Practical Semantic Interoperability recommendations for the EOSC Federation

Authorship community

Kurt Baumann (0000-0003-0627-8110)[1, 10]
Baptiste Cecconi (0000-0001-7915-5571)[2, 10]
Mark Doerr (0000-0003-3270-6895)[3, 9]
Esteban Gonzalez Guardia (0000-0003-4112-6825)[4]
Julien Homo (0009-0001-8997-1254)[5]
Clément Jonquet (0000-0002-2404-1582)[6]
Thomas Jouneau (0000-0001-5986-8128)[7, 10]
Joonas Kesäniemi (0000-0002-3770-0006)[8, 10]
Alexandra Kokkinaki (0000-0001-8042-6391)[9, 10]
Jiří Marek (0000-0003-2132-762X)[10,11]
Michal Med (0000-0002-3844-426X) [12]
András Micsik (0000-0001-9859-9186) [10, 13]
Christelle Pierkot (0000-0002-2591-3311)[14]
María Poveda-Villalón (0000-0003-3587-0367)[15]
Mattia Santoro (0000-0003-0401-3100) [10, 16]

[1] Switch, Werdstrasse 2, 8021 Zurich, Switzerland
[2] LIRA, Observatoire de Paris, Université PSL, Sorbonne Université, Université Paris Cité, CY Cergy Paris Université, CNRS, 92190 Meudon, France
[3] Universität Greifswald: Greifswald, Mecklenburg-Vorpommern, Germany
[4] Universidad Politécnica de Madrid, Madrid, Spain
[5] Foxcub, Issy-Les-Moulineaux, France
[6] INRAE, Montpellier, France
[7] Plateforme Recherche Data Gouv (UMS 1516), Université de Lorraine, Metz, France
[8] CSC – IT Center for Science, Finland
[9] British Oceanographic Data Centre, National Oceanography Centre, Southampton, UK
[10] Member, EOSC Association Task Force: Technical and Semantic Interoperability
[11] Institute of Computer Science, Masaryk University, Brno, Czechia
[12] Czech Technical University in Prague, Prague, Czechia
[13] HUN-REN SZTAKI, Budapest, Hungary
[14] Member, EOSC Association Secretariat
[15] Universidad Politécnica de Madrid, Madrid, Spain
[16] CNR (National Research Council of Italy)

Marcin Wichorowski (0000-0002-1498-9492)[17]

Contributors

Patrycja Antosz (0000-0001-6330-1597)[10, 18]
Javier de la Cueva (0000-0001-9152-2071) [10]
Daniel Garijo (0000-0003-0454-7145) [15]
Sveinung Gundersen (0000-0001-9888-7954)[10, 19]
Parul Tewatia (0000-0002-3096-1318) [10, 18, 20]

---

[17] Institute of Oceanology Polish Academy of Sciences (Polish National EOSC Node EOSC-PL)
[18] Center for Modeling Social Systems, NORCE Research AS, Bergen, Norway
[19] ELIXIR Norway, Department of Biosciences, University of Oslo, Norway
[20] Science for Life Laboratory, Uppsala University, Uppsala, Sweden

Table of contents

Executive summary

As the European Open Science Cloud (EOSC) evolves into a federation of autonomous nodes (EOSC Federation), semantic interoperability is essential to ensure that research data and metadata are consistently understood by both humans and machines across diverse systems and disciplines. This document establishes practical recommendations intended for semantic professionals of EOSC Nodes and the EOSC Federation, with one recommendation also for the EOSC Association, based on the analysis of the current status in the Landscape Overview for the EOSC Federation Semantic Interoperability in 2025. The main goal of the recommendations is to achieve semantic interoperability across the federation of EOSC Nodes on the level of datasets, metadata, mappings, catalogues, validation and governance.

The recommendations promote semantic interoperability in the EOSC Federation by providing data that are both human and machine interpretable, through annotations with RDF based ontologies and well described with DCAT based metadata. This allows understanding and combination of data. Semantic interoperability advances through the creation of mappings between semantic artefacts or across heterogeneous schemas, along with the publication of these artefacts in Semantic Artefact Catalogues. Another important part is the need for governance regarding semantic interoperability between Nodes of the Federation, as they are becoming independent. It is nevertheless important to keep in mind that all semantic artefacts shall remain compliant with the FAIR principles.

Additionally, this document provides recommendations on prospective issues of semantic interoperability regarding architecture of semantic services and approach of semantic interoperability towards AI.

Taken together, the recommendations in this document show how it is possible for the EOSC Federation to align technical solutions with sustainable governance and targeted funding, the EOSC Federation aims to ensure that semantic interoperability, understood as "what is sent is what is understood", can become a key characteristic of the system.

List of abbreviations

- **AGROVOC**: A multilingual thesaurus covering all areas of interest to the Food and Agriculture Organization of the United Nations.
- **AI**: Artificial Intelligence.
- **API**: Application Programming Interface; a set of protocols for building and integrating software applications.
- **CC**: Creative Commons
- **CDIF**: Cross-Domain Interoperability Framework; a set of guidelines for aligning diverse metadata standards.
- **CSVW**: CSV on the Web; a standard for describing column-level semantics in tabular data.
- **DCAT**: Data Catalogue Vocabulary; a baseline schema for metadata profiles describing datasets and services.
- **DCAT-AP**: DCAT Application Profile for data portals in Europe.
- **DQV**: Data Quality Vocabulary; a W3C standard for attaching quality assessments to data.
- **EOSC**: European Open Science Cloud.
- **EOSC-A**: EOSC Association.
- **EOSC-IF**: EOSC Interoperability Framework.
- **FAIR**: Findable, Accessible, Interoperable, and Reusable principles for data management.
- **GCMD**: Global Change Master Directory (Earth Sciences vocabulary).
- **GEMET**: GEneral Multilingual Environmental Thesaurus.
- **I-ADOPT**: Interoperability and FAIRness through Adoption of Conceptual Modelling; a framework for scientific variable descriptions.
- **IRI**: Internationalized Resource Identifier.
- **ISO**: International Organization for Standardization.
- **JSKOS:** JSON for Knowledge Organization Systems
- **JSON-LD**: JavaScript Object Notation for Linked Data.
- **LinkML**: Linked Data Modeling Language
- **ML**: Machine learning
- **MOD**: Metadata for Ontology Description.
- **MOD-API**: An API specification for accessing semantic descriptions of semantic artefacts.
- **NCC**: Node Coordination Committee.
- **NFDI:** German National Research Infrastructure
- **NOC**: Node Operations Committee

- **OLS**: Ontology Lookup Service
- **ORCID**: Open Researcher and Contributor ID.
- **OWL**: Web Ontology Language.
- **PID**: Persistent Identifier.
- **PROV-O**: Provenance Ontology; a standard for provenance tracking.
- **QA**: Quality Assurance
- **RDA**: Resource Description and Access
- **RDF**: Resource Description Framework.
- **RDFS**: RDF Schema; a set of classes with certain properties providing basic elements for the description of ontologies.
- **RML**: RDF Mapping Language.
- **RO-Crate**: Research Object Crate, a community effort to establish a lightweight approach to packaging research data with their metadata.
- **ROR**: Research Organization Registry.
- **SA**: Semantic Artefact.
- **SAC**: Semantic Artefact Catalogue.
- **SHACL**: Shapes Constraint Language; used for validating RDF metadata against defined profiles.
- **SKG:** Scientific Knowledge Graph
- **SKG-IF**: Scientific Knowledge Graphs Interoperability Framework.
- **SKOS**: Simple Knowledge Organization System.
- **SKOSMOS**: A web-based SKOS browser and publishing tool
- **SLA**: Service Level Agreement
- **SOSA/SSN**: Sensor, Observation, Sample, and Actuator / Semantic Sensor Network ontologies.
- **SPDX**: The System Package Data Exchange
- **SSSOM**: Simple Standard for Sharing Ontological Mappings.
- **SWRL**: Semantic Web Rule Language
- **URI**: Uniform Resource Identifier
- **W3C:** World Wide Web Consortium

