| | |
|---|---|
| Project Title | Fostering FAIR Data Practices in Europe |
| Project Acronym | FAIRsFAIR |
| Grant Agreement No | 831558 |
| Instrument | H2020-INFRAEOSC-2018-4 |
| Topic | INFRAEOSC-05-2018-2019 Support to the EOSC Governance |
| Start Date of Project | 1st March 2019 |
| Duration of Project | 36 months |
| Project Website | www.fairsfair.eu |

# D2.5 FAIR Semantics Recommendations Second Iteration

| | |
|---|---|
| Work Package | WP2 |
| Lead Author (Org) | Yann Le Franc, Wim Hugo (e-Science Data Factory) |
| Contributing Author(s) (Org) | Gerard Coen (DANS), Jessica Parland von Essen (CSC), Luiz Bonino (LUMC - DTL) |
| Due Date | 31.12.2020 |
| Date | 21.12.2020 |
| Version | 1.0 |
| DOI | https://doi.org/10.5281/zenodo.4314320 |

Dissemination Level

| | |
|---|---|
| X | PU: Public |
| | PP: Restricted to other programme participants (including the Commission) |
| | RE: Restricted to a group specified by the consortium (including the Commission) |
| | CO: Confidential, only for members of the consortium (including the Commission) |

## Abstract

This document is the second iteration of recommendations for making semantic artefacts FAIR. These recommendations result from initial discussions during a brainstorming workshop organised by FAIRsFAIR as a co-located event with the 14th RDA Plenary meeting in Helsinki. Based on the feedback received there, 17 preliminary recommendations related to one or more of the FAIR principles, and 10 best practice recommendations on semantic artefacts were documented. These recommendations were first published as Deliverable 2.2 at the beginning of 2020[1]. Following various dissemination and stakeholder engagement activities, including an evaluation workshop held in October 2020, this second iteration of recommendations and best practices are proposed.

**Deliverable 2.5 is a complete, reviewed and improved version of D2.2.**

**D2.5 supersedes D2.2**

1 https://doi.org/10.5281/zenodo.3707984



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

## Versioning and Contribution History

| Version | Date | Authors | Notes |
|---|---|---|---|
| 0.4 | 23-Nov-2020 | Wim Hugo (e-SDF) | Revision and simple improvements to iteration 1 recommendations and best practice |
| 0.5 | 25-Nov-2020 | Wim Hugo (e-SDF) | Added conceptual models, architecture, best practice, minimum metadata proposals |
| 0.6 | 30-Nov-2020 | Wim Hugo (e-SDF) | Finalisation for project delivery, removed conceptual models, architecture, best practice, minimum metadata proposals to separate document |
| 0.7 | 09-Dec-2020 | Gerard Coen (DANS) | Revision of the document. Input from the RDA VSIG FSRTG added. |
| 0.8 | 14-Dec-2020 | Robert Huber (UniHB)<br>Eva Méndez (UC3M)<br>Mustapha Mokrane (DANS) | Internal review and review by PCO |
| 0.9 | 15-Dec-2020 | Wim Hugo (e-SDF) | Revision and response to internal review |
| 1.0 | 21-Dec-2020 | Wim Hugo (e-SDF) | Final for release |

## Disclaimer

FAIRsFAIR has received funding from the European Commission's Horizon 2020 research and innovation programme under the Grant Agreement no. 831558 The content of this document does not represent the opinion of the European Commission, and the European Commission is not responsible for any use that might be made of such content.



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

## Abbreviations and Acronyms

| | |
|---|---|
| API | Application Programming Interface |
| BFO | Basic Formal Ontology |
| BP | Best Practice |
| CODATA | Committee on Data of the International Science Council |
| DCAT | Data Catalog Vocabulary |
| DOI | Digital Object Identifier |
| DOLCE | Descriptive Ontology for Linguistic and Cognitive Engineering |
| DOOR | Descriptive Ontology of Ontology Relations |
| ELIXIR | ELIXIR the European life-sciences Infrastructure for biological Information |
| EMMO | European Materials Modelling Ontology |
| EOSC | European Open Science Cloud |
| ESFRI | European Strategy Forum on Research Infrastructures |
| FAIR | Findable, Interoperable, Accessible and Reusable |
| FDP | FAIR Data Point |
| FOAF | Friend Of A Friend |
| GUPRI | Globally Unique, Persistent and Resolvable Identifier |
| HTTP | Hypertext Transfer Protocol |
| IOF | Industrial Ontology Foundry |
| IRI | Internationalized Resource Identifier |
| JSON-LD | JavaScript Object Notation for Linked Data |
| KOS | Knowledge Organisation System |
| LD | LD Linked Data |
| LOV | Linked Open Vocabularies |
| MIREOT | Minimum Information to Reference an External OnTology |
| MIRO | Minimum Information for the Reporting of an Ontology |
| MOD | Metadata for Ontology Description and publication |
| NERC | Natural Environment Research Council (of UK) |
| OBO | Open Biological and Biomedical Ontology |
| ODRL | Open Digital Rights Language |
| OMV | Ontology Metadata Vocabulary |
| OWL | Web Ontology Language |
| P-Rec | Preliminary Recommendations |
| PID | Persistent Identifier |
| PURL | Persistent Uniform Resource Locator |
| RDA | Research Data Alliance |



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

| | |
|---|---|
| RDA VSIG | Research Data Alliance Vocabulary Services Interest Group |
| RDF | Resource Description Framework |
| RDFS | Resource Description Framework Schema |
| Rec | Recommendation (See. P-Rec, BP) |
| RIF-CS | Registry Interchange Format - Collections and Services |
| SKOS | Simple Knowledge Organisation System |
| SPARQL | SPARQL Protocol and RDF Query Language |
| TFiR | Turning FAIR into reality |
| UFO | Unified Foundational Ontology |
| URI | Uniform Resource Identifier |
| URL | Uniform Resource Locator |
| W3C | World Wide Web Consortium |
| XML | Extensible Markup Language |

## Definition of Important Terms

| | |
|---|---|
| Controlled vocabulary | A controlled vocabulary is a flat, normalised, restricted list of terms for a specific use or context. Thesauri and taxonomies are types of controlled vocabularies, but not all controlled vocabularies are thesauri or taxonomies. |
| Formal Semantic Artefact | Artefacts for which relations and/ or logic are properly defined and standardised (taxonomies, thesauri, ontologies) |
| Glossary | A glossary is an alphabetical list of terms in a particular domain of knowledge with the definitions for those terms. |
| Ontology | An ontology is a formal version of a thesaurus where relations are described using a formal system such as Description Logic (DL) to mathematically classify individuals of classes and properties |
| Semantic artefact | A semantic artefact is defined in this work as a machine-actionable and -readable formalisation of a conceptualisation, enabling sharing and reuse by humans and machines. These artefacts may have a broad range of formalisation, from loose sets of terms, taxonomies, thesauri to higher-order logics. Moreover, semantic artefacts are serialised using a variety of digital representation formats, e.g., RDF Turtle, and OWL, using XML (RDF) and JSON-LD. |



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

| | |
|---|---|
| Semantic Registry | A semantic registry is a catalogue that contains metadata about semantic artefacts. |
| Semantic Repository | A semantic repository is defined in this recommendation as a service that stores and offers access to both the metadata of semantic artefacts and their content, i.e. offers search and access to get individual terms (including their metadata) both for humans and for machines. |
| Taxonomy | A taxonomy is a controlled vocabulary with a hierarchical structure used to classify things or concepts. Terms within a taxonomy have relations to other terms (parent/broader term, child/narrower term). |
| Term/class/concept | A term/class/concept is an individual element with a unique semantic interpretation, represented with a unique identifier. |
| Thesaurus | A thesaurus is essentially a controlled vocabulary following a standard structure, where all terms have relationships of three kinds to each other: hierarchical (broader term/narrower term), associative, and equivalent (use/used from or see/ seen from). In addition, it is common in thesauri for some or all terms to have additional metadata such as scope notes (brief explanations of how the term should be used in indexing) or history notes. |



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

## Executive Summary

Semantic artefacts (i.e. controlled vocabularies, ontologies, thesauri, and other knowledge organisation systems) are key building blocks for the implementation of the FAIR principles, specifically as emphasized in the Interoperability principle I2 "*(Meta)data use vocabularies that follow FAIR principles*". However, most of these artefacts are actually not FAIR themselves.

The main objective of our work within the Task "FAIR Semantics" of the FAIRsFAIR project is to support the creation of a federated semantic space by harmonising practices in the development and usage of semantics in representing and describing information and knowledge. For this purpose, we are working to establish guidelines for practitioners, repositories, the community, and any related stakeholders. To ground these recommendations in reality, we are collecting recommendations and practical information from practitioners through an open consultation and dedicated workshops, and we are reusing/ referring to existing recommendations built by different communities of practice.

This document summarizes a second iteration of recommendations for making semantic artefacts FAIR. The initial recommendations resulted from discussions during a brainstorming workshop organised by FAIRsFAIR as a co-located event with the 14th RDA Plenary meeting in Helsinki. Based on the feedback received there, 17 preliminary recommendations related to one or more of the FAIR principles and 10 best practice recommendations to improve the global FAIRness of semantic artefacts were documented. These were published as D2.2 at the beginning of 2020.

A follow-up workshop was held in October 2020 to obtain input from experts and practitioners in respect of the recommendations. The results of this workshop are documented in detail in the workshop report. The FAIR Semantics team have also been engaged in active participation in a number of international and pluri-disciplinary initiatives such as RDA, CODATA, and GO FAIR in order to foster grassroots engagement in the recommendation development and ensure that the output delivered is aligned to the needs of its stakeholders. Under the RDA Vocabulary Services Interest Group a 'FAIR Semantic Repository Task Group' has been set up which works on evaluating the recommendations from the perspective of semantic repositories and service providers. This Task Group provides both, expert feedback to the FAIR Semantics team and simultaneously works across institutions on common ways to implement the recommendations. Based on the feedback received during the October workshop, and that provided by other stakeholder groups, this second iteration of recommendations and best practices are proposed.



