al., 2022; Nyberg Åkerström & Maccallum, 2023).

Case studies covered in the Annex I:

- European coordination to contribute to international collaborations (Euro Virtual observatory - CNRS/CDS, ARI/U. Heidelberg, INAF.)
- Coordination across projects to agree on common standards ('CMIP6 governance' - IS-ENES/DKRZ)
- Machine-actionable research data/tools management - a Research Ecosystem Approach
- Semantic interoperability in the Humanities
- Provenance information and variables customised for specific user needs ('Semantic mapping of climate variables')
- Access to data from a National Statistical Agency - the ODISSEI portal
- Consistency in the face of changing technologies ('Semantic mapping of plant phenotyping variables")
- Cross-disciplinary models of research information and their representations ('Semantic mapping of Highly Pathogenic Agents variables")
- Virtual graphs for harmonised ontology-based data access (Semantic data mapping to RESCS.org (Ontology) and data validation with SHACL-Shapes)

EOSC Association AISBL
Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

Recommendations

The following recommendations aim to improve on and support future implementations of the EOSC-IF. They are directed towards the EOSC initiatives that are governing, conceptualising, building and operating various components of the envisioned web of FAIR data and services and focus on aspects that would maximise the value of a common reference architecture for interoperability solutions. The suggested actions should support convergence on common solutions and identify gaps and directions for improved implementations.

The recommendations are organised under five broad categories reflected in the body of this document:

1. Align emerging adaptations and implementations to the semantic view of the EOSC-IF reference architecture.
2. Identify and consolidate different approaches to representing and exchanging (meta)data with the FDO model described in the EOSC-IF.
3. Extend the EOSC-IF to include a research process perspective that can support convergence on solutions for common use cases.
4. Extend the set of semantic objects to include artefacts such as mappings and crosswalks.
5. Recognise the semantic artefact catalogue as a critical part of the long-term viability of any research data infrastructure.

1. Align emerging adaptations and implementations of the EOSC-IF reference architecture

Objective: Establish the semantic view of the EOSC-IF reference architecture as a shared frame of reference to exchange and converge on shared practices and solutions for semantic interoperability.

Indicators of success:

- Increased awareness of the range of available solutions and good practices associated with the different components of the EOSC-IF across the wider EOSC community.
- Shared framework for alignment with other data spaces and the internet at large.
- Support for cross-domain applications such as discovery and data integration.
- Shared vocabulary to support discussions across projects, professions, and domains.

Suggested actions:

- Track and consolidate all work done to expand on and implement the EOSC-IF across EOSC initiatives into a single collection of documents / resources.
- Support continued efforts to bring together experts from diverse communities to create awareness of these efforts and opportunities to converge on shared practices and solutions around the EOSC-IF.

EOSC Association AISBL
Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- The federation of EOSC Nodes should adopt solutions that integrate well with other data spaces and the internet at large and align their architectural descriptions with the EOSC-IF.
- Reference and elaborate on the semantic view of the EOSC-IF to describe and assess mediation across the emerging EOSC Nodes and the wider EOSC stakeholder community.
- Coordinate an exerted effort to expand the number of terms related to semantic interoperability defined in the EOSC-IF to support alignment across adaptations and implementations.

2. Identify and consolidate different approaches to representing and exchanging (meta)data with the FDO model described in the EOSC-IF

Objective: Promote successful interpretations and implementations of the FAIR Digital Objects (FDO) model described in the EOSC-IF to support adoption across the wide range of data types, distributions and models of existing digital repositories.

Indicators of success:

- Increased availability of interoperability and community guidelines that specify how they define the boundaries around (meta)data and the relations to semantic artefacts, metadata schema and other semantic business objects / semantic functional content.
- (Meta)data is increasingly served with resolvable references to the complete set of semantic artefacts necessary to decode and make sense of their contents.
- (Meta)data increasingly makes use of qualified references to concepts defined in semantic artefacts.

Suggested actions:

- Converge on a common way to describe and share semantic interoperability specifications that define the relevant aspects of the FDO model adopted by data providers and services across EOSC.
- Develop and implement processes to enrich and describe existing (meta)data using concepts defined in openly available and FAIR semantic artefacts.
- Improve and promote the use of editor services to better support creators and curators of (meta)data using concepts defined in semantic artefacts.

3. Extend the EOSC-IF to include a research process perspective that can support convergence on solutions for common use cases

Objective: Promote common use cases and context specific case studies that effectively demonstrate the value of the EOSC-IF components and their implementations to serve as input to EOSC initiatives and as examples including lessons learned to a wide range of EOSC stakeholders.