1. Introduction

The present document provides practical recommendations to increase the semantic interoperability of scientific resources during the ”Build-up Phase” of the EOSC Federation. The report is based on the *2025 Landscape Overview*[21] recently published, that retrieved the view about semantic interoperability from EOSC-related projects and relevant initiatives. The recommendations follow these results and are anchored in the results of EOSC-related projects aimed to mature the **EOSC Interoperability Framework (EOSC-IF)**[22]. The core principle driving this effort is the concept that "**what is sent is what is understood**". This document puts this principle into practice by defining semantics at two distinct but interdependent levels: The **Scientific/Conceptual Level** addresses the actual meaning and intent of the concepts used to describe data and metadata. It ensures that scientific variables, values, and their relationships are interpreted consistently across different organizational and disciplinary contexts, preventing "misunderstandings" when research outputs are reused. The **Technical/Structural Level** provides the framework required to make that meaning machine-actionable. It involves the use of semantic artefacts serialized in formats like JSON-LD or Turtle to provide a formal, shared language for knowledge representation. While technical implementations often rely on RDF models and IRI-identified vocabularies to eliminate ambiguity, the ultimate goal of "Semantic" interoperability in the EOSC Federation is to preserve the precise intent and context of Digital Objects for both humans and machines throughout the research lifecycle.

Achieving this requires semantic alignment at the data and metadata levels, cross-referencing across catalogues and various resources, enabling interoperability of resources from different origins, and ultimately establishing the conditions under which cross-disciplinary research may be carried out.

The landscape analysis reveals that the research community is strongly engaged in topics related to semantic interoperability, with nearly all surveyed projects and EOSC Nodes recognising it as a key ingredient to improve data discovery and FAIRness. However, current efforts remain fragmented, with semantic interoperability remaining solved only locally, at the level of repositories and organisational implementations. There is a strong consensus that EOSC must transition towards a coordinated, standards-based ecosystem. This involves a shift toward machine-actionability through the use of semantic artefacts in the form of RDF graphs serialized

[21] Landscape overview for the EOSC Federation Semantic Interoperability in 2025, https://doi.org/10.5281/zenodo.19134873 (last accessed 24/03/2026)

[22] *EOSC interoperability framework – Report from the EOSC Executive Board Working Groups FAIR and Architecture*,, 2021, https://data.europa.eu/doi/10.2777/620649 (last accessed 25/03/2026)

in formats like Turtle or JSON-LD, and the replacement of free-text keywords with IRI-identified controlled vocabularies. By adopting standardised profiles such as DCAT-AP and aligning with cross-domain frameworks, the EOSC Federation can achieve consistent and interoperable metadata management. The validation of semantic artefacts further ensures their coherence before inclusion in catalogues. Together, these practices support the entire research lifecycle.

The feedback from participants in the survey done for the *2025 Landscape Overview* highlights that to scale effectively, technical solutions must be matched with sustainable governance and targeted funding. This report identifies several areas that are critical for progress, including interoperability of metadata and data, the implementation of robust mappings, and the operation of Semantic Artefact Catalogues (SACs). Furthermore, validation of these artefacts through tools like SHACL and LinkML is essential for maintaining a high-performance research environment. By aligning these efforts with the evolving requirements of the Common European Data Spaces[23], EOSC can transition from a collection of federated nodes into a cohesive, interoperable ecosystem.

The remaining of the document is organised as follows:

Chapter 2 contains the practical recommendations considered critical for the future development of semantic interoperability within the EOSC Federation.

The recommendations are addressed to the EOSC Nodes or the EOSC Federation, with one recommendation to the EOSC Association. They are categorised in three levels of importance: high-priority (crucial for ensuring semantic interoperability, require immediate action), medium-level (specific to some use cases), and low-priority (to enhance semantic interoperability).

Chapter 3 describes two areas of emerging importance: these are topics that have not received enough attention so far, but are considered critical to ensure semantic interoperability in the EOSC Federation keeps up with current developments.

The recommendations can serve as a self-assessment tool for current and future EOSC Nodes to evaluate their own maturity on semantic interoperability. The EOSC Association and Federation Governance are invited to use this input to provide the normative guidance and coordination necessary to overcome remaining bottlenecks. Ultimately, the goal is to foster a research environment where semantic alignment is a cross-cutting concern, from discovery and access to governance and provenance, enabling that scientific data can be seamlessly interpreted and reused by the research community.

[23] https://digital-strategy.ec.europa.eu/en/policies/data-spaces

2. Practical recommendations for the EOSC Federation

The recommendations in this chapter are intended to support EOSC Nodes, the EOSC Federation, and the EOSC Association in their advance from fragmented, domain-specific implementations of semantic interoperability towards a coordinated and standards-based interoperability ecosystem.

Semantic interoperability must be addressed across the whole research lifecycle to ensure that research outputs can be interpreted consistently by humans and machines across different systems, disciplines, and organisational contexts, with the principle that "what is sent is what is understood" as the core objective to be achieved in the EOSC Federation.To facilitate broad adoption, we recommend the use of JSON-LD (JSON Linked Data), which bridges the divide between developer-friendly formats like JSON or YAML and the graph-based requirements of RDF. By leveraging JSON-LD features, the Federation enables developers to interact with hierarchical structures while maintaining the semantic integrity necessary for interoperability at both the data and metadata levels

The chapter is structured into six sections:

2.1 Semantic interoperability at the data level addresses how the meaning of data values, variables, relationships, and dataset structures can be made explicit and machine-processable.

2.2 Semantic interoperability at the metadata level focuses on shared metadata models, application profiles, controlled vocabularies, identifiers, and validation mechanisms that support discovery and reuse.

2.3 Semantic mappings describe how correspondences between schemas, vocabularies, semantic artefacts, and data structures can be expressed, maintained, and reused.

2.4 Semantic artefact catalogues addresses the role of catalogues in making vocabularies, ontologies, mappings, and other semantic artefacts discoverable, accessible, and reusable.

2.5 Validation and FAIRness of semantic artefacts focuses on the technical quality, consistency, and FAIR management of semantic artefacts.

2.6 Semantic interoperability governance pathways within the Federation describes the organisational and governance mechanisms needed to coordinate semantic interoperability across the autonomous EOSC Nodes that form the EOSC Federation.

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

Each section includes a description of the key area, followed by a set of recommendations. The recommendations are addressed to the relevant actor: EOSC Nodes (N), the EOSC Federation (F), or the EOSC Association (A). Further, they are prioritised according to their expected relevance for achieving interoperability to the EOSC Federation.