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

# Table of Contents


Executive Summary 7

Table of Contents 8

1. Introduction and Scope 10

1.1. Defining Semantic Artefacts 12

1.2. What is meant by FAIR Semantics? 17

1.3. Stakeholder Groups 18

2. Recommendations 19

P-Rec. 1: Globally Unique, Persistent and Resolvable Identifiers must be used for Semantic Artefacts, their content (terms/ concepts/ classes and relations), and their versions. 21

P-Rec. 2: Globally Unique, Persistent, and Resolvable Identifiers must be used for Semantic Artefact Metadata Records. Metadata and data must be published separately, even if it is managed jointly. 23

P-Rec. 3: A common minimum metadata schema must be used to describe semantic artefacts and their content 24

P-Rec. 4: Semantic Artefact and its content should be published in a trustworthy semantic repository 26

P-Rec. 5: Semantic repositories must offer access to Semantic Artefacts and their content using community standard APIs and serializations to support both use/ reuse and indexation by search engines 28

P-Rec. 6: Build semantic artefact search engines that operate across different semantic repositories 30

P-Rec. 7: Repositories must offer a secure access protocol, and appropriate user access control functionalities 31

P-Rec. 8: Human and machine-readable persistence policies for semantic artefacts metadata and data must be published 32

P-Rec. 9: Semantic artefacts must be made available as a minimum portfolio of common serialization formats 33

P-Rec. 10: Foundational Ontologies may be used to align semantic artefacts 34

P-Rec. 11: A standardized knowledge representation language should be used for describing semantic artefacts 35

P-Rec. 12: Semantic mappings between the different elements of semantic artefacts should be published in machine-readable formats 36

P-Rec. 13: Crosswalks, mappings and bridging between semantic artefacts should be documented, published and curated 38

P-Rec. 14: Standard vocabularies should be used to describe semantic artefacts 39

P-Rec. 15: Provenance information regarding the reuse of components from third-party semantic artefacts should be made explicit 41

P-Rec. 16: The semantic artefact must be clearly licenced for use by machines and humans 42

P-Rec. 17: Provenance must be clear for both humans and machines 43




FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

3. Recommendations Beyond the FAIR Principles: Best Practices for Semantic Artefacts 44

BP. 1: Use a unique naming convention for concept/class and relations 44

BP. 2: Use an Ontology Naming Convention 45

BP. 3: Use defined ontology design patterns 45

BP. 4: Allow mappings to reflect validation by domain experts 46

BP. 5: Define crosswalks between different formats 46

BP. 6: Harmonize the methodologies used to develop semantic artefacts 47

BP. 7: Interact with the designated community and manage user-centric development 47

BP. 8: Provide a structured definition for each concept 48

BP. 9: The underlying logic of semantic artefacts should be grounded in the domain it intends to be used in. 49

BP. 10: Define a set of governance policies for the semantic artefacts 50

BP. 11: Use TRUSTed and FAIR compliant repositories to persist Semantic Artefacts 50

BP. 12: Semantic Artefacts, if developed with public funding, should be published with open licenses unless one of the overriding conditions discussed below are true 51

BP. 13: A standard architecture for semantic artefact management, services for content and metadata, protocols, and serialisations/ content negotiation should be proposed, and semantic repositories are urged to use this architecture as a design pattern 52

BP. 14: User registration and authentication should be as simple as possible and adhere to legislative requirements 52

4. FAIR Principles Coverage 53

5. Final Remarks 54

6. References 56

7. Appendix A: Initial list of recommendations compiled from internal and from workshop discussions 59

8. Appendix B: Stakeholder views 62

8.1. Practitioner view 62

8.2. Repository view 63

8.3. Community view 64



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

# 1. Introduction and Scope

The FAIR principles have evolved into a set of technology-agnostic guidelines to make digital assets Findable, Accessible, Interoperable, and Reusable, defined originally by Wilkinson et al. (2016) in the context of research data. Semantic artefacts (i.e., controlled vocabularies, thesauri, ontologies, etc.) are machine readable models of knowledge organisation systems. They are used to facilitate the extraction and representation of knowledge within datasets using annotations or assertions. These annotations and assertions enable discovery, interoperability, integration and data retrieval. Both, artefacts and services that support and offer them, play an increasing role in the implementation of the FAIR principles (particularly principle 2 of Interoperability (I2): (Meta)data - use vocabularies that follow the FAIR principles) and in building FAIR Scientific Knowledge Graphs, in order to express and link scientific contributions and related artifact in a semantically rich FAIR graphical model.. This role has been acknowledged by the European Commission Expert Group on FAIR Data in Recommendation 7 within their final report and action plan "Turning FAIR into reality" (European Commission Expert Group on FAIR Data, 2018):

"Support semantic technologies - Semantic technologies are essential for the interoperability and need to be developed, expanded and applied both within and across disciplines" is the seventh recommendation in the seminal *Turning FAIR into reality* report (European Commission Expert Group on FAIR Data, 2018, p. 42) According to the expert group, semantic artefacts and registries have been developed within almost all scientific disciplines). However, they have often been built using different formats (SKOS, XML, RDF, OWL), different levels of complexity/ expressiveness (codelists, reference data, controlled vocabularies, taxonomies, thesauri, ontologies, formal ontologies, etc.) and are scattered on the web. Indeed, in many cases, semantic artefacts are not interoperable and not easy to find and are therefore accessible only to the community of practice within which they were developed, which clearly hampers their reuse (Goldfarb and Le Franc, 2017). The emergence of semantic registries such as BARTOC[2], FAIRsharing[3] and repositories and other vocabulary services, such as Bioportal[4] (Whetzel et al., 2011), EBI-OLS[5] (Jupp and al., 2015), Ontobee[6] (Ong et al., 2017), Research Vocabularies Australia[7], the NERC Vocabulary Service[8], Linked Open Vocabulary[9] and others provide means to improve discoverability and enable reusability. The importance of such semantic registries/repositories and the issue of the findability of semantic artefacts is already being worked

2 BARTOC https://bartoc.org/

3 FAIRsharing https://fairsharing.org/

4 Bioportal https://bioportal.bioontology.org/

5 Ontology Lookup Service https://www.ebi.ac.uk/ols/index

6 Ontobee: http://www.ontobee.org/

7 Research Vocabularies Australia https://ardc.edu.au/services/research-vocabularies-australia/

8 NERC Vocabulary Server https://www.bodc.ac.uk/resources/products/web_services/vocab/

9 LOV https://lov.linkeddata.es/dataset/lov/



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

on by various groups such as the DCMI/NKOS Interest Group[10] and d'Aquin & Noy (d'Aquin and Noy, 2012). Despite these existing changes, a large number of semantic resources (i.e. artefacts and repositories) do not comply with most of the FAIR principles.

Semantic Web technologies and standards were built to connect and add meaningfulness to data silos and create a web of data next to a web of documents as the current World Wide Web. Unfortunately, in the past decades, the isolated development of semantic artefacts and the lack of common practices to foster interoperability and reusability of semantic artefacts lead to the creation of semantic silos. There is therefore a clear need for a harmonized framework to build, share, publish and reuse semantic artefacts which will provide a harmonised semantic landscape easing reuse and integration for practitioners.

The main goal of **task 2.2 "FAIR Semantics"** is to build such harmonized framework by proposing a set of recommendations and good practices that enable domain specific specialist and data professionals to design FAIR semantic artefacts from the start and therefore de facto supporting the usage of semantics in the FAIRification of data, cross-disciplinary semantic interoperability and the creation of FAIR Scientific Knowledge Graphs. The current situation is characterized by a lack of communication and cross fertilization between the semantic web and knowledge (ontology) engineering practitioners across various domains of application. Our goal is to develop general recommendations that could be applied by all domains of knowledge to create FAIR semantic artefacts from the start. Our approach is to consider that such generic recommendations should be designed considering input at the grassroots level and with the support of as many experts as possible. This will foster validation through a large diversity of use-cases.

Our approach relies on establishing a platform for discussion and collaboration between all stakeholders, to propose a common approach to define recommendations for FAIR Semantics and to promote existing domain-specific efforts, such as OBO foundry[11] for the biomedical domain or the recently created Industry Ontology Foundry[12]. For this purpose, we are organising dedicated workshops to gather a large audience. These recommendations were proposed by experts during our first brainstorming session organised as a workshop co-located with RDA Plenary 14 (2019) in Helsinki, and then over the course of 2020 they have been refined and adapted based on stakeholder feedback which culminated in a half-day evaluation workshop in October 2020. Following the publication of this report the new release of FAIR Semantics recommendations will once again be disseminated to the communities to gather feedback.

10 DCMI/NKOS Interest Group https://dublincore.org/groups/nkos/

11 OBO Foundry http://www.obofoundry.org/

12 Industrial Ontology Foundry https://www.industrialontologies.org



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

## 1.1. Defining Semantic Artefacts

Initially, we were considering using the common term “ontology” to encompass the different types of semantic models. However, during our discussions both within the project and with our colleagues, we realized that the term “ontology” had different meanings for different communities of practice. This ambiguity of the concept “ontology” has been discussed largely in the scientific literature, for example by Guarino et al. (Guarino et al., 2009) and is still debated (see Neuhaus, 2017).

The original definition has been given by Gruber in 1993: “An ontology is an explicit specification of a conceptualization” (Gruber, 1993). In this context, *“A conceptualization is an abstract, simplified view of the world that we wish to represent for some purpose.”* Based on these two key definitions, we can consider ontologies as semantic models of a part of the real world.

Due to the problem of ambiguities with the use of the term “ontology”, we decide to distance ourselves from this debate by proposing and using a more generic umbrella term: **Semantic Artefact**. Semantic Artefact is defined here as a machine-actionable and -readable formalisation of a conceptualisation enabling sharing and reuse by humans and machines. These artefacts may have a broad range of formalisation, from loose set of terms, taxonomies, thesauri to higher-order logics, and include the concepts/terms/classes constituting these. . Moreover, semantic artefacts are serialised using a variety of digital representation formats, e.g., RDF Turtle, OWL-RDF, XML, JSON-LD. In current practices, these artefacts share a common structure encapsulating its metadata, the data i.e. the semantic artefact content comprising of concepts/terms/classes and relations among them, and their (artefact's content) associated metadata (see fig. 1).



FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

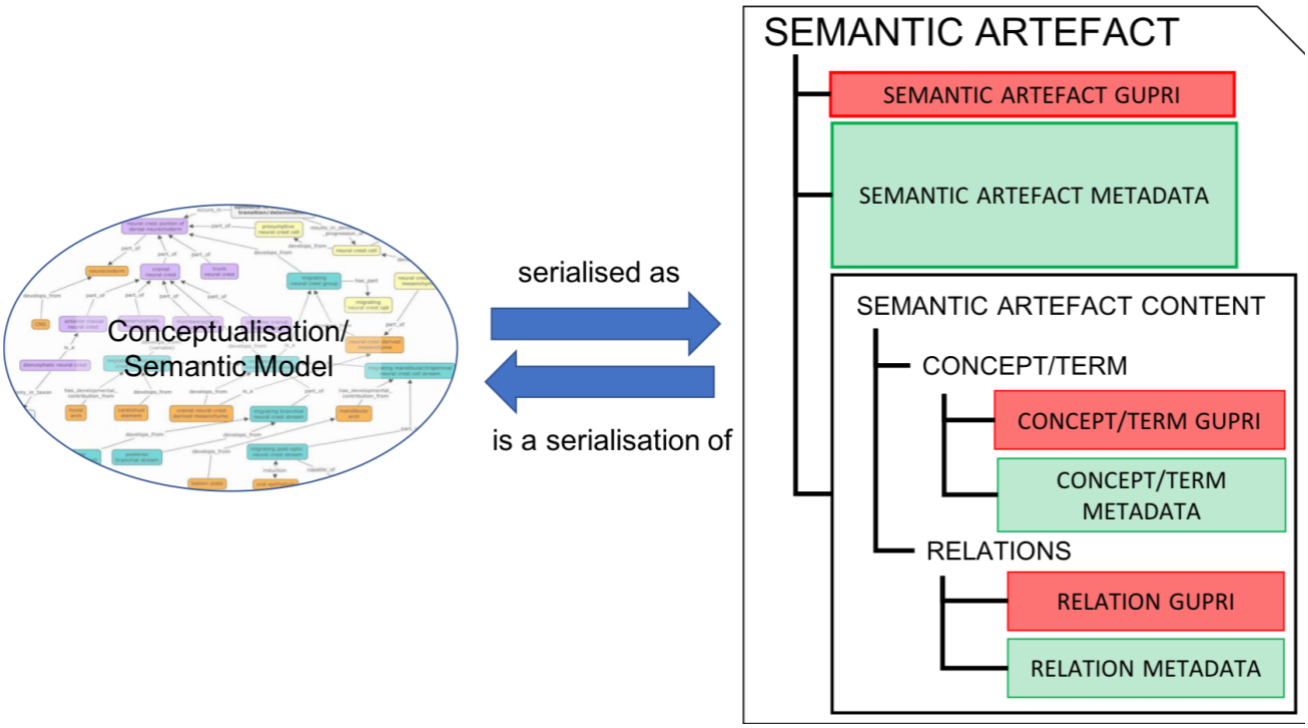


*Figure 1: Common structure of semantic artefacts*

*Semantic artefacts are often structured text files. They have a common structure encapsulating a GUPRI (Globally Unique, Persistent and Resolvable Identifier) for the semantic artefact, the metadata describing the semantic artefact and the semantic artefact content i.e., concept/term and relations. Both are also encapsulating a GUPRI and associated metadata.*

As we mentioned previously, Semantic Artefacts are created in different formats and at different levels of complexity/expressiveness. To classify Semantic Artefacts based on their complexity, an "ontology spectrum" has been proposed and can be used to identify the different types of semantic artefacts and associate them with common formats used to serialise these models (Obrst, 2003, 2010). In fig. 2, we are presenting a simplified version of the spectrum which represents the different types of Semantic Artefacts along a semantic strength axis (ranging from weak semantics to strong semantics). The classification starts with simple lists of terms/concepts (code list, glossary, catalogue ID, controlled vocabulary). These lists of terms/concepts are the simplest building block of semantics and provide a minimal set of information for each item such as a definition, context information and provenance information without relations of any kind.