EOSC Association AISBL
Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

Indicators of success:

- Increased availability of compelling demonstrators that illustrate how stakeholders, systems and other actors interact with semantic artefact catalogues, mapping repositories and other functional content.
- Stakeholder validation of EOSC's value proposition through common use cases and exemplified by case studies that use examples and language that they can relate to.
- New opportunities for collaboration and engagement across and beyond EOSC initiatives, focusing on tools, services and specifications to support common use cases.

Suggested actions:

- Complement the EOSC-IF with a framework for and references to curated resources with interoperability case studies and common use cases.
- Provide incentives and offer support to stakeholder groups within EOSC initiatives and the wider EOSC user community to share case studies and validate common use cases and their implementations (explore options such as the SIP survey, FAIR Cookbook).
- Liaise with communities beyond EOSC that are working on solutions to support similar use cases, including the research communities related to semantic technologies, standardisation bodies, and public sector initiatives such as Interoperable Europe / SEMIC.
- Identity and index recommendations from EOSC initiatives and the wider EOSC stakeholder community that can be associated with the common use cases and EOSC-IF components, such as the deliverables of the FAIRsFAIR project that describe several processes related to semantic artefacts and interoperability, and the FAIR Cookbook etc.

4. Extend the set of semantic objects described in the EOSC-IF to include artefacts such as mappings and crosswalks

Objective: Recognise the importance of the mapping repository component and the value of sharing and reusing the related semantic objects to support mediation across semantic artefacts.

Indicators of success:

- Increased engagement across EOSC initiatives in finding common solutions and good practices for sharing and making mappings and crosswalks FAIR.
- Increased availability of compelling demonstrators, such as common use cases and case studies, involving implementations of the mapping repository component, mappings and crosswalks.
- Emerging shared vocabulary to support discussions around the mapping repository component across projects, professions, and domains.
- Convergence towards descriptions of requirements and desirable characteristics for the mapping repository component

EOSC Association AISBL
Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- Convergence towards descriptions of requirements and desirable characteristics for the semantic objects mappings and crosswalks.

Suggested actions:

- Define and standardise representations for qualified relationships between semantic concepts (e.g. entity mappings, schema crosswalks) that can support data discovery and data integration.
- Ensure the adherence of mappings, crosswalks and other interoperability enabler semantic artefacts to FAIR (Findable, Accessible, Interoperable, and Reusable) principles (e.g. by leveraging the recommendations developed in the context of the FAIR Impact project). This includes making these artefacts accessible in catalogues of semantic artefacts.
- Develop and support best practices for implementing and sharing toolsets that enable semantic interoperability. This should involve entity mappings and schema crosswalks with standardisation labels, ensuring consistency and reliability in their usage.
- The establishment and implementation of effective governance mechanisms to oversee and manage the process of creating, maintaining, and using mappings and crosswalks, to guarantee their relevance and utility within EOSC.
- Develop practices to enforce consistency, reduce redundancy and implement version control for mappings, and crosswalks.

5. Recognise the semantic artefact catalogue as a critical part of the long-term viability of any research data infrastructure

Objective: Promote and emphasise the role of the semantic artefact catalogue component as a pivotal element in the long-term effectiveness and resilience of research data infrastructures to ensure the long-term access to and the discoverability of semantic artefacts.

Indicators of success:

- Adherence of semantic artefacts to the FAIR Principles for increased findability, enhanced usability and interoperability. Rich and accurate metadata for semantic artefacts are available in catalogues.
- Semantic artefact catalogues are used to find existing artefacts and to share newly created ones; the use of these catalogues becomes part of research workflows / processes.
- Widely accepted methods to assess semantic artefact catalogues and to guide their improvement.

Suggested actions:

- Develop and promote a maturity model to assess the catalogues of semantic artefacts. Integrate the maturity model with recommendations from other EOSC task forces for a comprehensive approach.

EOSC Association AISBL
Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

- Develop strategies to address found improvement areas for semantic artefacts catalogues (specific areas are, for example, to make catalogues dependable, machine-actionable and to support cross-domain interoperability).
- Ease the registration and maintenance of semantic artefacts in semantic artefact catalogues for enhanced discoverability and accessibility.
- Promote the uptake of established semantic artefacts catalogues.
- Select a common metadata standard for semantic artefacts and mandate its use across EOSC Nodes.

EOSC Association AISBL
Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

References

Bardi, A., Assante, M., & Mangiacrapa, F. (2023). ARIADNE: A Data Infrastructure for the Archaeological Research Community. *ERCIM News*, *133, Special theme Data Infrastructures and Management*, 8–10.