The priority levels are defined as follows:

- **Critical** - Recommendations considered crucial for enabling semantic interoperability across the EOSC Federation and which must be implemented as soon as possible,
- **Important** - Recommendations that are important for specific use cases, communities, or implementation scenarios and that significantly improve interoperability where applicable,
- **Useful** - Recommendations that further enhance semantic interoperability, support maturity, or prepare for future developments, but are not immediately needed for basic interoperability.

Together, the recommendations collected in this chapter provide a practical "roadmap" for improving semantic interoperability across EOSC Nodes, and serve as a self-assessment instrument for current and future EOSC Nodes to evaluate semantic maturity and identify areas where further alignment and implementation or governance support is needed.

2.1 Semantic interoperability at the data level

Semantic interoperability at the data level refers to the ability to interpret data values (and not just data variables) consistently across systems, organisations, and scientific domains. This is a non-trivial problem, since data (understood as the combination of variables and values) is always context-dependent. Therefore, such a 'translation' can only succeed if enough information about this context is provided to allow for its correct interpretation. Data-level semantic interoperability ensures that the meaning of the data values and their relationships and the understanding across systems is explicitly stated and machine-processable - the "What" and "Why".

The practical recommendation should therefore be selective: repository owners should first describe the dataset and its structure through metadata, and then add data-level semantics for the parts of the dataset where meaning and relationship is most important for reuse.

1. **Semantic Interoperability at data level can be expressed on various ways:**
    a. Using RDF-based representations, in which datasets are modelled as RDF graphs, entities are identified by IRIs, and semantic relations are expressed explicitly, where possible. Such datasets should be accompanied by an RDF schema or ontology that

defines their structure and meaning in a machine-readable manner. This can be realised by publishing data directly in RDF, or by transforming data from other formats into RDF. Where appropriate, tools such as RML[24], LinkML[25], or OpenRefine[26] should be used to align existing data structures with semantic models.

b. Using schemas, controlled vocabularies, semantic annotations, mappings or contexts such as JSON-LD or CSVW[27] for tabular data. In these cases, interoperability is often limited by the original data structure, which may still need to be adapted or transformed to achieve full machine-readable semantic integration.

c. Enable systems to understand, compare, combine and reuse datasets consistently, especially in Machine Learning (ML) and Artificial Intelligence (AI) workflows that use structured, machine-readable approaches to describe internal content, such as Croissant ML[28].

2. **Implement semantic interoperability for non-RDF datasets** by providing machine-readable descriptions of the dataset structure and content, including files, tables, fields, variables, data types, meanings, relationships, controlled vocabulary references, and mapping where needed. These descriptions should make the meaning of the data explicit and processable without requiring the dataset itself to be published directly as RDF.

3. **To ensure FAIRness and trustworthiness and interoperability of data**, it is crucial to accompany each dataset by machine-readable metadata that describe its structure, content, and meaning. General requirements for metadata descriptions are provided in Section 2.2.

4. **Interoperability mappings:** To enable data exchange across systems and domains, the use of mappings is recommended to relate data structures, values, and semantic artefacts, in case there is no shared agreement on common concepts and their relations.. Further guidance on the description of mappings is provided in Section 2.3.

5. **Bundle and semantic artefacts:** Where appropriate, datasets should be published as semantic bundles (see ANNEX) rather than as isolated data files. Such bundles should include machine-readable artefacts describing the dataset's structure, semantics, and context, so that the data becomes more understandable and actionable for humans and machines. The semantic description should also be treated as a first-class artefact and made available as a separate, retrievable resource, preferably with its own persistent identifier. Bundle approaches such as

[24] RML - https://rml.io/ (last accessed 15.03.2026)
[25] LinkML - https://linkml.io/ (last accessed 15.03.2026)
[26] OpenRefine - https://openrefine.org/ (last accessed 15.03.2026)
[27] CSVW - https://csvw.org/ (last accessed 16.03.2026)
[28] CroissantML - https://openreview.net/forum?id=a0WAM6q6fV#discussion (last access 04.05.2026)

RO-Crate[29] or SciDatS[30] may be used to package and expose the dataset together with its associated semantic artefacts in an interoperable manner.

Table 1: Recommendations at the data level

Recommendation	Short description	Priority
2.1.1 (N)	Implement full RDF datasets for RDF data or semantic interoperability for non-RDF datasets.	Critical
2.1.2 (N)	Support machine-readable dataset descriptions for ML and AI workflows	Useful
2.1.3 (N)	Ensure FAIRness and trustworthiness through metadata.	Critical
2.1.4 (N)	Use mappings to support interoperability across systems and domains	Important
2.1.5.(N)	Publish datasets as semantic bundles where appropriate	Important

2.2 Semantic interoperability at the metadata level

Semantic interoperability at the metadata level refers to the capacity for metadata created in one system, domain or schema to be understood and processed meaningfully by other federated systems through shared semantics, formal models and machine-readable definitions. It is crucial for transforming static descriptions into dynamic, machine-actionable information, enabling automated data discovery and cross-resource linking.

1. EOSC Nodes should adopt, where applicable, the latest version of DCAT[31] as the baseline vocabulary for metadata profiles describing datasets (dcat:Dataset), data catalogues (dcat:Catalog) and data services (dcat:DataService), thereby establishing a shared foundation for resource discovery[32].

[29] RO-Crate - https://www.researchobject.org/ro-crate/ - (last accessed 13.03.2026)
[30] SciDatS - https://gitlab.com/opensourcelab/scientificdata/scidats - (last accessed 04.05.2026)
[31] Data Catalogue Vocabulary - https://www.w3.org/TR/vocab-dcat-3/ (last accessed 24/03/2026)
[32] This approach was also explicitly adopted by the DSSC in the specifications for Data Spaces. As the European Commission considers EOSC to be a Data Space, compliance with this requirement is therefore mandatory for EOSC.

a. EOSC Nodes are encouraged to adopt existing DCAT application profiles (AP) where possible (e.g. DCAT-AP[33], GeoDCAT-AP[34], HealthDCAT-AP[35], DCAT-AP-PLUS[36]) or by defining new ones when necessary. Before developing new profiles, the Nodes should assess whether existing profiles already meet their requirements.The Semantic Interoperability Centre (SEMIC) Team has published guidance for creating new profiles of DCAT-AP 3.0.1[37].
b. When DCAT or the profile adopted is not enough to describe a dataset in detail, additional ontologies or schemas should be added to provide richer, domain-relevant information (such as PROV-O for provenance tracking or I-ADOPT for scientific variable descriptions, Data Quality Vocabulary (DQV) for quality etc.).
    i. DCAT 3.0 specification provides guidance[38] on how to include additional information, such as provenance and quality, by integrating complementary ontologies like PROV-O[39] and DQV[40].
    ii. For observable properties or scientific variables semantic artefacts that are decomposed using dedicated frameworks such as I-ADOPT[41], SOSA/SSN[42], OBOE[43] should be used to enable semantic interoperability of observable properties.
    iii. For social and economic data, we recommend using DDI and DDI-CDI (Cross-Domain Integration) as key models for achieving interoperability at the metadata and data levels.
    iv. For data access or reuse constraints as well as legal basis and purpose for data collection and consent or privacy statements, create Open Digital Rights Language (ODRL) policies utilizing the Data Privacy Vocabulary (DPV)[44].
c. For EOSC Nodes that do not natively implement DCAT (e.g., ISO 19115 metadata standard[45] for georeferenced data, or DataCite[46] to describe persistent, citable research