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

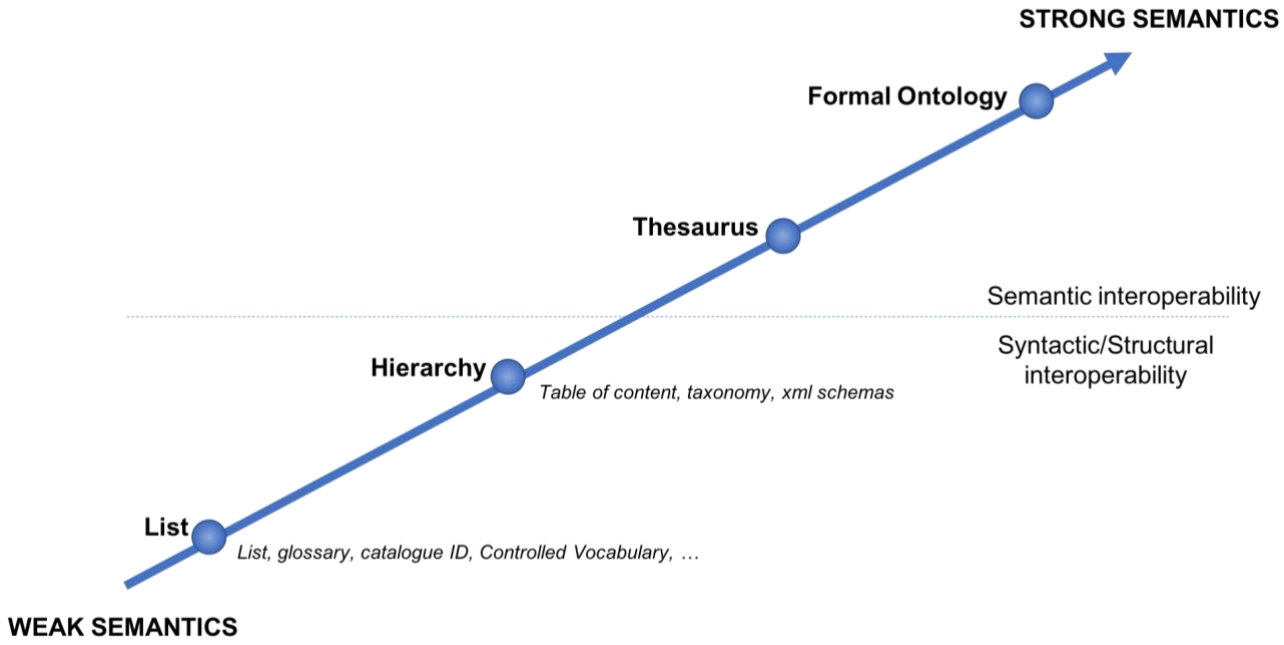


*Figure 2: Semantic artefact spectrum. Derived from Leo Obrst, 2010*

*Semantic artefacts are classified into 4 main types: list, hierarchy, thesaurus and formal ontology. These 4 different types of semantic artefacts are represented along an axis going from "weak semantics" to "strong semantics". Examples of subtypes are provided on the right side of the axis. A dichotomy can be made between hierarchy and thesaurus. On the one side the simplest types are supporting syntactic interoperability allowing machines to process information due to compatible syntax. On the other side, semantic interoperability is being achieved allowing machines to interpret and reason over the data.*

The second block corresponding to hierarchical models (informal hierarchies and taxonomies) builds upon a list of terms/concepts organised hierarchically using either "loose" parent/child or the more formal "*is a*" relations. These hierarchies can then be enriched with additional relations such as synonyms and association relations therefore becoming thesauri. Thesauri can be used as a basis to create formal ontologies by adding axioms and rules. This type of "Russian doll" like organisation is shown in fig. 3. It allows us to visualize a path of transformation between semantic artefact types.



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

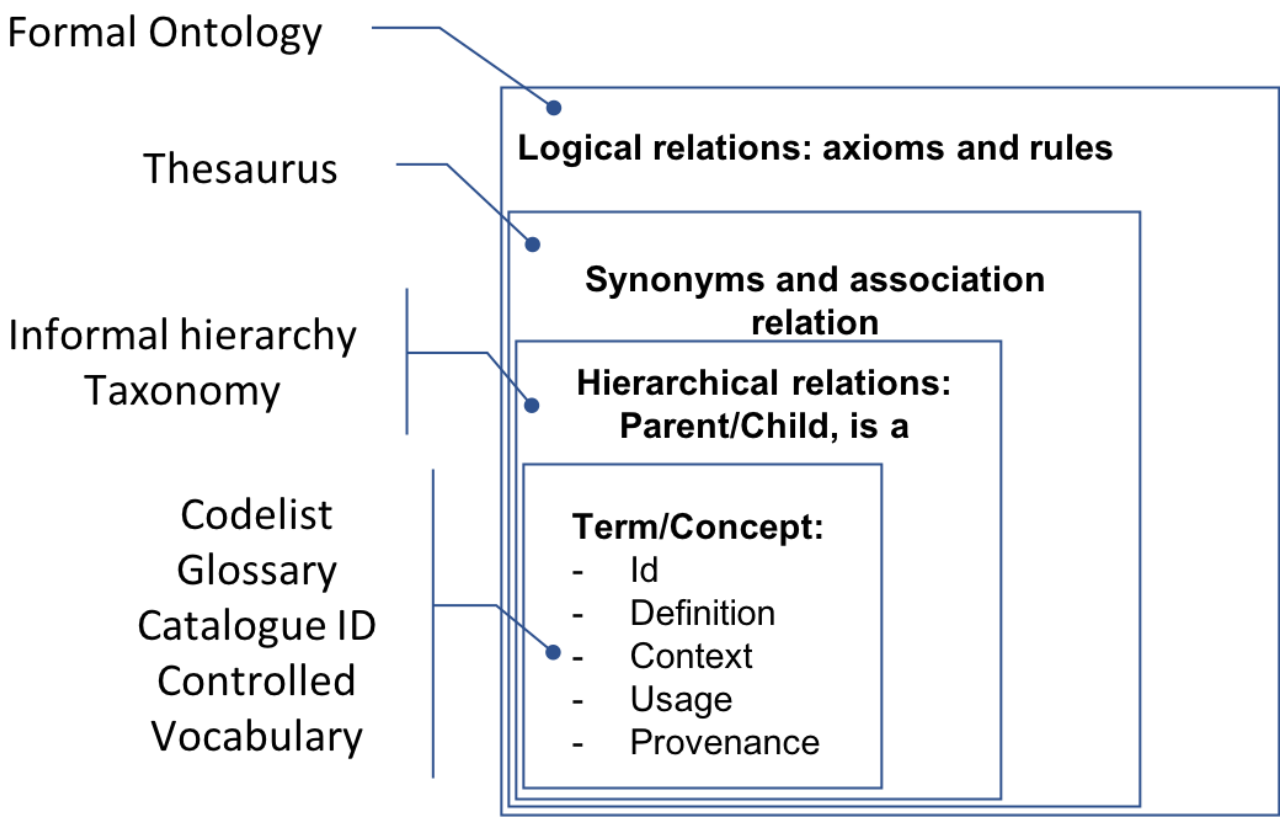


*Figure 3: From list to formal ontology: a transformation path.*

In addition to the complexity, semantic artefacts are also heterogeneous in nature due to the diversity of the data models and standards used to serialise semantic models. Various standards for data models (RDF[13], RDFs[14], OWL[15], SKOS[16]...) and serialization formats (XML[17], XML Schema[18], JSON, RDF/XML[19], OWL/XML[20], Manchester Syntax[21], JSON-LD[22], Turtle[23], N-Triples[24]…) have been proposed by W3C. For simple models, common formats such as XML, XML Schema and JSON are typically used.

13 RDF https://www.w3.org/TR/rdf11-concepts/

14 RDFs https://www.w3.org/TR/2014/REC-rdf-schema-20140225/

15 OWL https://www.w3.org/OWL/

16 SKOS https://www.w3.org/TR/skos-primer/

17 XML https://www.w3.org/TR/xml/

18 XML Schema https://www.w3.org/TR/xmlschema-0/

19 RDF/XML https://www.w3.org/TR/rdf-syntax-grammar/

20 OWL/XML https://www.w3.org/TR/2012/REC-owl2-xml-serialization-20121211/

21 OWL Manchester Syntax https://www.w3.org/TR/2012/NOTE-owl2-manchester-syntax-20121211/

22 JSON-LD https://www.w3.org/TR/json-ld11/

23 Turtle https://www.w3.org/TR/turtle/

24 N-Triples https://www.w3.org/TR/2014/REC-n-triples-20140225/



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

As the model becomes more complex, more expressive data models have been proposed. The Resource Description Framework RDF is one of them. It is a formal language for describing information as a very simple graph-oriented data schema. Based on a URI to identify the resources, RDF enables the exchange of data on the Web between applications while preserving their original meaning and facilitating the processing and re-combination of the contained information. There are many serialisations of RDF such as RDF/XML, Turtle, JSON-LD, N3, etc. RDF is complemented by RDF Schema denoted as RDFs. This extra layer provides additional data-modelling elements for RDF data thus extending the expressivity of the supported models.

To support a higher level of expressivity and logic to represent complex semantic models, W3C proposed the Web Ontology Language (OWL). This language extends the couple RDF/RDFs with additional reasoning options grounded in formal logic. OWL exists in various flavours and expressivity profiles (e.g., OWL-Lite, OWL-full OWL-DL). Despite these powerful additions enabling reasoning and automated processing, OWL suffers a limitation due to the initial working hypothesis used to formalise the logic. Indeed, based on the Open World Assumption, OWL cannot represent closed logic.

Finally, another standard used to build semantic artefacts is the Simple Knowledge Organisation System (SKOS)[25]. This standard, less formal and constrained that RDF and OWL is also a W3C recommendation to build knowledge organisation systems (KOS) (i.e., semantic artifacts). SKOS is quite popular for building thesauri, classification schemes, subject heading systems and taxonomies (as shown in Table 1). The main reason for such popularity lies in the fact that it has no formal grounding and people use it to express all kinds of containment relations. For example the skos: broaderProperty is used to express a subclass relation (mammal skos:broader animal), a subregion (Texas skos:broader USA), subperiod (baby-boom-period skos:broader 20thCentury) etc. Despite the lack of formal grounding, most humans do understand the inherent reasoning and can develop in retrospect applications that properly deal with these mappings. SKOS provides a standard way to represent knowledge organisation systems using RDF, allowing them to be passed between computer applications in an interoperable way and to be used in distributed, decentralised metadata applications, where metadata is harvested from multiple sources.