Bardi, A., Kraker, P., Mathiak, B., Widmann, H., Flügel, A.-L., Culina, A., Colomb, J., Goble, C., Heger, T., Hiseni, V., & Juty, N. (2022). *The Open Ecosystem of e-Infrastructures for Data Discovery: A Review*. https://doi.org/10.5281/ZENODO.7468089

Baumann, K., Corcho, O., Horsch, M. T., Jouneau, T., Molinaro, M., Peorni, S., Scharnhorst, A., Vancauwenbergh, S., & Vogt, L. (2021). *Task Force Charter Semantic Interoperability*.

Borgman, C. L. (2015). *Big Data, Little Data, No Data: Scholarship in the Networked World*. The MIT Press. https://doi.org/10.7551/mitpress/9963.001.0001

Broeder, D., Budroni, P., Degl'Innocenti, E., Le Franc, Y., Hugo, W., Jeffery, K., Weiland, C., Wittenburg, P., & Zwolf, C. M. (2021). *SEMAF: A Proposal for a Flexible Semantic Mapping Framework*. Zenodo. https://doi.org/10.5281/zenodo.4651421

Busse, C., Corcho, O., Ekaputra, F. J., Goble, C., Heibi, I., Jonquet, C., Lange, C., Le Franc, Y., Micsik, A., Palma, R., Peroni, S., Storti, E., & Widmann, H. (2023). *Raw data for the creation of a maturity model for Catalogues of Semantic Artefacts* (1.1) [dataset]. Zenodo. https://doi.org/10.5281/zenodo.8304972

Chambers, S., Palkó, G., Morselli, F., Ferguson, K., & Scharnhorst, A. (2023). *Book of Abstracts, DARIAH Annual Event 2023: Cultural Heritage Data as Humanities Research Data?* https://doi.org/10.5281/ZENODO.8340671

Corcho, O., Ekaputra, F. J., Heibi, I., Jonquet, C., Micsik, A., Peroni, S., & Storti, E. (2023a). *A maturity model for catalogues of semantic artefacts*. https://doi.org/10.48550/ARXIV.2305.06746

Corcho, O., Ekaputra, F. J., Heibi, I., Jonquet, C., Micsik, A., Peroni, S., & Storti, E. (2023b). *Catalogues of Semantic Artefacts—Maturity Dimensions and Features* (1.1) [dataset]. Zenodo. https://doi.org/10.5281/zenodo.8304959

Daga, E., Daquino, M., Fournier-S'niehotta, R., Guillotel-Nothmann, C., & Scharnhorst, A. (2023). *Documenting the research process. Opportunities and challenges for Bibliometrics and Information Retrieval*. https://doi.org/10.5281/ZENODO.8249512

Daga, E., Meroño Peñuela, A., Daquino, M., Ciroku, F., Musumeci, E., Gurrieri, M., Scharnhorst, A., Admiraal, F., & Fournier-S'niehotta, R. (2021). *D1.3: Pilots development – collaborative methodology and tools (V1.0)*. https://doi.org/10.5281/ZENODO.7712909

David, R., Baumann, K., Le Franc, Y., Magagna, B., Vogt, L., Widmann, H., Jouneau, T., Koivula, H., Madon, B., Åkerström, W. N., Ojsteršek, M., Scharnhorst, A., Schubert, C., Shi, Z., Tanca, L., & Vancauwenbergh, S. (2023). *Converging on a Semantic Interoperability Framework for the European Data Space for Science, Research and Innovation (EOSC)*. https://doi.org/10.5281/ZENODO.8042997

David, R., Mabile, L., Specht, A., Stryeck, S., Thomsen, M., Yahia, M., Jonquet, C., Dollé, L., Jacob, D., Bailo, D., Bravo, E., Gachet, S., Gunderman, H., Hollebecq, J.-E., Ioannidis, V., Bras, Y. L., Lerigoleur, E., & Cambon-Thomsen, A. (2020). FAIRness Literacy: The

EOSC Association AISBL
Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

Achilles' Heel of Applying FAIR Principles. *Data Science Journal*, *19*, 32. https://doi.org/10.5334/dsj-2020-032

David, R., Ohmann, C., Boiten, J.-W., Abadía, M. C., Bietrix, F., Canham, S., Chiusano, M. L., Dastrù, W., Laroquette, A., Longo, D., Mayrhofer, M. Th., Panagiotopoulou, M., Richard, A. S., Goryanin, S., & Verde, P. E. (2022). An iterative and interdisciplinary categorisation process towards FAIRer digital resources for sensitive life-sciences data. *Scientific Reports*, *12*(1), 20989. https://doi.org/10.1038/s41598-022-25278-z

Emery, T., Braukmann, R., Wittenberg, M., van Ossenbruggen, J., Siebes, R., & van der Meer, L. (2020). *The ODISSEI Portal: Linking Survey and Administrative Data*. https://doi.org/10.5281/ZENODO.4302096

EOSC Association. (2022). *The EOSC Partnership Monitoring Framework V6.6*. https://eosc.eu/monitoring-reporting/