[33] DCAT-AP - https://semiceu.github.io/DCAT-AP/ (last accessed 27/03/2026)
[34] GeoDCAT-AP - https://semiceu.github.io/GeoDCAT-AP/ (last accessed 27/03/2026)
[35] HealthDCAT-AP Release 6 - https://healthdataeu.pages.code.europa.eu/healthdcat-ap/releases/release-6/ (there seems to be no nice URL)
[36] DCAT-AP+ - https://zenodo.org/records/17702370 (last accessed 27/03/2026)
[37] How to create your DCAT-AP profile, A guide for creating new DCAT-AP profiles - https://semiceu.github.io/DCAT-AP-reuse-guidelines/ (last accessed 05/05/2026)
[38] W3C recommendation, Quality information - https://www.w3.org/TR/vocab-dcat-3/#quality-information (last accessed 27/03/2026)
[39] https://www.w3.org/TR/prov-o/ (last accessed 27/03/2026)
[40] W3C recommendation, Data Quality Vocabulary - https://www.w3.org/TR/vocab-dqv/ (last accessed 27/03/2026)
[41] https://i-adopt.github.io/ (last accessed 27/03/2026)
[42] https://www.w3.org/TR/vocab-ssn/ (last accessed 27/03/2026)
[43] https://bioportal.bioontology.org/ontologies/OBOE (last accessed 27/03/2026)
[44] DPV : https://w3c.github.io/dpv/2.3/dpv/ (last accessed 21/04/2026)
[45] ISO 19115-1:2014 - https://www.iso.org/standard/53798.html (last accessed 28.03.2026)
[46] DataCite metadata schema : https://datacite-metadata-schema.readthedocs.io/en/4.7/ (last accessed 16.04.2026)

objects, typically identified by a DOI, or communities relying on Schema.org[47]), it is recommended to provide mappings between their metadata profiles and DCAT. Further details on mappings and crosswalks are provided in Section 2.3.

2. **Domain vocabularies and standardised value sets** should be used as much as possible to populate metadata properties, thereby reducing ambiguities in interpretation and minimising lexical errors or the need for mappings, while also enabling cross-linking among datasets. Vocabulary terms should be referenced by their full IRI identifiers and human-readable label (if applicable):
    a. Standard identifiers should be used to describe contextual entities, such as SPDX for licenses[48], ROR for organizations and funding agencies[49], ORCID for people[50], RAID[51] for research activities, PIDINST[52] for instruments, etc.
    b. Frequently used controlled vocabularies commonly based on SKOS, recognized by thematic communities (e.g. AGROVOC in agronomy, GCMD in earth sciences, GEMET), should be used or enriched to populate metadata properties wherever applicable, e.g. keywords or themes.
    c. Universals[53] should be used to refer to domain-independent metadata elements. For each universal, values should be expressed together with their reference framework such as, coordinate reference system, calendar or temporal standard, or recognised unit vocabulary. The CDIF practical guidance should be followed to ensure consistent and machine-interpretable representations of universals.
3. **Dedicated semantic artefacts should be used to richly describe other kinds of research outputs:**
    a. CodeMeta[54] should be adopted to standardise the description of research software across repositories, improving citation, discovery, and reuse within scientific workflows.
    b. MOD[55] should be used for describing semantic artefacts, and MOD-API to ensure interoperability across Semantic Artefact Catalogs (discussed in Section 2.4)

---

47 Schema.org - http://schema.org (last accessed 28.03.2026)
48 SPDX License list - https://spdx.org/licenses/ (last accessed 28.03.2026)
49 Research Organisation Registry (ROR) - https://ror.org/ (last accessed 28.03.2026)
50 ORCID - https://orcid.org/ (last accessed 28.03.2026)
51 RAID : https://www.raid.org/ (last accessed 16.04.2026)
52 PIDINST : https://www.pidinst.org/ (last accessed 16.04.2026)
53 Cross-Domain Interoperability Framework (CDIF) https://cross-domain-interoperability-framework.github.io/cdifbook/universals/univintro.html (last accessed 28.03.2026)
54 https://codemeta.github.io/ (last accessed 28.03.2026)
55 Ontology Specification Draft - https://fair-impact.github.io/MOD/index-en.html (last accessed 28.03.2026)

c. The SKG-IF[56] framework and its associated data model should be applied to enhance the exchange of metadata about research products and other contextual information, ensuring both technical and semantic interoperability across diverse Scientific Knowledge Graphs (SKGs).

4. **For multi-domain or cross-disciplinary scenarios** within the EOSC Federation, CDIF[57] implementation guidelines (v.1.1[58]) should be used to establish a domain-neutral metadata bridge that aligns diverse standards into interoperable “exchange” profiles while preserving domain-specific semantics.

5. **To ensure consistent implementation and machine-processable validation**, EOSC Nodes should define data modelling and validation mechanisms such as SHACL[59] or LinkML[60] shapes alongside their application profile. SHACL constraints enable formal validation of metadata against the defined profiles, enabling automated compliance checks, and strengthening semantic interoperability across EOSC services (Further details on validation checks are provided in Section 2.5).

Table 2: Recommendations at the metadata level

Recommendation	**Short Description**	**Priority**
2.2.1 (N)	Use the latest version of DCAT as the baseline schema for datasets, catalogues, and data services, reusing existing application profiles	Critical
2.2.2 (N)	Use standardized domain vocabularies and identifier schemes.	Critical
2.2.3 (N)	Use dedicated semantic artefacts and frameworks to provide rich descriptions and exchange metadata for other research outputs than data.	Important
2.2.4 (F)	Use CDIF guidelines to build domain-neutral exchange	Useful

[56] Scientific Knowledge Graphs Interoperability Framework (SKG-IF) https://skg-if.github.io/ (last accessed 28.03.2026)
[57] For an introduction to the Cross-Domain Interoperability Framework (CDIF), see. https://cross-domain-interoperability-framework.github.io/cdifbook/introduction.html (last accessed 28.03.2026)
[58] CDIF model version 1.1 : https://cross-domain-interoperability-framework.github.io/cdifbook/
[59] W3C Documentation - SHACL - https://www.w3.org/TR/shacl/ (last accessed 28.03.2026)
[60] General Purpose Modeling Language - LinkML - https://linkml.io/ (last accessed 28.03.2026)

	metadata profiles for multi-domain and cross-disciplinary scenarios in EOSC.	
2.2.5 (N)	Define data modelling and validation mechanisms alongside application profiles.	Useful

2.3 Semantic mappings

Mappings are an essential building block of interoperability when data or metadata crosses domain or system boundaries. In addition to enhancing semantic interoperability by aligning meaning, mappings facilitate technical interoperability by enabling transformation of data structure. The correspondences between data formats or structures required to achieve semantic interoperability can be expressed through predicate-based mapping relationships which work in one or both directions (e.g. skos:broadMatch vs. skos:exactMatch) depending on the semantics of the predicate. For distribution, a set of predicate-based mapping relationships must be formalized using a machine readable mapping specification.