In table 1 below, we are listing the common formats/standards used to build each of the 4 types of semantic artefacts.

| Type of Semantic artefact | Currently used standards (serialisation formats and data models) |
|---|---|
| List (terminologies, glossaries, vocabularies) | CSV, XML, JSON, SKOS |

25 SKOS, Simple Knowledge Organisation System, https://www.w3.org/2004/02/skos/



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

| | |
|---|---|
| Hierarchical list | XML-schema, RDF-schema, SKOS |
| Thesaurus | RDF/RDFs, SKOS |
| Formal ontology | OWL, OntoUML, FOL, Modal logic |

*Table 1: Association between the type of semantic artefact and standards used.*

## 1.2. What is meant by FAIR Semantics?

FAIR Semantics, in the context of this project, means that semantic artefacts should adhere to the FAIR principles. For this, we are considering semantic artefacts as a specific type of data, used to describe or annotate other data, i.e., as metadata. This also reflects a transition from the common understanding of metadata based on controlled vocabularies, tags, and labels, towards 'next generation metadata' expressed as Linked Data or Permanent Identifiers (Smith-Yoshimura, 2020). This approach allows us to consider each individual FAIR principle in the context of semantic artefacts. This implies the following:

- usage of globally unique persistent and resolvable identifiers for semantic artefacts, their content (i.e., concept/term/class and relation) and their version,
- machine-readable metadata to describe the semantic artefacts themselves and their content,
- usage of repositories to share, publish and retrieve semantic artefacts and their content
- defining common API(s) to access and index semantic artefacts and their content,
- interoperability approaches to make sure that semantic artefacts of various degrees of complexity and encoding format should work together including publishing mappings and crosswalks between semantic artefacts,
- semantic artefacts and their content should be retrievable through search engines.

Solutions to address part or all these issues have been developed within domain specific communities. As our goal is not to reinvent but rather reuse, we are providing with our recommendation pointers to existing community-specific recommendations. As of now, we have included the following recommendations:

- OBO Foundry (Smith et al., 2007)
- Industry Ontology Foundry (Kulvatunyou et al., 2018)



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

- Agrisemantic Working Group recommendations: 39 Hints to Facilitate the Use of Semantics for Data on Agriculture and Nutrition[26].
- Metadata for Ontology Description and Publication Ontology [MOD] (Dutta et al., 2017)
- Ontology Metadata Vocabulary [OMV] (Hartmann et al., 2005)
- Minimum Information for Reporting an Ontology [MIRO] (Matentzoglu et al., 2018)
- Minimum Information to Reference an External Ontology Term [MIREOT] (Courtot et al., 2009)
- Linked Open Vocabulary [LOV] (Vandenbussche et al., 2017)
- Best practices for implementing FAIR vocabularies and ontologies on the Web. (Garijo & Poveda-Villalón, 2020)

## 1.3. Stakeholder Groups

The goal of this deliverable and the FAIRsFAIR FAIR semantics task is to co-create both recommendations for making semantic artefacts FAIR and a set of agreed best practices to follow together with the community of semantics at large. While working on this initial set of recommendations, we realised that recommendations are often targeted to particular stakeholders. Some recommendations are very specific about the format, structure and content of semantic artefacts, therefore useful for practitioners. Some others are also directed towards developers and maintainers of semantic repositories, while some recommendations actually highlight the need for a community wide consensus to fill gaps in the current landscape of standards and data models for semantic interoperability.

Therefore, for this phase of the work, we are considering three main stakeholder groups:

1. Expert vocabularies managers, practitioners dealing with the creation and maintenance of the semantic artifacts
2. Repositories managers, i.e. development team and curators of community specific semantic repositories;
3. Semantic Web Community at large, dealing with semantic artifacts in general, in different contexts, including research data infrastructures, etc. This includes expert end-users, as well as systems and systems developers incorporating semantic artefacts into their processes.

26 Agrisemantic Working Group https://agrisemantics.org/



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

For each of the recommendations, we listed the impacted/ concerned stakeholder groups and we are providing summary tables for each of stakeholders, listing the recommendations of interest.

> These recommendations remain preliminary. Community experts are invited to contribute to extend our view of community practices by adding any missing recommendation. Please do so by adding requests and suggestions in GitHub:
>
> https://github.com/FAIRsFAIR/FAIRSemantics

## 2. Recommendations

In this section, we are providing a list of individual recommendations. The recommendations were originally derived from the material we gathered during the workshop that took place as a co-located event to the RDA 14 Plenary in Espoo on 23 October 2019. In the workshop more than 20 experts brainstormed and discussed the different criteria of FAIR, elaborating on the implications of these requirements on semantic artefacts. From this material the FAIRsFAIR task group formulated more than 40 recommendations/requests/requirements, which were then analyzed individually. We evaluated how such input relates to one or more particular FAIR principles and aggregated them whenever possible into one recommendation. The result was published as D2.2: 1st set of Recommendations for FAIR Semantics[27]. Feedback has then been collected in discussions in different expert fora, such as the RDA Plenaries 15 and 16, as well as through the fairsfair.eu web page for community review, the RDA VSIG 'FAIR Semantic Repositories Task Group' and on GitHub. The most extensive feedback was collected in a workshop on October 15th 2020. The feedback is presented in the report published separately. (Hugo et al., 2020).

Based on this feedback, the 17 preliminary recommendations presented in the previous iteration have been reviewed, completed and updated, and some of the text of the previous iteration (D2.2) have been improved. The recommendations are each enriched with a description providing some context, and whenever possible, existing recommendations. In addition, we have been considering the FAIR principle(s) addressed by recommendations, and we also consider which stakeholders are impacted/ responsible for such recommendations.

The recommendations that could not be directly aligned to the FAIR principles were aggregated into a set of suggested best practices presented in the next section. These Best Practices are not directly linked to a particular recommendation but contribute to improve the overall FAIRness.

27 https://doi.org/10.5281/zenodo.3707985



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

There are several suggested mergers and splits of recommendations and best practices that were made by the community and the internal reviewers. These have been documented, and will require a restructuring of the recommendations and best practices, to be reviewed by the community in the first half of 2021.

The recommendations are presented in a standard schema, as follows:

| Imperative (Mandatory, Optional, Recommended) | Recommendation Text | FAIR Principles Supported | Broad Topic |
|---|---|---|---|

| **Description** |
|---|
| Short Description of the Recommendation |
| **Related Recommendations** |
| If Any |
| **Stakeholders** |
| Description, with typical use cases if these have been specified[28] |
| **Examples** |
| If any[28] |

28 For some recommendations, this is still work in progress



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

| | | | |
|---|---|---|---|
| Mandatory | P-Rec. 1: Globally Unique, Persistent and Resolvable Identifiers must be used for Semantic Artefacts, their content (terms/ concepts/ classes and relations), and their versions | F1 | Identifiers |

**Description**

Semantic artefacts are typically structured text files. They are *de facto* digital objects as well as creative works, and should be unambiguously identified by globally unique, persistent and resolvable identifiers (GUPRI). In the context of a web of FAIR data, these identifiers should be resolvable and support the retrieval of both the semantic artefact itself and also its metadata (see Rec. 2 regarding metadata).

Semantic artefacts are composite digital objects[29] requiring at least three levels of identifiers:

- one for the semantic artefact itself,
- one for its content
- and one for the metadata (including both the global metadata and the metadata associated with the content). The latter is described in the following recommendation (Rec. 2).

Semantic artefacts are living digital objects by nature, evolving over time. Each version of a semantic artefacts must be uniquely identifiable, allowing access to the latest version by default, but also providing access to previous versions in use in existing information systems.

Semantic artefacts can be considered as collections of terms/ concepts/ classes. Each term/ concept/ class should be uniquely identifiable.

This recommendation emphasizes the need for reliable and persistent identification systems without any technical constraints.

**Related Recommendations**

- W3C Data on the Web - Best Practice 9: Use persistent URIs as identifiers of datasets namespace[30]
- OBO Foundry - Principle 3[31]
- OBO Foundry - Identifier Policy[32]
- OBO Foundry - Principle 4[33]
- Industrial Ontology Foundry - principle 11 IRI and identifier space
- Industrial Ontology Foundry - principle 12 Identifier and naming conventions
- EOSC PID policy recommendation (Hellström et al., 2019)

29 As shown in Fig. 1

30 W3C Data on the Web - Best Practice 9 https://www.w3.org/TR/dwbp/#DataIdentifiers

31 OBO Foundry principle 3 http://www.obofoundry.org/principles/fp-003-uris.html

32 OBO Foundry ID policy http://www.obofoundry.org/id-policy

33 OBO Foundry principle 4 http://www.obofoundry.org/principles/fp-004-versioning.html



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

| **Stakeholders** |
| --- |
| Practitioners (Creators of semantic artefacts - referencing external content)<br>Repository (Ensuring compliance for metadata and content services) |
| **Examples** |
| Both OBO Foundry and Industry Ontology Foundry are proposing to use special conventions to define URI based identifiers (see BP. 1 and BP. 2).<br><br>A unified identifier schema must be used to identify each version of semantic artefact. This can be done using versioned URI as proposed by OBO Foundry. Using GUPRI for the different version allows information systems to retrieve automatically the latest version and older versions of the semantic artefact. |



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

| | | | |
|---|---|---|---|
| Mandatory | P-Rec. 2: Globally Unique, Persistent, and Resolvable Identifiers must be used for Semantic Artefact Metadata Records. Metadata and data must be published separately, even if it is managed jointly | F1, F3 | Identifiers<br>Metadata |

**Description**

Semantic artefacts are often built as containers including both their descriptive metadata and data. Commonly, semantic artefacts contain a set of concepts and their descriptions and are identified by a URL pointing to a file to download which can be parsed to access the content, including the metadata. If the metadata contained in these files are not standardised and machine-readable, such practices contribute to the lack of findability. For this purpose, it is necessary to publish the ontology metadata separately, allowing potential users to find it. This mirrors common practice in respect of other FAIR research outputs - data and scholarly publication.

This metadata record must have a GUPRI (Globally Unique, Persistent and Resolvable Identifier - FAIR principle F1), a human readable landing page, and an explicit reference to the semantic artefact it describes (FAIR principle F3). Alternatively, it must be available at the same GUPRI via content negotiation. In this way, search engines can retrieve and index metadata that uniquely point to their related semantic artefacts. This recommendation puts an emphasis on the necessity to publish metadata separately from the semantic artefact and have services to share/ publish ontologies which should support the extraction and the publication of their metadata as suggested in P-Rec. 4.

**Stakeholders**

Practitioner (indicate and reference metadata elements if managed in the same file as content)
Repository (ensuring that metadata can be found as a separate GUPRI or GUPRI with content negotiation)



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

| | | | |
|---|---|---|---|
| Mandatory | P-Rec. 3: A common minimum metadata schema must be used to describe semantic artefacts and their content | F2, R1.1, R1.2 and R1.3 | Metadata |

**Description**

As with any type of data, semantic artefacts should be described by metadata[34] to allow users to cite them, retrieve them, and to understand and apply their content. In addition, it is important to have general information regarding the scope of the semantic artefact (at least which domain is covered by the ontology), provenance information, and many other details. Metadata must be appropriate to the life cycle stage and application of the artefact.

This metadata must be available in popular encodings, and must be accessible for harvesting and discovery by search engines, semantic service providers and metadata aggregators, registries, and catalogues.

As for semantic artefact themselves, the concept /term/ class and relation that compose them should also have a common metadata schema that provide information such as label, definition, examples of usage, author, version, multilingual labels, and similar.

Reaching an agreement at this level will ease the process of working with concepts from multiple heterogeneous semantic artefacts. It is important to note that proper definitions are necessary to be able to evaluate the difference between similar classes from different ontologies (see BP. 8).