EOSC Association. (2023). *Strategic Research and Innovation Agenda (SRIA) of the European Open Science Cloud (EOSC): Version 1.2 – 1 November 2023* (p. 212). https://eosc.eu/sria-mar

European Commission (DG DIGIT). (2017). *New European interoperability framework: Promoting seamless services and data flows for European public administrations*. Publications Office of the European Union. https://data.europa.eu/doi/10.2799/78681

European Commission (DG RTD). (2018). *Turning FAIR into reality: Final report and action plan from the European Commission expert group on FAIR data.* Publications Office. https://data.europa.eu/doi/10.2777/1524

European Commission (DG RTD), EOSC Executive Board, Corcho, O., Eriksson, M., Kurowski, K., Ojsteršek, M., Choirat, C., Sanden, M. van de, & Coppens, F. (2021). *EOSC interoperability framework: Report from the EOSC Executive Board Working Groups FAIR and Architecture*. Publications Office of the European Union. https://data.europa.eu/doi/10.2777/620649

Guillotel-Nothmann, C., De Berardinis, J., Bottini, T., Cathé, P., Daga, E., Daquino, M., Davy-Rigaux, A., Gurrieri, M., Van Kranenburg, P., Marzi, E., McDermott, J., Meroño Peñuela, A., Mulholland, P., Scharnhorst, A., & Tripodi, R. (2022). *D1.2 Roadmap and pilot requirements 2nd version*. https://doi.org/10.5281/ZENODO.7116561

Keith, J., & Broeder, D. (2024). *EOSCfuture Metadata Working Group Recommendations* (1.0). Zenodo. https://doi.org/10.5281/ZENODO.10497034

Kuhn, T., Magagna, B., & Schultes, E. (2023). *FAIR Implementation Profile (FIP) Ontology*. https://w3id.org/fair/fip/

Lacagnina, C., David, R., Nikiforova, A., Kuusniemi, M. E., Cappiello, C., Biehlmaier, O., Wright, L., Schubert, C., Bertino, A., Thiemann, H., & Dennis, R. (2023). *Towards a Data Quality Framework for EOSC*. https://doi.org/10.5281/ZENODO.7515816

Le Franc, Y., Parland-von Essen, J., Bonino, L., Lehväslaiho, H., Coen, G., & Staiger, C. (2020). *D2.2 FAIR Semantics: First recommendations*. https://zenodo.org/records/5361930

Magagna, B., Moncoiffé, G., Devaraju, A., Stoica, M., Schindler, S., Pamment, A., Environment Agency Austria, Austria/University of Twente, NL, National Oceanography Centre/British Oceanographic Data Centre, UK, Terrestrial Ecosystem Research Network (TERN), University of Queensland, Australia, University of Colorado, Boulder,

EOSC Association AISBL
Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

USA, Institute of Data Science, German Aerospace Centre (DLR), Germany, & National Centre for Atmospheric Science/UKRI, UK. (2022). *InteroperAble Descriptions of Observable Property Terminologies (I-ADOPT) WG Outputs and Recommendations* (Version 1). Research Data Alliance. https://doi.org/10.15497/RDA00071

Magagna, B., Schultes, E., Suchánek, M., & Kuhn, T. (2022). FIPs and Practice. *Research Ideas and Outcomes*, *8*, e94451. https://doi.org/10.3897/rio.8.e94451

Maineri, A. M., Morselli, F., & Braukmann, R. (2022). *Assessing FAIR Data Support Needs Amongst Data Supporters in the ODISSEI Community* (1.0). Zenodo. https://doi.org/10.5281/ZENODO.6524821

Nyberg Åkerström, W., & Maccallum, P. (2023). *Case studies and use cases as means of effectively demonstrating value and engaging stakeholders*. 5831708 Bytes. https://doi.org/10.17044/SCILIFELAB.21542313.V2

Nyberg Åkerström, W., Orten, H., Perseil, I., Jouneau, T., & Andersson, L. (2022). *Case study template for the EOSC-A Semantic Interoperability Task Force*. https://doi.org/10.5281/ZENODO.10508363

Ojsteršek, M. (2021). *Crosswalk of most used metadata schemes and guidelines for metadata interoperability* (1.0) [dataset]. Zenodo. https://doi.org/10.5281/ZENODO.4420116

Riley, J. (2004). *Understanding metadata*. NISO Press.