1. **Recommendations for mappable semantic artefacts:**
   a. Mappings should be created between maintained versions of semantic artefacts.
   b. Semantic artefacts being mapped must be identified with identifiers which are globally unique, should be persistent, and preferably resolvable.
   c. Individual mappable elements within the semantic artefact should follow the recommendations from the previous step.
2. **Recommendations for creating mapping specifications:**
   a. Simple predicate-based semantic mappings should be published using serialized SSSOM[61]. Predicates can be chosen based on the level of ontological commitment required. For this task SKOS should be used as default, and RDFS and OWL should be used if support for inference is needed. RDF graph representation of the SSSOM mapping should also be used when appropriate.

[61] SSSOM a Simple Standard for Sharing Ontological Mappings: https://mapping-commons.github.io/sssom/dev/ (last accessed 22.06.2026)

b. More complex mappings between Semantic Artefacts can be expressed with languages such as JSKOS, SWRL, or OWL. SSSOM mapping of metadata properties should be used with other models if possible.

c. Schema mappings transforming data structures and values between schemas (e.g. crosswalks and schema matching) should include the necessary metadata and documentation for execution[62] (further details on metadata are provided in section 2.2), and be described using the FAIR mappings metadata model recommended by the RDA FAIR Mappings Working Group[63] to enable harmonised harvesting by mapping registries.

3. **Recommendations for sharing mapping specifications:**

   a. In order to support the proliferation of FAIR mappings as curated and maintained Semantic Artefacts, the EOSC Federation must support mapping registry that adhere to a standard data schema and provide an API for harvesting/retrieving mappings.

   b. Mapping sets should be published in mappings registries that should ideally be federated.

4. **Recommendation for mapping reuse:**

   a. Mappings should be executable at query time, not just published. A federation-layer mediation service that uses registered mappings to dynamically translate user queries,[64] ensuring consistent cross-node results without requiring custom user logic.

Table 3: Recommendations for Semantic Mappings

Recommendation	Short Description	Priority
2.3.1(N)	Ensure all mapping specifications, semantic artefacts, and their individual elements are uniquely identified.	Critical
2.3.2 (N)	Utilize SSSOM as the default and use a FAIR mapping metadata model to describe mappings.	Important
2.3.3 (F)	Provide Federation-level support for mapping registry	Critical

[62] FAIR mapping schema - https://mapping-commons.github.io/fair-mappings-schema/ (last accessed 28.03.2026)
[63] RDA FAIR Mappings working group - https://mapping-commons.github.io/rda-fair-mappings/ (last accessed 30.03.2026)
[64] For example resolving a search for 'Licence=CC-BY' across a node using the CC URI and another using a local 'CCBY4' string

	services that adhere to standard data schemas and offer APIs.	
2.3.4 (N)	Publish mappings in the mapping registry.	Important
2.3.5 (F)	Enable dynamic, real-time query translation across nodes through an executable mapping and federation-layer mediation service.	Useful

2.4 Semantic Artefact Catalogues

Semantic Artefact Catalogues (SACs) are the primary interface for accessing and managing Semantic Artefacts, and enable reaching a comprehensive view of SAs operated in the EOSC Federation. Leveraging SACs is key to facilitate discoverability and the uniform use of SAs across the Federation[65]. The following recommendations are conceived with the goal of maximising their impact and ensure a cohesive, interoperable and sustainable EOSC ecosystem:

1. **Centralising Management of SACs:**
    a. Each EOSC Node should publish the Semantic Artefacts (SAs) it produces, manages, endorses, or uses in a Semantic Artefact Catalogue (SAC). To promote autonomy, best community practices and avoid bottlenecks, EOSC Nodes should ensure a harmonised and homogeneous access to SAs by contributing to manage any SA they leverage.
    b. To accelerate implementation, adoption and federation, we recommend leveraging existing solutions and technologies for the implementation of SACs, such as OntoPortal, SKOSMOS, VocPrez or OLS.
2. **SA representation, PIDs and queries in SAC:**
    a. To enable consistent access and simplify integration with other services and platforms in the EOSC Federation, SAs in SACs should be represented in an RDF-compliant format, leveraging relevant standards like SKOS, RDF(S) or OWL that support semantic richness and enable advanced reasoning.
    b. SAs must support URIs functionalities to ensure long-term access and provenance, and should support relevant Persistent Identifiers (PIDs).

[65] Some of these recommendations are directly inspired from FAIR-IMPACT's WP4: https://doi.org/10.5281/zenodo.14643278 (last accessed 27.03.2026)

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

c. SACs should have a publicly accessible SPARQL endpoint to enable efficient semantic queries on SAs data and metadata and facilitate cross queries within the EOSC Federation.

3. **API, Standards and compliance with the EOSC Federation:**
    a. To ensure seamless integration and interoperability, endorsed SACs must provide an API according to agreed standards to access and use their content. At a minimum, community schemas such as Dublin Core (minimal requirement) should be used. More advanced ones, such as Metadata for Ontology Description Application Programming Interface (MOD-API)[66], which is fully compliant with DCAT,should be used to support semantic descriptions of SA, SA distributions, SA catalogs and services.
    b. SAs themselves should be treated as specific resources and be included as a distinct research product type in the EOSC Federation to ensure they are more easily discoverable throughout the Federation.
4. **SAC management and improvement:**
    a. Content curation and governance[67] should be enabled when relevant.
    b. SA management teams should include domain experts to provide the necessary technical oversight and guidance. A human review process is essential; all submissions and the registration of new editors must be reviewed by a gatekeeper before publication
    c. EOSC Nodes should implement a process for continuous monitoring and improvement, incorporating feedback from users and stakeholders. To ensure proper SAs evaluation by users at the discovery stage, SACs should track and report on key metrics, including artefact usage frequency, update frequency, and FAIRness assessment, to identify and address any potential issues.

By prioritizing these aspects, SACs will become a cornerstone of the EOSC Federation, promoting data discoverability, fostering collaboration, and unlocking the full potential of European research data.

Table 4: Recommendations for Semantic Artefact Catalogues

[66] FAIR-IMPACT MOD-API - https://github.com/FAIR-IMPACT/MOD-API (last accessed 27.03.2026)
[67] Content governance refers to the policies, workflows, and quality control mechanisms that determine what concepts are included in an SA, how definitions are curated, how changes are reviewed, and how domain correctness, consistency, and provenance are maintained.

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

Recommendation	Short Description	Priority
2.4.1 (N)	Ensure that all semantic artefacts produced or used by a node are discoverable and managed through a SAC, leveraging existing platforms.	Critical
2.4.2 (N)	SACs should represent all semantic artefacts using RDF-based standards, assign persistent identifiers and stable URIs, and provide a publicly accessible SPARQL endpoint.	Critical
2.4.3 (N)	SACs should provide a well-defined standard API	Critical
2.4.4 (N)	SACs should implement robust content governance and curation process	Critical
2.4.5. (N)	SACs should establish continuous monitoring and improvement processes, supported by a statement of sustainability and an editorial charter.	Useful

2.5 Validation and FAIRness of Semantic Artefacts

For Semantic Artefacts to support cross-domain discovery, integration, and reuse of research outputs effectively, they must be aligned with the FAIR principles. Validation is essential to ensure the structural and logical consistency of semantic artefacts (e.g., RDF, OWL, SKOS compliance). FAIRness assessment complements validation by ensuring that SAs themselves are managed as FAIR Digital Objects. Quality and FAIRness aspects of SAs have different focus; while FAIR metadata tells how Semantic Artefacts are managed and exposed, FAIR compliance does not guarantee that they are well designed, logically consistent, or semantically meaningful. Quality refers to the intrinsic characteristics of the Semantic Artefact itself: its correctness, coherence, completeness, and suitability for purpose. With so many SAs appearing in science, the validations and assessment mentioned can provide guidance when selecting the appropriate one for a given purpose. A clear and holistic approach is necessary to integrate these aspects into services offered in the EOSC Federation, and our following recommendations articulate this direction.