**Existing Recommendations**

- Force 11 Citation Recommendations[35]
- OBO Foundry - Principle 8 Documentation[36]
- OBO Foundry - Principle 5 Scope[37]
- OBO Foundry - Principle 6 Textual definition[38]
- Industry Ontology Foundry - Requirement 9 Documentation[39]
- Industry Ontology Foundry - Requirement 5 Scope[40]
- LOV - DCAT based metadata schema
- VOAF[41]

34 There is currently no consensus on a common set of metadata elements to describe semantic artefacts. See existing recommendations, and refer to the examples for inputs to be considered when defining a minimum metadata schema.

35 https://www.force11.org/datacitationprinciples

36 OBO Foundry principle 8 http://www.obofoundry.org/principles/fp-008-documented.html

37 OBO Foundry principle 5 http://www.obofoundry.org/principles/fp-005-delineated-content.html

38 OBO Foundry principle 6 http://www.obofoundry.org/principles/fp-006-textual-definitions.html

39 IOF Technical Principles https://www.industrialontologies.org/?page_id=87

40 IOF Technical Principles https://www.industrialontologies.org/?page_id=87

41 VOAF https://lov.linkeddata.es/vocommons/voaf/



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

- Ontology Metadata Vocabulary[42]
- Metadata for Ontology Description and Publication Ontology[43]
- W3C Data on the web best practices – BP. 1, BP. 2 and BP. 3[44]
- Networked Knowledge Organization Systems Dublin Core Application Profile (NKOS AP)[45]

**Stakeholders**

Practitioner (ensure that minimum metadata is available for each semantic artefact created or maintained)
Repository (perform quality assurance in respect of minimum metadata)
Community (participate in and contribute to forums for definition and maintenance of minimum metadata)

**Examples**

Several initiatives are proposing their recommendations such as OBO Foundry and IOF. Several metadata schemata have been developed, such as LOV (Vandenbussche et al., 2017), Ontology Metadata Vocabulary (OMV)[46], Metadata for Ontology Description and Publication Ontology (MOD)... (see list of related recommendations below). However, the heterogeneity of these metadata schema hampers indexing, and retrieval, as well as reuse of the semantic artefacts.

The initial workshop also produced a set of minimum metadata elements[47].

For other research outputs, there are ample examples to consider as inputs, for example DataCite.

42 Ontology Metadata Vocabulary http://omv2.sourceforge.net/

43 MOD-Ontology https://github.com/sifrproject/MOD-Ontology

44 Data on the Web Best Practices https://www.w3.org/TR/dwbp/

45 NKOS AP https://nkos.slis.kent.edu/nkos-ap.html

46 OMV http://mayor2.dia.fi.upm.es/oeg-upm/index.php/en/downloads/75-omv/index.html

47 https://docs.google.com/document/d/1ot2K_u1xzLQz1FUuZfxpP6h0um3elJPR0jxnAktsbpc/edit



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

| Optional | P-Rec. 4: Semantic Artefact and its content should be published in a trustworthy semantic repository | F4 | Repository |
|---|---|---|---|

**Description**

Semantic artefacts are made accessible using a wide variety of mechanisms, including publication in open repositories such as Zenodo or Figshare, deployment to GitHub, availability as a downloadable file or object in a website, or a response offered by a web service. Most of the time semantic artefacts need to be downloaded and parsed in order to have access to its content, such as concepts/ terms, relations, and metadata.

This hampers the findability of the semantic artefact and makes reuse more difficult (see Rec. 2). To solve these issues, specific repository technologies have been developed to support the publication of semantic artefacts, their content and the metadata associated with the semantic artefacts. These "semantic repositories" provide interfaces for both humans and machines to consume semantic artefacts. They are an important piece of the infrastructure underlying the implementation of FAIR principles and FAIR Semantics as pointed out in the "Turning FAIR into reality" report and action plan (European Commission Expert Group on FAIR Data, 2018).

The number of such repositories is currently increasing with domain specific repositories and registries such as Bioportal, EBI-OLS, Ecoportal[48], Agroportal[49], BODC NERC vocabulary service[50] or more generic services such as Finto.fi[51], BARTOC[52] or Research Vocabularies Australia[53].

In addition, several such semantic repositories have ceased operations or do not exist any longer, and raises the requirement for sustainability. Based on analogies in the data landscape, where trustworthy repositories are defined, and recommended for long-term preservation, we propose a similar approach for semantic repositories.

Such repositories should act as a trustworthy long-term archive, should provide GUPRIs, publish metadata making the semantic artefact findable for humans through a dedicated user interface, and for machines through an API. Trustworthy data repositories, focused on specific disciplines, are available for data curation and preservation, and these repositories may be amenable to preservation of semantic artefacts in the short term. In the medium term, development of criteria for Trustworthy Semantic Repositories is a community responsibility, as is the establishment of mechanisms for certification of compliance. See BP. 11.

This recommendation does not aim to support any particular technology, but emphasizes the necessity to share, publish, and preserve semantic artefacts in such repositories to improve both findability and reuse over time.

48 Ecoportal http://ecoportal.lifewatchitaly.eu/

49 AgroPortal http://agroportal.lirmm.fr/

50 BODC vocabulary service http://seadatanet.maris2.nl/v_bodc_vocab_v2/welcome.asp

51 finto.fi http://finto.fi/en/

52 BARTOC https://bartoc.org

53 Research Vocabularies Australia https://vocabs.ands.org.au/



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

| **Existing technologies** |
|---|
| • SKOSMOS[54]<br>• Bioportal[55]<br>• Research Vocabularies Australia |
| **Stakeholders** |
| Practitioner (ensure that a certified repository is used for publication of the semantic artefact whenever possible)<br>Community |
| **Examples** |
| Nothing precludes the use of trustworthy repositories for data to public and preserve semantic artefacts, provided that such repositories include semantic artefacts into the scope of research outputs that are in their scope for curation. Refer to the current CoreTrustSeal list of certified trustworthy repositories as an example. |

54 SKOSMOS https://www.kansalliskirjasto.fi/en/services/system-platform-services/skosmos

55 Bioportal https://bioportal.bioontology.org/;



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

| | | | |
|---|---|---|---|
| Mandatory | P-Rec. 5: Semantic repositories must offer access to Semantic Artefacts and their content using community standard APIs and serializations to support both use/ reuse and indexation by search engines | F4, A1, A1.1 | Repository |

**Description**

Semantic artefacts are distributed across the web in a variety of locations and formats. Semantic repositories act as aggregators of semantic artefact publishing both the metadata and the content of the semantic artefacts and providing a search engine and an API to search and access the content through dedicated services. However, the APIs to search and access content is specific to each repository which hampers the possibility to access content from multiple sources for use and reuse but also for indexing by search engine. Part of the API heterogeneity is linked with the diverse metadata schema used by repositories to describe semantic artefacts.

To enable federated searches across repositories, it is necessary to harmonize the API landscape by defining a common set of API features[56], based on a common minimum set of metadata for describing semantic artefacts (see P-Rec. 3). See examples below for more details.

This recommendation does not aim to support any particular solution, but emphasizes the need for the community of semantic repositories to agree on a portfolio of common solutions.

**Existing Recommendations**

- The FAIR Data point specification[57]
- Registry Interchange Format - Collections and Services (RIF-CS) (ISO 2146)[58].
- Linked Data Platform[59]
- OpenAPI[60]
- smartAPI[61]

**Stakeholder**

Repository, Community

**Examples**

56 Work in progress

57 FAIR Data point specification https://github.com/FAIRDataTeam/FAIRDataPoint-Spec

58 RIF-CS https://vocabs.ands.org.au/viewById/1

59 Linked Data Platform https://www.w3.org/TR/ldp/

60 OpenAPI https://www.openapis.org/

61 smartAPI https://smart-api.info/



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Enabling automated indexing across repositories will require agents to access a machine-readable description of the API and the description of content that can be accessed. Therefore, repositories should consider publishing at least the description of their API using, for example, OpenAPI specifications which will provide human readable API documentation in a machine-readable format. A recent extension of the OpenAPI specification, called smartAPI[62] has been proposed to provide semantically annotated API description to make an API FAIR. Such semantically enriched description could enable automated workflows for indexing semantic repositories.

Several other possible solutions exist i.e. publishing directly the content as Linked Data and to be compliant with the LD standards (Linked Data Platform, …), use a common inter-exchange metadata format such as RIF-CS or publish metadata and content using the FAIR Data Point service.

If one uses current practices in the data landscape as blueprints[63], then the methods required are typically as a minimum

- Capabilities: understanding what the API offers in terms of scope and methods;
- Collections: lists of semantic artefacts and their UIDs filtered by one or more metadata elements (authors, topics, institutions, and so on). These can also be the result of a query;
- Specific record by UID: a metadata record for a semantic artefact.

These methods are supplemented in some cases by content negotiation.

62 smartAPI https://smart-api.info/

63 For example OAI-PMH and OGC CSW.



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

| | | | |
|---|---|---|---|
| Optional | P-Rec. 6: Build semantic artefact search engines that operate across different semantic repositories | F4 | Repository |

**Description**

To be able to reuse existing semantic artefacts in part or in full, it is necessary to be able to find them across a large number of distributed and heterogeneous semantic repositories. For this, semantic artefact search engines are required that can operate across different semantic repositories. These search engines should

- enable federated queries across the semantic artefacts;
- provide means to gather analytics across all ontologies (overlaps, mappings, reuse);
- and support the use of large-scale automated mappings to resolve semantic ambiguity.

The indices resulting from such federated queries can be composed in real time (strongly dependent on infrastructure and connectivity), or persisted with periodic updates and synchronisation. Such indices could be directly integrated within semantic artefact authoring tools and software to provide access to the existing resource at the time of the creation of a new semantic artefact, or be used to populate lookup lists and vocabulary resources for applications such as metadata editors.

This recommendation emphasizes that such services are an important element of the infrastructure to support FAIR data and FAIR Semantics.

**Existing Recommendations**

None

**Stakeholder**

Community (confirm the need for search engine indices, and fund/ develop such services)
Repositories (enable harvesting of metadata via standard APIs to support federated search engines, apply index-driven selection and referencing of semantic artefacts in software and procedures)
Practitioner (make use of search engine indices to refer to and find semantic artefacts).

**Examples**

GBIF developed a partially manually maintained taxonomy index to support its species occurrence data sets. This task will be automatable in part should federated semantic repositories be available. https://www.gbif.org/dataset/d7dddbf4-2cf0-4f39-9b2a-bb099caae36c

Swoogle provided an early example of such a federated semantic artefact search engine, but did not survive, probably due to missing community support.



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

| Mandatory | P-Rec. 7: Repositories must offer a secure access protocol, and appropriate user access control functionalities | A1.2 | Repository |
|---|---|---|---|

**Description**

Semantic artefacts should be openly shared to support reuse and to avoid concept redundancy and semantic ambiguities. There are, however, several aspects in respect of HTTP protocols, user authentication, and openness to consider.

- Semantic artefacts might be developed under specific copyrights with paywalls (e.g., Dewey Decimal Classification) preventing direct access for use. In such cases, user authentication and a secure access protocol will both be required.
- Openly available semantic artefacts may be managed in an environment where reuse is monitored. In such cases, user authentication is required but such authentication should be based on third-party services via protocols such as OpenID whenever possible, recording as little personal information as possible.
- Even if services are open, and do not require authentication, service endpoints should be made available via a secure HTTP protocol to enable trust in machine-to-machine exchanges.

It is recognised that open access should be the norm (See BP. 12). Personal information should be managed according to applicable legislative frameworks (See BP. 14).

**Existing Recommendations**

N/A

**Stakeholder**

Repository (provide secure access to services)

**Examples**

To be developed and extended in Iteration 3



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

| Mandatory | P-Rec. 8: Human and machine-readable persistence policies for semantic artefacts metadata and data must be published | A2 | Metadata |
|---|---|---|---|

**Description**

Once published in a semantic repository, semantic artefacts will be reused by others to build their information systems. In the eventuality where the semantic artefact, a concept/ term or a relation is deprecated or simply replaced, the repository must publish a persistence policy for the metadata (specifically the duration of archiving of metadata). Such a policy must be both human and machine readable.