Scharnhorst, A., Flohr, P., Tykhonov, V., De Vries, J., Hollander, H., Touber, J., Hugo, W., Smiraglia, R., Le Franc, Y., Siebes, R., & Meijers, E. (2023). Knowledge Organisation Systems in the Humanities—Semantic Interoperability in Practice. *Proceedings of the Association for Information Science and Technology*, *60*(1), 1113–1115. https://doi.org/10.1002/pra2.962

Scharnhorst, A., Van Horik, R., Daga, E., Daquino, M., Musumeci, E., Van Kranenburg, P., Guillotel-Nothmann, C., Gurrieri, M., Presutti, V., Clementi, M., Meroño Peñuela, A., Turci, M., Marzi, E., Puglisi, A., & Fournier-S'niehotta, R. (2023). *D7.2 Data Management Plan (Second Version)*. https://doi.org/10.5281/ZENODO.7660299

Schweizer, T. J., & Baumann, K. (2023). *Developing a Flexible Linked Data Pipeline For Open Data Discovery*. SciDataCon2023, part of IDW2023, Salzburg, Austria. https://www.scidatacon.org/IDW-2023-Salzburg/sessions/570/paper/1140/

Sheth, A. P., & Larson, J. A. (1990). Federated database systems for managing distributed, heterogeneous, and autonomous databases. *ACM Computing Surveys*, *22*(3), 183–236. https://doi.org/10.1145/96602.96604

Soiland-Reyes, S., Goble, C., & Groth, P. (2023). *Evaluating FAIR Digital Object and Linked Data as distributed object systems*. https://doi.org/10.48550/ARXIV.2306.07436

Van Muijden, S. (2023). *Verrijk je collectie met termen: Handleiding Reconciliation met OpenRefine en het Termennetwerk*. https://doi.org/10.5281/ZENODO.7728880

Wilkinson, M. D., Sansone, S.-A., Méndez, E., David, R., Dennis, R., Hecker, D., Kleemola, M., Lacagnina, C., Nikiforova, A., & Castro, L. J. (2022). *Community-driven Governance of FAIRness Assessment: An Open Issue, an Open Discussion* (Final). Zenodo. https://doi.org/10.5281/ZENODO.7390482

EOSC Association AISBL
Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

Annex I: Brief case studies

Case study (Euro Virtual observatory - CNRS/CDS, ARI/U. Heidelberg, INAF):

European coordination to contribute to international collaborations

Euro-VO is the European coordination to contribute to the International Virtual Observatory Alliance (IVOA) that defines standards for interoperability for the astrophysical global community. The efforts connected with many projects and initiatives of the EU framework programme focus on maintaining and updating standards that for more than a decade have helped that an interoperable federation of infrastructures worldwide provides research in astrophysics a common way to integrate and analyse data.

Successful parts of the continuing effort is the definition of metadata schema or digital resources and service in astrophysics flanked by vocabularies and semantics.

Both the resources metadata schemas and the vocabularies are based on general standards, like OAI-PMH, Dublin Core, RDF (and XML in general). Operable solutions in metadata mapping were developed to provide DataCite metadata schema based translation of the resources.

A measure of the success can be seen both in the worldwide take-up of the standards in big research infrastructures in astrophysics (ESA, ESO, CTA, SKA, NASA, CSIRO, CADC) and in the attempts in onboarding specific resources in the EOSC.

Topics that might need further work and discussion relate to how recognizable the above semantics and metadata modelling solution are in terms of FAIR metrics (that are domain declined FAIR principle adherence), and in specific topics like usage of licences and definition of policy rights within a community that started from a completely public dissemination solution or its data resources and products.

EU projects attached to the above scenario goes from FP6/FP7 "Euro-VO" projects to ASTERICS to the H2020 ESCAPE cluster (now continuing as an in kind collaboration), part of EOSC Future and the oncoming OSCARS.

Case study (CMIP6 governance - IS-ENES/DKRZ):

Coordination across projects to agree on common standards

CMIP6 is an international effort coordinating huge Climate Model Intercomparison Projects (MIPs) where a lot of ESModelling centres are involved. The output of the CMIP6 simulations serves not only to compare and improve the Earth system models but provide experiments and climate projection datasets as a base for the assessment reports of IPCC-AR6[38] and other studies and publications. For this purpose, it is crucial that re-users of the CMIP6 data are given the means to identify and interpret the data sets they need for their studies and analysis. This could include, for example, requesting a specific version of a dataset or all data associated with a particular experiment. If the relevant information is available in machine-readable form as metadata—for example, in the PID kernel information—the desired data collections could be created automatically, avoiding tedious searching by the user and finally return a provenance record comprising all needed information in an interpretable and

[38] https://www.ipcc.ch/assessment-report/ar6/

EOSC Association AISBL
Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

understandable form. For this, we investigate the utilisation of components developed in the FAIRCORE4EOSC project: Specification of the PID kernel information as a detailed data type in the DTR[39] will allow machines to perform requested processes. In addition, we intend to use the PIDGraph[40] to interlink the dataset with other metadata needed for the processing; this comprises information on use constraints, which tools can be applied, and a list of variables it makes sense to apply the operation.