1. **The EOSC Federation should publish FAIRness guidelines for Semantic Artefacts** including examples, learning materials and QA services. The EOSC Federation should

select among suitable approaches available, and specify how recommendations[68,69,70] should be applied. The guidelines should detail what is expected from SAs in terms of FAIRness (i.e. identifiers, formats, machine actionability, publishing, licensing and dissemination).

2. **Support FAIRness assessment services in the EOSC Federation**. The EOSC Federation should define expectations and support FAIRness validation. This may be enabled by promoting existing technologies for FAIR assessment (e.g. FOOPS![71] or O'FAIRe[72]), and integrating them into the EOSC Federation's managed services and service catalogue. Assessment results should be preserved and coupled with the SA in a machine-findable way, for example using MOD[73] or the FAIR Testing Interoperability Framework (FAIR-IF) developed in the OSTrails project[74].

3. **Provide a FAIR-by-design development methodology for Semantic Artefacts**[75]. Select, improve and suggest methodologies to maintain or improve the FAIRness of SAs during the whole SA lifecycle.

4. **Support validation of Semantic Artefacts**. The EOSC Federation should plan a validation ecosystem for SAs that includes validator tools, validation constraints, outputs and other components. However, these are yet to be defined. Validation may target the SA itself or the data conformance to the SA. The ecosystem of validation services must be designed to allow the use of the services by humans or by automated workflows. Currently available validators are mostly syntactic (e.g. syntax validators[76], common pitfalls detection for ontologies[77], or SHACL[78] validation for RDF data), but the emergence of AI-based solutions

---

[68] Garijo, Daniel, and María Poveda-Villalón (2020). "Best practices for implementing fair vocabularies and ontologies on the web." Applications and practices in ontology design, extraction, and reasoning. IOS Press, 39-54. https://doi.org/10.3233/SSW200034 - (last accessed 26.03.2026 (Pre-print https://arxiv.org/abs/2003.13084 - (last accessed 26.03.2026))

[69] Cox SJD, Gonzalez-Beltran AN, Magagna B, Marinescu M-C (2021) Ten simple rules for making a vocabulary FAIR. PLoS Comput Biol 17(6). https://doi.org/10.1371/journal.pcbi.1009041 (last accessed 30.04.2026)

[70] Clement Jonquet, Biswanath Dutta, Luiz O. Bonino da Silva Santos, Robert Pergl, Yann Le Franc (2023). Common Minimum Metadata for FAIR Semantic Artefacts. Onto4FAIR 2023, https://hal.science/hal-04106533v2 (last accessed 30.04.2026)

[71] OEG FAIR Ontologies Assessment (FOOPS!) https://w3id.org/foops/ (last accessed 30.04.2026)

[72] O'FAIRe: Ontology FAIRness evaluator https://github.com/agroportal/fairness (last accessed 30.04.2026)

[73] Ontology Specification Draft - https://w3id.org/mod#hasEvaluation (last accessed 26.03.2026)

[74] FAIR Interoperability Framework - OSTrails - https://ostrails.eu/fair-if (last accessed 16.03.2026)

[75] Poveda-Villalon, M., Garijo, D., Gonzalez-Beltran, A., le Franc, Y., Jonquet, C., Aubin, S., Soiland-Reyes, S., Goble, C., & Kechagioglou, X. (2023). M4.2 - Processes & tools to engineer FAIR semantic artefacts. Zenodo. https://doi.org/10.5281/zenodo.10551054 (last accessed 26.03.2026)

[76] W3C RDF Validation Service - https://www.w3.org/RDF/Validator/ (last accessed 16.03.2026)

[77] OOPS! - OntOlogy Pitfall Scanner - https://oops.linkeddata.es/ (last accessed 16.03.2026)

[78] Shapes Constraint Language - https://www.w3.org/TR/shacl/ (last accessed 16.03.2026)

for validation has already started[79], so the extension of validation towards requirements compliance, SA element re-use, etc. is anticipated.

5. **Promote further research on quality assessment and validation of Semantic Artefacts.** While FAIRness concerns only certain aspects of the metadata, quality assessment of SAs focuses on the SA itself from a broader perspective. We expect good quality of SAs can lead to better semantic interoperability, and we consider therefore that this area deserves to be improved. There are a lot of research questions yet to be answered in the SA quality domain, from the foundations to quality metrics and methods[80].
6. **Plan the use of quality metrics in the EOSC Federation for Semantic Artefacts**, based on the W3C Data Quality Vocabulary (DQV)[81]. DQV provides a unified way to attach various kinds of quality assessments to SAs, including among others quality measurements, questions and answers, conformance statements and quality classifications.

Table 5: Recommendations for Validation and FAIRness of Semantic Artefacts

Recommendation	Short Description	Priority
2.5.1 (F)	Publish FAIRness guidelines for Semantic Artefacts in the EOSC Federation	Important
2.5.2 (F)	Support FAIRness assessment services in the EOSC Federation	Important
2.5.3 (F)	Provide a FAIR-by-design development methodology for Semantic Artefacts	Important
2.5.4 (F)	Plan a validation ecosystem for Semantic Artefacts	Important
2.5.5 (A)	Promote further research on quality assessment and validation of Semantic Artefacts	Useful
2.5.6 (F)	Plan the use of quality metrics in EOSC Federation for Semantic Artefacts	Useful

[79] Tufek, N. et al. (2025). Validating Semantic Artifacts with Large Language Models. DOI:10.1007/978-3-031-78952-6_9 (last accessed 16.03.2026)
[80] Wilson RSI, Goonetillake JS, Indika WA, Ginige A (2023). A conceptual model for ontology quality assessment: A systematic review. Semantic Web: – Interoperability, Usability, Applicability. 2023;14(6) DOI:10.3233/SW-233393
[81] W3C working Group - Data Quality Vocabulary =- https://www.w3.org/TR/vocab-dqv/ (last accessed 16.03.2026)

2.6 Semantic interoperability and the governance of the EOSC Federation

Semantic interoperability within the EOSC Federation must be supported by its governance. The governance bodies of the EOSC Federation should define the rules, processes, and responsibilities that enable EOSC Nodes to share metadata, data, services, and other research products, including Semantic Artefacts, while preserving consistency of meaning.

This will enhance the semantic interoperability in the Federation by aligning vocabularies, interoperability framework modeling (modeling APIs and protocols) and EOSC Node descriptions.