Machine readable policies will allow services to automatically detect the change, to either warn the user or to directly integrate the change whenever it is possible. For humans, repositories could use a tombstone page with redirect to the new page when the semantic artefact or the element has been replaced.

Semantic Artefact components (terms/ concepts/ classes) with minimum metadata and GUPRI structures will not, under typical circumstances, be deprecated, but replaced by new versions (P-Rec 1, 2, 3, 4), especially if hosted by trustworthy repositories.

**Existing Recommendations**

RDA - Recommendations on Citation of Evolving Data (Rauber et al., 2015)
RDA - Practical Policy (Moore et al., 2015)
DataCite: Best Practices for Tombstone Pages[64]

**Stakeholder**

Repository (define a publish a preservation policy)
Community (agree a standard for preservation policy encoding and best practices)
Practitioner (agree preservation policy provisions with repository)

**Examples**

If community recommendations similar to those for data are followed for version management, older versions of objects will remain available - both as content and as metadata (Rauber et al., 2015).

RDA has recommended a wide variety of machine-actionable template policies for repositories that can be used as a basis for policy definition and encoding (Moore et al., 2015).

One could also consider an approach similar to the W3C P3P[65] specification, aimed at matching and comparing privacy policies.

64 https://support.datacite.org/docs/tombstone-pages

65 https://www.w3.org/TR/P3P/



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

| Mandatory | P-Rec. 9: Semantic artefacts must be made available as a minimum portfolio of common serialization formats | I1 | Metadata |
|---|---|---|---|

**Description**

Semantic repositories must provide a Linked Data compliant API to enable the creation of a semantic graph for analysis and reuse. This API must support at least one of a portfolio of community-endorsed serialisations.

Semantic artefacts can be serialised in the formats developed in the context of the Semantic Web and Linked Data, i.e. OWL, OBO, RDF, and SKOS. However, these standards have limited capabilities for complex logical models. A good practice should be to share a simplified serialization of the model to provide at least access to the concepts/terms. In P-Rec. 11, we are recommending a portfolio of common serializations to describe complex logical relations. Also see BP. 13.

**Existing Recommendations**

- RDF and RDFS[66]
- W3C-OWL (OWL 2 / Full; OWL 2 / EL; OWL 2 / QL; OWL 2 / RL) stack[67]
- SKOS[68]
- The OBO Foundry Principle 2 “Common Format”.[69]
- Industry Ontology Foundry - requirement 3[70]
- OBO
- Linked Data Platform

**Stakeholder**

Practitioner (use one of the recommended serialisations as a basis for publishing semantic artefacts)
Repository (support at least one serialisation, optionally support crosswalks to others)

**Examples**

The OBO Foundry Principle 2 “Common Format” requires the OWL file in RDF-XML format. Legacy formats are automatically converted to OWL.[69]

66 RDF and RDFS https://www.w3.org/TR/rdf11-primer/

67 W3C-OWL https://www.w3.org/TR/owl2-profiles/

68 SKOS https://www.w3.org/TR/skos-primer/

69 The OBO Foundry principle 2 http://www.obofoundry.org/principles/fp-002-format.html

70 IOF Technical Principles https://www.industrialontologies.org/?page_id=87



FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

| Optional | P-Rec. 10: Foundational Ontologies may be used to align semantic artefacts | I1, I2, I3 | Semantic alignment |
|---|---|---|---|

**Description**

Foundational ontologies are complex logical representations of the basic concepts of the world of discourse. Grounding domain-specific semantic artefacts in foundational ontologies allows the alignment of various domain specific semantic artefacts around a common hypothesis about the world. These semantic artefacts are built to support the integration and the interoperation[71] of domain specific semantic artefacts acting as a language bridging them.

This recommendation is appropriate for formal semantic artefacts (ontologies, thesauri, taxonomies) where such alignment is strongly recommended. Vocabularies and linked lists may not be formal and the recommendation is less critical in such cases.

This recommendation does not make any claim regarding which foundational ontology to use but emphasizes the value of being aligned with one.

**Existing Recommendations**

- Industry Ontology Foundry - requirement 8[72]

**Stakeholder**

Practitioners (use foundational ontologies whenever possible to define new semantic artefacts)

**Examples**

Several foundational ontologies exist, such as UFO[73], BFO[74], DOLCE[75], EMMO[76].

71 In other words: Semantic Interoperability is the main benefit from this recommendation

72 IOF Technical Principles https://www.industrialontologies.org/?page_id=87

73 UFO https://ontouml.readthedocs.io/en/latest/intro/ufo.html

74 BFO https://basic-formal-ontology.org/

75 DOLCE http://www.loa.istc.cnr.it/dolce/overview.html

76 EMMO https://emmc.info/taxonda/emmo-european-materials-modelling-ontology/



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

| Optional | P-Rec. 11: A standardized knowledge representation language should be used for describing semantic artefacts | I1 | Semantic alignment |
|---|---|---|---|

**Description**

The semantic web community should define a common language for high expressivity model representation as per FAIR principle I1.

As discussed in P-Rec. 9, knowledge representation languages such as OWL and RDF, commonly used to represent semantic artefacts by the Semantic Web and Linked Data communities, cannot express all characteristics of more complex/ expressive semantic models. This lack of standard expression and encoding impedes interoperability, since complex semantic artefacts have to be simplified, resulting in a loss of information and expressivity.

Use of such a language, once it has been defined, will be added to the architecture best practice recommendations (BP. 13).

**Existing Recommendations**

- SHACL[77]
- SWRL[78]
- OntoUML[79]

**Stakeholders**

Community (define and maintain, or select and extend, a common standardised language for expressing and encoding complex semantic artefacts

**Examples**

SHACL provides a possible foundation for development of a more expressive language. SWRL serves as an example of extension of OWL, adding rules based on the Rule Markup Language[80].

77 SHACL https://www.w3.org/TR/shacl/

78 SWRL https://www.w3.org/Submission/SWRL/

79 OntoUML https://ontouml.readthedocs.io/en/latest/intro/ontouml.html

80 http://www.ruleml.org/



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

| | | | |
|---|---|---|---|
| Optional | P-Rec. 12: Semantic mappings between the different elements of semantic artefacts should be published in machine-readable formats | I1, I3, R1.3 | Semantic alignment |

**Description**

As we discussed in P-Rec. 10, semantic artefacts are often developed to describe a specific aspect of a scientific domain. Despite this reduced scope, several models of the same aspects can co-exist. They are either developed *de novo* or developed as a part of another ontology.

This leads to potential for divergence, often due to a lack of knowledge regarding existing semantic artefacts. To manage the impacts of such divergence, it is necessary to allow mapping of relations between the elements of such semantic artefacts (ontological alignments[81]).

In many cases, these mappings can be based on existing relations (such as *sameAs* from SKOS). However, mappings can become complex especially when considering logical relations. Risks include representing content drift, as well as context insensitive use of semantic artefacts, and there are no common descriptions for such complex mappings. Mappings are often created by individuals for satisfying a specific need. Information regarding the provenance and usage of these mappings are of importance for any practitioner who would be interested in reusing them.

Principle I3 states that "I3. (Meta)data include qualified references to other (meta)data" - the "qualified references" referred to here are essentially mappings. This recommendation aims at highlighting a gap in relation to the achievement of principle I3, and to emphasize the need for harmonisation. It will require the implementation of machine readable descriptions of mappings in order to foster interoperability.

**Existing Recommendations**

- DOOR[82]
- SSSOM[83]

**Stakeholder**

Practitioner (make use of mappings to manage and understand divergence and content drift)
Repositories (aggregate distributed resources aligned with these different models)
Community (develop a common mapping language that is machine readable)

**Examples**

81 https://en.wikipedia.org/wiki/Ontology_alignment

82 DOOR http://oro.open.ac.uk/24326/1/keod9.pdf

83 https://github.com/OBOFoundry/SSSOM/blob/master/SSSOM.md



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

The DOOR initiative, with an illustrative implementation (KANNEL) provides a recent and extensive overview of the needs for mapping, and proposes a language based on OWL. It allows extensions for versioning.

The Ontology Alignment Evaluation Initiative[84] is in the process of evaluating automated methods for alignment of ontologies, and the formats used for documenting such alignments would provide a good candidate for encoding of mappings.

SSSOM (Simple Standard for Sharing Ontology Mappings) is a more mature, but nevertheless emerging standard, and could be reviewed and extended by the community to support ontological alignments.

Historically, there have been many implementations of automated alignment services (OntoMerge, MAFRA[85]). These can also be used to evaluate and define a common mapping definition language.

The SEMAF project[86] will be determining a framework for semantic mappings and this will result in up-to-date recommendations on the operationalisation of semantic mapping, in time for the FsF project to consider.

84 http://oaei.ontologymatching.org/

85 https://web.archive.org/web/20120423030358/http://sourceforge.net/projects/mafra-toolkit/files/

86 https://drive.google.com/file/d/1vA096ggFP9SIeeHgT-lv7p_thMp-XVvt/view



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

| | | | |
|---|---|---|---|
| Optional | P-Rec. 13: Crosswalks, mappings and bridging between semantic artefacts should be documented, published and curated | R1.2, R1.3 | Semantic alignment |

**Description**

Mappings, crosswalks, content negotiations, and semantic bridges[87] discussed in the previous recommendations (P-Rec. 12) should be made publicly available to allow the reuse by others. Mappings are semantic artefacts in themselves, and therefore should be shared and published in semantic repositories, following the recommendations for such artefacts (for example in respect of GUPRI and metadata).

Sharing these resources in a standardised way will improve interoperability. The main requirement is for such mappings to be machine readable for reuse purposes (P-Rec 11) , and to be described with minimum metadata for human interaction.

**Existing Recommendations**

- SSSOM[88]

**Stakeholder**

Practitioner (provide minimum metadata for semantic mappings)
Repository (ensure the semantic mappings are described, published, and are made available via services)
Community (assist with the definition of a mapping language, and use metadata to cite the reuse of mappings)

**Examples**

SSSOM has made some progress in respect of metadata associated with mappoings, and can be used as a basis for future community

87 These are all mediators in the language of the RDA Brokering Interest Group

88 https://github.com/OBOFoundry/SSSOM/blob/master/SSSOM.md



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

| Optional | P-Rec. 14: Standard vocabularies should be used to describe semantic artefacts | I2 | Metadata |
|---|---|---|---|

**Description**

As stated in P-Rec. 3, the semantic artefact metadata is important for both findability and reusability. Standard vocabularies used by metadata schema, such as Dublin Core, schema.org, and DataCite, for example, should be used to describe such semantic artefacts. Agreeing on a common set of standard vocabularies would allow to improve the interoperability of the metadata descriptions. This should also apply to the metadata associated with the content.

**Existing Recommendations**

- LOV recommendations[89]
- FDP metadata scheme[90]
- OBO Foundry

**Stakeholder**

Practitioner (apply standard vocabularies proposed for minimum metadata elements)
Community (agree on a set of vocabularies to support minimum metadata)

**Examples**

DataCite provides a set of controlled vocabularies[91] for use with its metadata schema, and these are available for re-purposing to describe the elements of a minimum metadata schema for semantic artefacts in a controlled manner. These are currently published as XSD schema.

Schema.org provides vocabularies too[92], available as OWL, and allows community participation in definition of vocabularies - this may also be a useful vehicle for the semantic artefact community to use. It is potentially more aligned with community standards[93] and can be downloaded/ accessed in several popular serialisation formats.

Dublin Core[94] remains a major consideration for reuse of vocabularies to describe semantic artefacts.