Case study (The Polifonia project's Research Ecosystem):

Designing an ecosystem of machine-actionable research data/tools

There is ongoing work to develop ontologies which helps to trace the actual research process on the very fine-grained level of research objects. An example is the RO-CRATE specification[41], which aims to foster open and reproducible science. Recently, as part of the Polifonia project[42], it has been argued that projects with a heavy load on collective software engineering might profit from a formalisation of research assets on a middle range level - between the individual research objects and the known work package/task project organisation - called Research Ecosystem[43]. By formalising essential components of research of types such as data, tools, and reports in a machine-readable way, inner project links can be made visible and re-use of shared data and methods can be encouraged outside of a specific research consortium. The implementation of this idea relies on an agile annotation scheme, together with clear workflows to select and curate relevant research components. The annotation scheme can be adapted to the project needs, but is cross-linked to other schemas such as schema.org or PROV-O[44]. It enables a machine-based evaluation of components against criteria of FAIRness (e.g. licences) including components from type data. This way, research data management becomes formalised as part of a wider formal project management. Complemented by a website based on the github implementation, it also supports humans in navigating through complex research processes.[45] (Daga et al., 2023, 2021; Guillotel-Nothmann et al., 2022; Scharnhorst, Van Horik, et al., 2023)

As another example for a research ecosystem in preparation: the combination of Project CEDAR[46] , OntoPortal[47] and Describo[48] services is used to create a collaborative environment for the creation, maintenance and use of community metadata schemas in the Hungarian

[39] EOSC Data Type Registry (DTR), https://faircore4eosc.eu/eosc-core-components/eosc-data-type-registry-dtr

[40] EOSC PID Graph, https://faircore4eosc.eu/eosc-core-components/eosc-pid-graph-pid-graph

[41] https://www.researchobject.org/ro-crate/

[42] Polifonia is an EC funded project, situated between Cultural Heritage (music) and semantic web technologies, https://polifonia-project.eu/

[43] https://github.com/polifonia-project/ecosystem

[44] https://www.w3.org/TR/prov-o/

[45] https://polifonia-project.github.io/ecosystem/

[46] https://more.metadatacenter.org/

[47] https://ontoportal.org/

[48] https://describo.github.io/

EOSC Association AISBL
Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

Scientific Data Repository Platform[49]. It allows research communities to re-use or design new metadata schemas with visual templates, and then to use them to describe items in RO-Crate packages. The schemas and RO-Crate metadata are collected together in a knowledge graph, which can be used both as a discovery and a reference service. This adds transparency to the process of supporting FAIR principles.

Case study (The ARIADNEplus portal and the Dutch Digital Heritage Network):

Semantic interoperability in the Humanities

The 'generic use case' of creating access to heterogeneous resources also (re)applies to problems inside of a domain. Data in the domain of social sciences and humanities could be described as part of the 'long tail of data', with a large number of relatively small collections rather than a few large data flagships. (Chambers et al., 2023) Simultaneously to attempts of harmonisation and standardisation, we still see an increase of the variation in data formats and structures, due to new research questions, new empirical materials which lead to new knowledge ordering systems. Researchers still use widely varying metadata schemas, vocabularies and thesauri which are often not even accessible to other humans (scholars) let alone machine-readable, so that in practice the data is interoperable only with extensive mapping exercises. The situation is exacerbated by the lack of definition of essential variables, which drives semantic data interoperability in several other domains. Despite this 'complexity due to heterogeneity challenge' there are examples of platforms which connect various source data.

The **ARIADNEplus portal**[50] provides access to almost four million archaeological data sources. In order to address the complexity of archaeological data integration, ARIADNE (an ERIC in the making) uses CIDOC CRM[51] as the backbone of its data model. Datasets of project partners coming from different European countries were mapped to this model and enriched with dating information via PeriodO and subject terms via the Getty AAT thesaurus. Partners needed to map from their local terms and thesauri first. Each dataset was then mapped and transformed as Linked Open Data and included in the ARIADNEplus Knowledge Base (Bardi et al., 2023). GraphDB[52] allows researchers to explore this Knowledge Base with an interface or through queries.