1. **EOSC Nodes should ensure that policies to promote and implement semantic interoperability are in place**, and that the data and metadata they provide when joining the EOSC Federation are semantically annotated. For this purpose, EOSC Nodes should use well-established Semantic Artefacts already in use by the scientific community.
2. **The multidisciplinary nature of the EOSC Federation requires the use of different Semantic Artefact Catalogues** to address domain-specific requirements (more information in Section 2.4). SACs must be formally endorsed by the EOSC Federation to guarantee long-term semantic interoperability across the Federation and ensure effective governance.
    a. The process to endorse SACs should be developed in collaboration with domain experts to ensure alignment with community standards and practices.
    b. EOSC Nodes should reuse existing platforms and data models when possible, create new SACs only if needed, and ensure any new SACs integrate smoothly with federated systems like OntoPortal[82] or TIB-OLS[83] and with standards registries such as FAIRsharing[84]
    c. A clear assignment of roles and responsibilities for data management, access control, and metadata curation is necessary. Continuous monitoring and improvement based on user/stakeholder feedback should be considered.
3. **The decisions about semantic interoperability** in the Federation should be discussed in the Node Operations Committee (NOC) and approved by the Node Coordination Committee (NCC). In particular, the EOSC Federation should create a dedicated working group on semantic interoperability that formulates recommendations and best practices for EOSC

[82] OntoPortal: https://ontoportal.org/ (last accessed 29.06.2026)
[83] TIB terminology services: https://terminology.tib.eu/ts (last accessed 29.06.2026)
[84] FAIRSharing: https://fairsharing.org/ (last accessed 04.05.2026)

Nodes. The recommendations and other insights should be adopted by other working groups, such as those focused on catalogs or scientific repositories.

Table 6: Recommendations for Semantic interoperability governance pathways within the Federation

Recommendation	Short Description	Priority
2.6.1 (F)	Establish Semantic Interoperability Governance Body within the Federation	Critical
2.6.2 (F)	Establish a list of Semantic Artefact Catalogues supported/endorsed by the EOSC Federation	Important
2.6.3 (F)	Establish a working group on Semantic Interoperability within the EOSC Federation	Important

3. Prospective Semantic Interoperability Issues

The recommendations in this chapter are intended to support the EOSC Federation in its advance on federated semantic services and AI, two emerging topics in the context of semantic interoperability. Because of the novelty of these issues, there is only a description of recommendations in the following subchapters without assessing the priority level. These recommendations should be taken as initial steps towards future research of these topics..

3.1 Recommendations for federated semantic services

As the EOSC moves toward a federation of nodes, semantic interoperability should evolve from simply publishing and sharing Semantic Artefacts (SAs) in independent and not harmonized Semantic Artefact Catalogues (SACs), to operating semantics at runtime and across different infrastructures and communities. This federation should extend beyond SACs to semantic services as a whole.

Federated semantic services play a critical role in enabling scalable and uniform access to FAIR SAs; they cover the full SA lifecycles, including management (editing, curating, versioning, cataloguing), quality control (evaluating), and exploitation (accessing, discovering, mapping, recommending, or annotating data). Federated semantic services are often, but not necessarily, offered across distributed SACs.

Beyond aggregation, federation allows services to behave as a single logical ecosystem while preserving local governance and autonomy. At the programmatic level, federated SPARQL endpoints allow semantic queries to span multiple repositories, and federated API services provide programmatic access to distributed SAs through a single entry point, facilitating integration into data pipelines, annotation tools, and infrastructure or data spaces.

Federated browsing and search services enable to explore SAs hosted in different SACs through a unified user interface or API, supporting cross-domain discovery that would otherwise require manual navigation of multiple platforms. Federation also enables advanced functionalities such as global SA recommendation services (e.g., suggesting FAIR artefacts across catalogues), annotation services that can use terms independently of their serving infrastructures.

Together, these federated semantic services transform isolated services into an interoperable and uniform infrastructure, enabling cross-community reuse, reducing fragmentation, and supporting large-scale semantics in data-driven research and innovation. Furthermore, there is a need for contracts and shared governance of federated semantic services to keep the services provided by nodes comparable and testable at the federation scale.

Concrete recommendations are:

1. **Adopt a federation-first architecture.** Design semantic services assuming that artefacts are distributed across multiple autonomous SACs and can be involved in multiple services. Federation should be a core architectural principle, not an afterthought layered on top of isolated catalogues.
2. **Preserve local autonomy while enabling global access.** Allow each semantic service to retain its own governance and policies, while exposing standardized interfaces that enable federation at the service at UI and API levels.
3. Federated services (and their operators) must **respect a minimal agreed upon quality of services or SLA** corresponding to the commitment made by the federation and which should guarantee support and engagement.
4. **Expose federated services at multiple interaction levels**. Support federation not only at the user interface level, but also through REST APIs, SPARQL endpoints, and machine-to-machine services to ensure seamless integration into external tools, workflows, infrastructures.
5. **Standardize metadata and APIs across federation members.** Agree on a minimal shared metadata model and common API (such as the MOD-API for SACs) to ensure interoperability, while allowing local extensions when needed.
6. **Provide unified federated search and browsing services**. Implement cross-catalogue indexing and search capabilities that allow users to discover semantic artefacts through a single entry point, with clear provenance and links back to the source SACs.
7. **Implement cross-portal SA recommendation or annotation services**. Leverage federation to recommend FAIR semantic artefacts across catalogues, helping users identify existing resources and avoid unnecessary duplication. Ensure that semantic annotations reference terms hosted in SAs, with URIs independent of the artefact's physical location or SAC.
8. **Make provenance, versioning, and the curation process explicit**. Clearly expose where artefacts originate, how they are curated, and which versions are being accessed, especially in federated search, recommendations and query results.
9. **Implement active semantic brokerage services**. Enable the dynamic execution of mappings (query rewriting/mediation) at query time to ensure consistent discovery across autonomous Nodes.
10. **Facilitate onboarding of new federated semantic services** (provided by Nodes). Provide clear documentation, validation tools, and reference implementations to lower the technical, legal and organizational barriers for new catalogues to join the federation.

Note that in the context of EOSC, some initiatives have already explored the federation of semantic services, including efforts to interconnect multiple semantic artefact catalogues (e.g., OntoPortal Federation[85]) and to provide unified access to distributed terminology and ontology services through shared interfaces (e.g., TIB Terminology Service[86]) or service gateways (e.g., TS4NFDI API Gateway[87])."

3.2 Recommendations about semantic interoperability and AI

The use of AI models has become ubiquitous in recent years, and interoperability is no exception. In the European Union, the legal framework is set up by the European AI Act passed in 2024[88]. We consider here two topics: semantics for AI, and AI for semantics.

Semantics for AI refers to the role that well-structured, interoperable data can play in improving AI systems and open weight models. All the recommendations and best-practices outlined in this document are aimed at improving AI systems by adoption of shared vocabularies, ontologies, persistent identifiers, and rich metadata that provide models with clear and consistent information, and improve their overall quality. All efforts invested in semantic interoperability contributes directly to the robustness, transparency, and reproducibility of AI models.

We recommend a model such as CroissantML[89], which can be highlighted as a concrete effort to facilitate semantics for AI. CroissantML provides a structured way to describe dataset contents, including data resources, features, annotations, and relationships between data elements. Typically expressed using JSON-LD, CroissantML descriptions allow AI systems and data processing tools to interpret dataset features consistently and integrate datasets into machine learning pipelines.

To facilitate the discovery of all resources available, the data model for scientific resources should include AI-friendly resource profiles (esp. through the use of appropriate metadata schemas). Moreover, services should include AI-related resources in a separate category described with the appropriate semantics facilitating their discoverability to users building advanced AI tools[90].