The Fair Data Point Specification90, in addition to definitions of its own, combines elements from Dublin Core, RDF, and re3data and should be used as a basis for semantic repository metadata specifications.

89 LOV https://lov.linkeddata.es/dataset/lov/

90 FAIR Data point specification https://github.com/FAIRDataTeam/FAIRDataPoint-Spec

91 https://schema.datacite.org/meta/kernel-4.3/doc/DataCite-MetadataKernel_v4.3.pdf, p32

92 https://schema.org/docs/kickoff-workshop/sw1109_Vocabulary.pdf

93 https://schema.org/version/latest/schemaorg-current-http.jsonld

94 https://www.dublincore.org/specifications/dublin-core/dcmi-terms/



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

A criterium for vocabularies might be inclusion in Linked Open Vocabulary, which confirms availability in formats that are machine-readable and standardised.



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

| Optional | P-Rec. 15: Provenance information regarding the reuse of components from third-party semantic artefacts should be made explicit | I3, R1.2 | Metadata |
|---|---|---|---|

**Description**

New semantic artefacts can be built upon existing artefacts. In some cases, reuse involves inclusion of elements of one artefact to another. This mechanism does not allow automatic access to reused elements and their containing semantic artefact. In order to be able track, the element reused from other semantic artefacts, reference to third party semantic artefacts should be made explicit. For this, external semantic artefacts (in part or in full) should be imported using a specific metadata element to represent the import (e.g. <owl:import/>). When using such explicit references, it becomes possible to extract this dependency information automatically from the artefacts.

This recommendation emphasizes the need for the community of semantic web to define a common standard for referencing or importing external (third-party) semantic artefacts.

**Existing Recommendations**

- LOV[95]
- VoID[96]
- MIREOT[97]

**Stakeholder**

Practitioner (follow recommendations when using elements from third-party semantic artefacts)
Community (agree on standard for referencing external semantic artefacts)

**Examples**

The requirement here is for explicit references to third-party. he explicit reference can be made

- by providing a direct link to the used resource, preferably as a GUPRI
- or by describing the link as inbound and outbound links as proposed by LOV
- or by using VoID vocabulary to interlink the different ontologies
- or by considering the requirement of the Minimum Information to Reference an External OnTology, MIREOT (Courtot et al., 2009).

A concrete example is provided by BODC - references to external vocabulary sources are provided with small metadata extensions, stored in a separate file, and referenced by the instances using external semantic artefacts. For example, many of the statistical terms used in BODC vocabularies are defined elsewhere[98].

95 LOV https://lov.linkeddata.es/dataset/lov/

96 VoID https://www.w3.org/TR/void/

97 MIREOT http://precedings.nature.com/documents/3574/version/1

98 https://vocab.nerc.ac.uk/collection/P01/current/



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

| Mandatory | P-Rec. 16: The semantic artefact must be clearly licenced for use by machines and humans | R1.1 | Metadata |
|---|---|---|---|

**Description**

Proper reuse of digital objects requires a human and machine-readable licence. Well documented legal interoperability is a prerequisite for automatic distributed search, and the use of both the semantic artefacts and their component terms/concepts/classes and relations.

Although we are encouraging Open licences, preferably using Creative Commons 4.0 licencing, this recommendation doesn't impact the choice but emphasizes the need for adding this information explicitly for both human and machine to avoid ambiguities on the conditions for reuse. See BP. 12 for more information.

**Existing Recommendations**

- Creative Commons licences[99]
- ODRL[100]

**Stakeholder**

Practitioner (ensure that the appropriate license is communicated to the repository, and ensure that the right to decide the license vests with the practitioner)
Repository (ensure that a choice of machine readable licenses are offered to practitioners wishing to publish)

**Examples**

DataCite provides an example of licensing embedded into the metadata schema, and there is an option for using signposting[101] to do the same.

It is only required to reference a URI (IRI) to a machine-readable license. The license itself, to be machine readable, will need to be encoded using a standard such as ODRL [100].

99 Creative Commons licences https://creativecommons.org in rdf https://github.com/creativecommons/cc.licenserdf

100 ODRL https://www.w3.org/TR/odrl-model/

101 https://signposting.org/



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

| Mandatory | P-Rec. 17: Provenance must be clear for both humans and machines | R1.2 | Metadata |
|---|---|---|---|

**Description**

Semantic artefacts are living digital entities undergoing changes and revisions to cope with semantic drift and for improving/ extending the scope or granularity. Provenance information describing all these changes during the semantic artefact lifecycle should be provided to external users.

Provenance must be documented at an appropriate level of granularity to enable reuse of semantic artefacts and its components (class/ term/ concept and relation). This means that provenance is applicable both at the level of the semantic artefact as a container (where versioning may be the primary consideration) and components (where references to external artefacts are important in addition to versioning).

Provenance must be presented to the human user, but also should be expressed in a machine-readable way. All appropriate sources should be referred to (both source reference - semantic artefact, and creator reference) and the provenance should provide dates and lifecycle events.

Provenance information should be described using an appropriate standard model such as PROV[102]. PROV-based machine-readable description could be then used to provide means to automatically update any resource using the semantic artefact. The provenance information should contain all the necessary elements to build representations to the users such as changelogs and describe backward interoperability.

**Existing Recommendations**

- MIRO (Matentzoglu et al., 2018)
- OBO foundry - Principle 4[103]
- OBO Foundry - Principle 8[104]
- PROV [102]
- Metadata schema elements defining versioning and provenance (D-CAT, DataCite)

**Stakeholder**

Practitioner (ensure that provenance in respect of versioning and external references are properly made)
Repository (compose a metadata record compiling a list of external contributions from detailed third-party references, based on detailed information resulting from P-Rec 15).

**Examples**

DataCite implements provenance by including external PID references in respect of previous versions, and a vocabulary defining the inputs used to create a new object - it provides a comprehensive list of relationship

102 PROV data model https://www.w3.org/TR/prov-primer/

103 OBO foundry principle 4 http://www.obofoundry.org/principles/fp-004-versioning.html

104 OBO foundry principle 8 http://www.obofoundry.org/principles/fp-008-documented.html



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

| |
|---|
| types to describe provenance[105]. |

## 3. Recommendations Beyond the FAIR Principles: Best Practices for Semantic Artefacts

In this section, recommendations are listed that do not apply to a particular FAIR principle but contribute to the improvement of Findability, Accessibility, Interoperability and Reusability of the semantic artefacts

| BP. 1: Use a unique naming convention for concept/class and relations |
|---|
| **Description** |
| Concept/class and relations composing semantic artefact are associated with a human readable label which is a character string. This string can become complex and include several associated words (e.g. Hyperplastic and giant kidney**).** There are multiple conventions for naming semantic artefact elements such as CamelCase or the conventions proposed by OBO Foundry and Industry Ontology Foundry. Unfortunately, these existing conventions/recommendations are not harmonized which leads to the need for a search engine or an automated mapping service to comply with the different conventions/ recommendations. This hampers both searching capabilities and automated mappings.<br><br>This recommendation for Best Practice emphasizes the need to define a common unique naming convention by the community of practitioners. |
| **Existing Recommendations** |
| ● OBO Foundry - Principle 12 (Schober et al., 2009)[106]<br>● Industry Ontology Foundry - requirement 11[107] |
| **Stakeholders** |
| Practitioners, Community |

105 https://schema.datacite.org/meta/kernel-4.3/doc/DataCite-MetadataKernel_v4.3.pdf p 23.

106 OBO Foundry principle 12 http://www.obofoundry.org/principles/fp-012-naming-conventions.html

107 IOF Technical Principles https://www.industrialontologies.org/?page_id=87



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

## BP. 2: Use an Ontology Naming Convention

**Description**

Semantic artefacts often have a human readable name associated with an acronym. The name provides information about the general topic covered by the semantic artefact while the acronym can be used as prefix to create URI-based GUPRI for concepts and terms (see OBO Foundry naming convention). The governance of these names is managed by organisations such as the OBO Foundry or the Industry Ontology Foundry. As of now, there is no widespread consensus on the naming of semantic artefacts and the use of acronyms/ prefixes, which leads to ambiguities and non-uniqueness (see Goldfarb and Le Franc, 2017). The community of practice should consider addressing this issue by defining a common governance model for semantic artefacts.

**Existing recommendations**

- OBO Foundry - Principle 3[108]
- Industry Ontology Foundry[109]

**Stakeholders**

Community

## BP. 3: Use defined ontology design patterns

**Description**

To support interoperability, semantic artefacts should be designed based on well-supported patterns whenever relevant and possible. These patterns should be documented and published as a resource for practitioners following the example of OntologyDesignPatterns.org[110] that focuses on OWL design patterns.

Such patterns cover elements of construction of and representation of classes, axioms, and the standard representation of relations.

**Existing recommendations**
OntologyDesignPatterns.org [110]

**Stakeholders**

108 OBO Foundry principle 3 http://www.obofoundry.org/principles/fp-003-uris.html

109 IOF Technical Principles https://www.industrialontologies.org/?page_id=87

110 OntologyDesignPatterns.org http://ontologydesignpatterns.org/



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

| |
|---|
| Practitioner<br>Community |

## BP. 4: Allow mappings to reflect validation by domain experts

**Description**

Semantic artefacts, and in particular concept definitions, vary within and between communities. This diversity of definitions generates semantic ambiguities which hampers interoperability between ontologies. To support such interoperability, explicit mappings should *ideally* be generated by knowledge experts, and *validated* by domain experts.

It is, however, increasingly likely that validation will be automated, for example using the approach of Ontology[111]. using (standardised) validation tests to evaluate ontology.

**Existing recommendations**

None

**Stakeholders**

Practitioner
Community

## BP. 5: Define crosswalks between different formats

**Description**

Semantic artefact can be serialized in various formats (SKOS, RDF, OWL, XML, …). This diversity of formats makes it complicated to integrate and work with heterogeneous semantic artefacts. Practitioners should describe or reference the particular crosswalk they used to convert the semantic artefact from one format to another. These workflows could be defined using machine readable mappings.

**Existing recommendations**

N/A

111 https://github.com/OnToology/OnToology



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

| Stakeholders |
| --- |
| Practitioner<br>Community |

## BP. 6: Harmonize the methodologies used to develop semantic artefacts

| Description |
| --- |
| Semantic artefacts can be built with very different methodologies (processes) depending on the available resources and the expertise of the practitioner. These methodologies should be documented or references should be provided to published third-party methodologies. This will allow users of the artefact to assess the quality and applicability of the semantic artefact.<br><br>To support interoperability and prevent poorly formed ontologies (in terms of structure and content), the community of practice should work toward harmonising, documenting, and publishing methodologies. These can be used as a training resource for newcomers and guidelines for expert practitioners. |
| **Existing Recommendations** |
| • OBO Foundry - Principle 8[112]<br>• Industry Ontology Foundry - Principle 9[113]<br>• OntologyDesignPatterns.org [110] |
| **Stakeholders** |
| Community<br>Practitioner |

## BP. 7: Interact with the designated community and manage user-centric development

| Description |
| --- |
| A semantic artefact needs to reflect the actual semantic model/ conceptualisation embraced and endorsed |

112 OBO Foundry principle 8 http://www.obofoundry.org/principles/fp-008-documented.html

113 IOF Technical Principles https://www.industrialontologies.org/?page_id=87



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

by the community that uses it. As science and language changes, changes and updates are inevitable in the long run or content drift will happen in the semantics. This has to be acknowledged and managed. As others (re)use semantic artefacts to enable interoperability there has to be even more clear processes for managing the necessary communication and negotiations.

Semantic artefacts have to be regarded as services to the community in their own right, not only service components, especially if they are reused (used by several systems or solutions). Thus ownership as well as continuous user centred development should be ensured.

Interaction and communication with the designated community has to be organised and managed as a process. The community should receive training and guidance in using and developing the semantic artifact.