In the Netherlands, and connected to roadmap projects as CLARIAH.NL (which in turn is connected to the CLARIN and DARIAH ERIC's), the **Dutch Digital Heritage Network** defines itself as a network of networks with the ambition to enable public access to digitised cultural heritage. One of its main platforms is a registry of cultural heritage data collections[53]. This is built on Linked Data principles and connects to Dublin Core (DC), the Europeana Data Model (EDM), and schema.org. At its heart there is a 'network of terms' to which content providers

[49] https://science-research-data.hu/en

[50] https://portal.ariadne-infrastructure.eu/

[51] https://www.cidoc-crm.org/

[52] https://ariadne.d4science.org/web/ariadneplus_lab/

[53] https://datasetregister.netwerkdigitaalerfgoed.nl/faq-beheerders.php?lang=en

EOSC Association AISBL
Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

can equally contribute, and which is published and made accessible through an API (Van Muijden, 2023). (Scharnhorst, Flohr, et al., 2023)

Case study (CMIP6 'Semantic mapping of climate variables'):

Provenance information and variables customised to user needs

In the context of semantic interoperability in particular, the need for interoperable usage of CMIP6 data by impact communities - such as forestry, agriculture or tourism requiring climate data for their specific purposes - is crucial. But often these users can not interpret and choose the appropriate data needed.

To illustrate this with a concrete user story, think about a region affected by extreme forest dieback where the responsible forest managers are facing the challenge of identifying tree species that are well suited for reforestation in the region. On top of climate parameters such as temperature and precipitation, other factors also play an important role, e.g. vegetation duration, radiation, soil moisture, extreme events and whether native species associated with the tree exist for a species-rich ecosystem. Indirect effects such as the occurrence of pests like bark beetles must also be taken into account. This requires a mapping of CF standard names to vocabularies that can be understood and interpreted by the affected communities.

To overcome this gap, we examine a machinery that provides the non-expert user provenance information and climate variables customised for specific needs. For this, we examine how we can use the components MSCR[54] and DTR[55] developed in the FAIRCORE4EOSC project. We also consider services such as the NERC Vocabulary server[56] that uses the I-ADOPT Interoperability Framework[57] (Magagna, Moncoiffé, et al., 2022)for mapping variables to automate the requested processing of the data, in order to return a customised and understandable provenance record to the requestor.

Case study (The ODISSEI portal):

Multilingual access to data from a National Statistics Agency

ODISSEI (Open Data Infrastructure for Social Science and Economic Innovations)[58] is the national research infrastructure for the social sciences in the Netherlands. Through ODISSEI, researchers have access to large-scale, longitudinal data collections as well as innovative and diverse new forms of data. These can be linked to administrative data at Statistics Netherlands (CBS). The platform ODISSEI Portal[59] represents an interesting combination of a metadata harvester, for which it combines metadata from a wide variety of research data repositories into a single interface and workflows to enrich metadata automatically. Currently, the portal

[54] Metadata Schema and Crosswalk Registry (MSCR), https://faircore4eosc.eu/eosc-core-components/metadata-schema-and-crosswalk-registry-mscr

[55] EOSC Data Type Registry (DTR), https://faircore4eosc.eu/eosc-core-components/eosc-data-type-registry-dtr

[56] NERC Vocabulary Server (NVS), https://vocab.nerc.ac.uk/

[57] I-ADOPT Framework ontology, https://i-adopt.github.io/

[58] ODISSEI, https://odissei-data.nl/en/

[59] ODISSEI Portal, https://portal.odissei.nl/

EOSC Association AISBL
Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

includes metadata from more than 7500 social science datasets available at CBS, DANS, LISS, DataverseNL and HSN (The Historical Sample of the Netherlands).[60] By mapping of Dutch keywords to English terms in the European Social Sciences Language Thesaurus (ELSST), those keywords are visible in the metadata as linked terms that are connected to the ELSST vocabulary in the ODISSEI Portal Skosmos environment, enabling machine-actions. Mapping to these translations allows users for instance, to find the Dutch CBS metadata records while searching with English terms. For CBS data, metadata are enriched by providing information about the frequency of use. In the future, more metadata providers will be added to the Portal with the ultimate goal of giving researchers access to information about all relevant social science datasets in the Netherlands. (Emery et al., 2020; Maineri et al., 2022)

Case study (Semantic mapping of plant phenotyping variables):

Consistency in the face of changing technologies

For intrinsic reasons related to innovation objectives, information systems for phenotyping research data are fed by constantly renewed protocols. Newly-produced data are increasingly heterogeneous, multi-source, and are represented in multiple and diverse formats. All experiments need the assessment of multiple, related and temporally contextualised parameters. However, for historical reasons, high-throughput plant phenotyping is based on heterogeneous and expensive automated platforms where variables are designed by separated research teams and projects. To face growing challenges on food sovereignty in the context of climate change, interoperating this highly-experimental equipment and all their data by mapping the most used vocabularies could substantially increase the efficiency of multi-site and international research projects.