---

[85] Federated FAIR Semantic Artefacts Discovery and Search with OntoPortal Federation https://dx.doi.org/10.1007/978-3-032-09530-5_25 (last accessed 27.03.2026)
[86]NFDI4ING Terminology Service - https://terminology.tib.eu/ts (last accessed 27.03.2026)
[87] TSNFDI API Gateway - https://ts4nfdi.github.io/api-gateway/ (last accessed 27.03.2026)
[88] http://data.europa.eu/eli/reg/2024/1689/oj (last accessed 28.04.2026)
[89] CroissantML - https://dl.acm.org/doi/10.1145/3650203.3663326 (last accessed 13.03.2026)
[90] For example, objects and products metadata could be improved by using search terms entered by users in queries, through empirical investigation. Assuring that the most popular search terms are reflected in metadata content requires setting up a regular, periodic and automated analysis. An up-to-date knowledge graph can also increase the quality of the resource search service for end users.

AI for semantics is a newer development with a lot of potential for innovation that explores how AI techniques can support and partially automate semantic interoperability. AI models can support the resolution of certain interoperability challenges by promoting the adoption of semantic practices. However, their application requires appropriate supervision. It is essential to incorporate a curation process to validate and ensure the reliability of the results.

Some of the challenges include:

- AI assistants for creating mappings between schemes[91]
- Ontology alignment
- AI assisted translation services
- Supporting non-harmonised interoperability at the (meta)data level
- Automated annotation of datasets with metadata and ontology terms
- Automated curation of datasets, including semantic metadata
- AI-supported entity resolution and linking across heterogeneous data sources
- Detection and harmonisation of inconsistencies in metadata records and vocabularies
- Creation, enrichment, and curation of semantic artefacts.
- Identification of gaps in existing schemas and suggest harmonization strategies across disciplines.

The use of AI in this domain must be approached with utmost caution. Automated semantic alignment can propagate hidden biases, hallucinated relations, or subtle errors at scale. Two safeguards are essential:

- systematic cross-validation between multiple models or methods,
- sustained human oversight.

Expert review remains necessary to ensure conceptual accuracy, contextual relevance, and general integrity.

Recommendations for the responsible adoption of AI tools include:

1. **Establish a "human in the loop" curation process for AI outputs** that allows the validation and modification of outputs produced by the AI.
2. **Develop and maintain benchmarks to assess the performance of AI tools over time.** Such benchmarks should support the evaluation of the accuracy and reliability of current and future solutions , enabling consistent comparison and continuous improvement.

[91] Guidelines for the usage of components for technical and semantic interoperability in cross-domain use cases - https://doi.org/10.5281/zenodo.15527557 (last accessed 27.03.2026)

3. **Ensure transparency and traceability of the outputs.** AI-supported mappings, annotations, or translations should be explainable and reproducible. It should be possible to document which tool or model produced a given output and under which conditions.
4. In addition, the EOSC Federation should **define a minimal disclosure profile for AI-assisted semantic outputs/artefacts** (mappings, annotations, translations): provenance of the tool/model and inputs, versioning and PID, validation status (human review), and publication in machine-processable formats (e.g., SSSOM/SHACL/RDF) so results remain reusable and auditable across nodes.

3. Conclusions

This document provides recommendations to ensure and/or enhance semantic interoperability in the EOSC Federation. The recommendations are categorized in six areas – data, metadata, mappings, catalogues, validation and governance, and are addressed to the EOSC Nodes (recommendations labeled with N), the EOSC Federation (F), and the EOSC Association (A). The recommendations are furthermore categorized by importance: "critical recommendations" are crucial to exchange meaning of data in the federation and need to be implemented as soon as possible, "important recommendations" are important to achieve interoperability goals in specific cases, and "useful recommendations" support and enhance interoperability, but do not affect its existence. The document further highlights two prospective issues: federated semantic services and artificial intelligence.

EOSC Node level

Recommendations to EOSC Nodes (there are 17 of them) mostly address the semantic interoperability of the resources offered by the Nodes, and of the Nodes' infrastructure. They concern the semantic interoperability of data (chapter 2.1) and metadata (chapter 2.2), as well as that of the services used by EOSC Nodes to publish information about those resources. Core recommendations to the Nodes are:

- Use DCAT as the baseline schema for describing datasets, catalogues, and data services, reusing existing application profiles.
- Use RDF graphs, IRIs, and RDF schemas or ontologies to make dataset semantics explicit.
- Ensure all mapping specifications, semantic artefacts, and their individual elements are uniquely identified.
- Ensure that all semantic artefacts are discoverable and managed through a SAC, leveraging existing platforms.

EOSC Federation level

The 11 recommendations to the EOSC Federation concern features of semantic interoperability related to the infrastructure needed to interconnect the EOSC Nodes to implement guidelines and Federation-level services. The recommendations refer mostly to validation and governance. Notable recommendations to ensure semantic interoperability at the level of EOSC Federation are:

- Provide Federation-level support for mapping registry services that adhere to standard data schemas and offer APIs.
- Publish FAIRness guidelines for semantic artefacts in EOSC Federation.
- Support FAIRness assessment service(s) in EOSC Federation.

- Establish Semantic Interoperability Governance Body within the Federation.

EOSC Association

There is only 1 recommendation for the EOSC Association, regarding the Validation and FAIRness of Semantic Artefacts, which requires further development by experts, e.g. through the current or new Task Forces or Opportunity Area Expert Groups. The Association should advocate for more funding in European and national funding programmes for research on quality assessment and validation of Semantic Artefacts.

Prospective issues

As the EOSC Federation continues to evolve, federated semantic services and Artificial Intelligence should be considered as key enablers for its future development and sustainability.

Federated semantic services that support the full lifecycle of FAIR artefacts provide unified access through standard interfaces such as SPARQL endpoints and APIs. This creates a cohesive ecosystem preserving local autonomy while enabling cross-domain discovery, recommendations, and annotations. A shared governance with a sustainable perspective remains essential to ensure scalability, consistency, and long-term interoperability across domains.

Regarding the dual role of Artificial Intelligence: Semantics for AI (using interoperable data like CroissantML to build better models) and AI for Semantics (leveraging AI to automate annotations and mappings). To ensure reliability, AI-assisted outputs must remain transparent, traceable, and subject to human-in-the-loop curation.

By aligning technical solutions with sustainable governance and targeted funding, the EOSC Federation can ensure that "what is sent is what is understood". Based on the recommendations, there is a need for targeted funding from the European Commission to develop further the necessary cross-federation services and activities related to the semantic interoperability.

ANNEX:

Semantic bundle (data-level semantics)

A semantic bundle minimally contains:

1. The dataset itself
2. A machine-readable schema describing the data
3. Semantic references to controlled vocabularies or ontologies

For full interoperability in EOSC, the bundle can additionally include:

4. Semantic mappings between vocabularies
5. Provenance information
6. Catalogue metadata for discovery

Example Conceptual Structure of a Semantic Bundle

```
semantic-bundle/
│
├── data/

│   └── dataset.csv
│
├── metadata/
│   └── dataset.csv-metadata.json          (CSVW schema or JSON-LD)
│
├── semantics/
│   ├── variable-definitions.ttl      (I-ADOPT / ontology definitions)
│   └── vocabulary-links.ttl               (SKOS concepts)
│
├── mappings/
│   └── vocabulary-crosswalk.sssom.tsv
│
├── provenance/
│   └── dataset-provenance.ttl        (PROV-O)
│
└── catalogue/
    └── dataset-dcat.jsonld                (DCAT discovery metadata)
```