Case study (Semantic mapping of Highly Pathogenic Agents variables – ERINHA):

Cross-disciplinary models of research information and their representations

BSL 4 (Biosafety Level) laboratories are high-sensitivity laboratories working on class 4 pathogens (Marburg, Ebola, Influenza, etc., i.e. pathogens with a strong potential impact on human health and for which no effective remedy exists). They produce heterogeneous types of sensitive data in different fields of life sciences (molecular, cellular studies, animal experiments and context data...). These laboratories are cost-intensive. Furthermore, the access of their services is enabled via federated European calls for projects. The BSL4 equipment is federated within a Research infrastructure environment called ERINHA. Last years have shown how crucial it is to respond rapidly to pandemic situations linked to unknown pathogens, and to prepare the interoperation of the types of data produced in these contexts through the validation of standards and mapping of used vocabularies (David et al., 2022). A better produced semantic interoperability of the data would significantly improve the quality of the response to these challenges as well as the consistency of the projects using them, which could thus call on several laboratory equipment at the same time at the European level.

[60] https://odissei-data.nl/en/2023/10/new-version-of-the-odissei-portal-contains-enriched-metadata/

EOSC Association AISBL
Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

Case study (RESCS.org Ontology and data validation with SHACL-Shapes):

Virtual graphs for harmonised ontology-based data access [61]

The project Connectome at Switch aims to build a discovery platform and tools, APIs and data pipelines, with validation processes that allow Data Service Providers to integrate them to their technical environment.

The Linked Data Pipeline (LDP) is designed with transparency and flexibility in mind. The mapping process itself is defined declaratively as an RDF Mapping Language (RML) mapping and thus independent of an implementation. The RML mapping describes how elements in the source data are converted to RDF and also serves as documentation of how the knowledge graph (KG) was constructed. There are subtle differences when it comes to different RML engines, e.g. handling of non-ASCII chars for URI/IRI construction or the handling of empty string values. The preprocessing is procedural and we are trying to reduce pre-processing-tasks in favour of RML functions. RML functions are a relatively new concept and their adoption in an implementation-independent manner is still a work in progress. As a challenge: RDF seems an ideal choice for enhancing the FAIRness of data. However, the mere availability of data as RDF does not make it automatically FAIRer. Generating RDF requires stable and unique identifiers in the source data to express relationships between entities correctly. Also, quality issues in the source data such as syntactically invalid URLs, inconsistent dates etc. make the generation of RDF challenging. Ultimately, some of the problems can only be addressed by the data providers themselves since they have the necessary domain knowledge and can fix problems at the source.

[61] Schweizer & Baumann, 2023, https://www.scidatacon.org/IDW-2023-Salzburg/sessions/570/paper/1140/

EOSC Association AISBL
Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

Annex II: Links to supporting task force outputs

This appendix lists some of the documents used and produced by the EOSC Semantic Interoperability Task Force (TF SI). It includes EOSC-related documents, responses to requests for input addressed to TF SI by the EOSC Association and supporting outputs, such as articles, presentations, workshop proposals, and reports. Relevant document will be added to the EOSC-A Semantic Interoperability community on Zenodo: https://eosc.eu/eosc-task-forces

Responses to requests for consultations

- Response from the SI TF to the EOSC IF consultation
- TF SI - MAR 2025-2027 Headlines document
- EOSC Semantic Interoperability TF Topics for alignment.docx
- TF SI Copy of MAR-2025-27-draft02.docx
- Response to call for topics of interest
- Copy of Topic: A federated structure (objective 3)

Supporting task force outputs

EOSC Symposium 2023: Semantic Interoperability for data and metadata (slides)

Theme 1: Converging on a Semantic Interoperability Framework for the European Data Space for Science, Research and Innovation (EOSC)

- Theme1_EOSC_TF SI Scope Landscape Overview (WP1)
- Glossary Theme 1 - TF Semantic interoperability.xlsx
- Semantic interoperability landscape references.docx
- EOSC minimum metadata set recommendation.xlsx
- Semantic interoperability Framework.docx

Theme 2: A maturity model for catalogues of semantic artefacts, https://doi.org/10.48550/arxiv.2305.06746

- Catalogues of Semantic Artefacts - Maturity Dimensions and Sub-Criteria (Version 1.1). Version 1.1. https://doi.org/10.5281/zenodo.8304959
- Raw data for the creation of a maturity model for Catalogues of Semantic Artefacts (Version 1.1). Version 1.1. https://doi.org/10.5281/zenodo.8304972

Theme 3: Case studies and use cases as means of effectively demonstrating value and engaging stakeholders, https://doi.org/10.17044/scilifelab.21542313

- Definition of use case terms
- Case study template for the EOSC-A Semantic Interoperability Task Force, https://doi.org/10.5281/zenodo.10508363 Initial list of case studies and use cases that we can contribute

Presentations and other outreach activities

- TS SI - Presentations and other contributions
- TF SI - Relevant conferences and events

EOSC Association AISBL
Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931