**Existing Recommendations**

- OBO Foundry - Principle 9 Users[114]
- OBO Foundry - Principle 11 Authority[115]
- OBO Foundry - Principle 16 Maintenance[116]
- IOF Principle 16[117]
- IOF Principle 15
- IOF Principle 14

**Stakeholders**

Practitioner

## BP. 8: Provide a structured definition for each concept

**Description**

Semantic artefacts are used to annotate data. Concepts are the key elements of the semantic artefacts used by the annotators. Annotators need both a human readable and logical definition to make a decision on using a specific term. Human readable definitions are therefore crucial for the reuse of semantic artefacts. When building a semantic artefact, one the main challenge is to write a structured definition for concepts. There are recommendations providing guidelines for writing human readable definition as OBO Foundry principle 6 and the IOF principle 10.

114 OBO Foundry principle 9 http://www.obofoundry.org/principles/fp-009-users.html

115 OBO Foundry principle 11 http://www.obofoundry.org/principles/fp-011-locus-of-authority.html

116 OBO Foundry principle 16 http://www.obofoundry.org/principles/fp-016-maintenance.html

117 IOF Technical Principles https://www.industrialontologies.org/?page_id=87



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

| |
|---|
| A set of 11 guidelines have been proposed following up a dedicated series of workshops and a survey on the usage of definition (Seppälä et al., 2017). A recent blog post from C. Mungall describe in more detail how to write simple and concise definitions[118]. |
| **Existing Recommendations** |
| • OBO Foundry - Principle 6[119]<br>• IOF - Principle 10[120]<br>• Guidelines for writing definition in ontologies[121] |
| **Stakeholders** |
| Practitioners |

## BP. 9: The underlying logic of semantic artefacts should be grounded in the domain it intends to be used in

| |
|---|
| **Description** |
| Semantic artefacts are developed within research communities and represent a specific and restricted domain of discourse. Reuse of such ontologies by stakeholders outside of the community raises questions regarding the relations between the ontologies and scientific domains. When reusing ontologies, practitioners should strive to choose semantic artefacts with highest precision (e.g., existence of information cardinality and value), the most well defined documentation including information regarding cardinality and value type when applicable.<br><br>Whenever the semantic artefact is reused, it should be extended in granularity depending on use-case. This implies that the reuser should adhere to the same design principle. |
| **Existing Recommendations** |
| N/A |
| **Stakeholders** |

118 OntoTip: Write simple, concise, clear, operational textual definitions https://douroucouli.wordpress.com/2019/07/08/ontotip-write-simple-concise-clear-operational-textual-definitions/

119 OBO Foundry principle 6 http://www.obofoundry.org/principles/fp-006-textual-definitions.html

120 IOF Technical Principles https://www.industrialontologies.org/?page_id=87

121 Guidelines for writing definitions in ontologies https://philpapers.org/archive/SEPGFW.pdf



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

| |
|---|
| Practitioner |

## BP. 10: Define a set of governance policies for the semantic artefacts

| **Description** |
|---|
| Semantic artefacts are living entities that are undergoing changes through their lifecycle. As an example, when used these artefacts are becoming part of a data service and therefore changes and updates should be published to warn users that the data service should be updated. It is therefore crucial to have well identified governance policies for the semantic artefact. These policies should be available in human readable format but also whenever possible in machine actionable format to allow the automation of change propagation. They should cover the various aspects of the semantic artefact life cycle i.e. versioning policy, deprecation policy, contribution policy. |
| **Existing Recommendations** |
| N/A |
| **Stakeholders** |
| Practitioner<br>Repository |

## BP. 11: Use TRUSTed and FAIR compliant repositories to persist Semantic Artefacts

| **Description** |
|---|
| Certification authorities for verification of the trustworthiness of digital repositories exist for data repositories, and it is likely that similar infrastructure will develop for certification of FAIRness of repositories, and for TRUST in respect of Semantic Artefacts.<br><br>TRUST deals with repository governance, policies, sustainability, expertise, quality assurance, and reliability of infrastructure. Development of criteria for such repositories is a community responsibility, and the criteria can be based on those developed for data repositories[122]. |

122 CoreTrustSeal Trustworthy Data Repositories Requirements 2020–2022, CoreTrustSeal Standards and Certification Board https://zenodo.org/record/3638211#.XrEoXhMzaRs



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

| |
|---|
| Committing semantic artefacts to code repositories such as github or gitlab is not recommended as a long-term preservation strategy. In the short term, in the absence of a semantic artefact repository infrastructure, these platforms will be adequate, provided that they are referenced through a GUPRI as per P-Rec 1. Any changes in location of the stored artefacts should be handled by the GUPRI resolution. |
| **Existing Recommendations** |
| TRUST Principles for Digital Repositories, https://doi.org/10.1038/s41597-020-0486-7 |
| **Stakeholders** |
| Practitioner<br>Repository<br>Community |

| BP. 12: Semantic Artefacts, if developed with public funding, should be published with open licenses unless one of the overriding conditions discussed below are true |
|---|
| **Description** |
| It is now common practice for publicly funded research outputs to be open by default. The exception will be if one of the conditions listed below are true:<br>1. Uncontrolled disclosure could harm an individual, an ecosystem, or a community.<br>2. It contains commercially sensitive information<br>3. It contains classified government information<br>Open licenses (for example Creative Commons[123]) address such limitations such as reuse, commercial use, and changes to license conditions, and generally requires proper citation. |
| **Existing Recommendations** |
| Horizon 2020: As open as possible, as closed as necessary[124] |
| **Stakeholders** |
| Practitioner<br>Repository<br>Community |

123 https://creativecommons.org/

124 https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

| BP. 13: A standard architecture for semantic artefact management, services for content and metadata, protocols, and serialisations/ content negotiation should be proposed, and semantic repositories are urged to use this architecture as a design pattern |
|---|
| **Description** |
| A standard architecture ('best practice') can be defined for implementation of infrastructure to support semantic artefacts, based on the recommendations. |
| **Existing Recommendations** |
| N/A |
| **Stakeholders** |
| Practitioner<br>Repository<br>Community |

| BP. 14: User registration and authentication should be as simple as possible and adhere to legislative requirements |
|---|
| **Description** |
| When authenticating users, the following best practice is recommended:<br>● If the only requirement is to log usage of semantic artefacts, it is best to base authentication on third party services compliant with OpenID (GeANT in the EU, ORCID, eduroam, etc.)<br>● If more information is required about the user (for example if a subscription service is involved), then the prescriptions of the legislative environment need to be met (for example GDPR in Europe). |
| **Existing Recommendations** |
| N/A |
| **Stakeholders** |



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

| |
|---|
| Repository<br>Community |

## 4. FAIR Principles Coverage

| | FAIR Principles | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| P-Rec | F1 | F2 | F3 | F4 | A1 | A1.1 | A1.2 | A2 | I1 | I2 | I3 | R1 | R1.1 | R1.2 | R1.3 |
| 1 | | | | | | | | | | | | | | | |
| 2 | | | | | | | | | | | | | | | |
| 3 | | | | | | | | | | | | | | | |
| 4 | | | | | | | | | | | | | | | |
| 5 | | | | | | | | | | | | | | | |
| 6 | | | | | | | | | | | | | | | |
| 7 | | | | | | | | | | | | | | | |
| 8 | | | | | | | | | | | | | | | |
| 9 | | | | | | | | | | | | | | | |
| 10 | | | | | | | | | | | | | | | |
| 11 | | | | | | | | | | | | | | | |
| 12 | | | | | | | | | | | | | | | |
| 13 | | | | | | | | | | | | | | | |
| 14 | | | | | | | | | | | | | | | |
| 15 | | | | | | | | | | | | | | | |
| 16 | | | | | | | | | | | | | | | |
| 17 | | | | | | | | | | | | | | | |



FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

## 5. Final Remarks

This document presents the second iteration of the FAIR Semantics recommendations and best practices.

An approach proposing a monolithic and static set of recommendations would clearly fail to capture the diversity of the semantic usages. Therefore, in the FAIRsFAIR project we are using a more agile approach which allows the modification, extension and update of the recommendations based on expert feedback. A final round of community engagement (see workshops proposed below) is planned prior to finalisation of the recommendations and best practices.

Based on the feedback received from the community during the Workshop in October 2020, and internal review of the second iteration of recommendations (D2.5), we have identified conceptual and design work to be done in support of the recommendations. The scope of this work, which is under way and being documented prior to community engagement, is the following:

- **Conceptual Model**: in essence an ontology of semantic artefacts, their context, repositories, stakeholders and users, classes, and properties. This is needed to address concerns in respect of better definition of terms and concepts, and to serve as a basis for a vocabulary for semantic artefacts (to be used in artefact and repository metadata, for example).
- **Design Considerations and Architecture:** defining the requirements and guidelines for:
    - ***Globally Unique, Persistent, Resolvable Identifiers***: In the world-wide web, URIs and URLs are unique, and hence each file or object referenced at a given URI is also unique (resources). Guaranteeing uniqueness is already built into the design of the semantic web, but persistence is not. We need community consensus on acceptable approaches.
    - ***Value Chain for Semantic Artefacts*: The question arises, at what point does a Semantic Artefact require a GUPRI?**, and what should be considered as minimum metadata appropriate to its context?
    - ***Metadata***
        - Semantic Artefacts - minimum metadata elements, based on inputs including OBO, D-CAT, DataCite, and norms for other research outputs, and considering the value chain implicit in production of such artefacts - not all artefacts require the same level of detail in metadata;
        - Repositories - describing semantic repositories, in line with existing infrastructure.
    - **Actors, Roles, and Responsibilities:** definition of the actors (people, institutions, systems) involved, and the relationships between them.



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

- **Vocabularies**: Based on the conceptual model and the design considerations discussed earlier, it is clear that a portfolio of standard vocabularies will be required to manage FAIR Semantic Artefacts - specifically in respect of artefact and repository metadata.
- **Canonical Artefacts**: During the next iteration (D2.8, due in month 34), a set of canonical artefacts will be identified, and where necessary developed to serve as examples of the details discussed in many of the recommendations.
- **Infrastructure and a Reference Implementation**: The third Iteration will also be used to provide a simple reference implementation with examples of important architecture elements.

For this purpose, we will engage stakeholders in collaboration in the following ways:

1. Organising workshops in the first half of 2021 to discuss the following:
    a. Conceptual basis for the recommendations, minimum metadata, design considerations, and architecture: a community workshop (Workshop 2021-1);
    b. Consultation with experts, including non-EU experts, in respect of best practice, recommendations, and implementations that reflect these (Workshop 2021-1a), and
    c. Final review of recommendations, proposed restructuring, and the FAIR Semantics roadmap.
2. Continuing the open consultation with the international community of semantic practitioners by publishing the recommendations on the FsF FAIR Semantics GitHub; and
3. Reinitiating discussions within the international and pluri-disciplinary networks/initiatives with which we have partnered as part of the evaluation of the 1st Set of FAIR Semantics Recommendations i.e., the Vocabulary Services & Semantics Interest Group (RDA), the GO INTER implementation network[125] (GO FAIR)

125 GO INTER https://www.go-fair.org/implementation-networks/overview/go-inter/



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

# 6. References

Sources like services, web resources, policy documents and specifications are given in footnotes as full http links to enable easy access while reading this document.

Courtot, M, Gibson, F, Lister, AL, Malone, J, Schober, D, Brinkman, RR, Ruttenberg, A, 2009. 'MIREOT: the Minimum Information to Reference an External Ontology Term', *Nature Precedings*. https://doi.org/10.1038/npre.2009.3574.1

d'Aquin, M, Noy, NF, 2012. 'Where to publish and find ontologies? A survey of ontology libraries', *Journal of Web Semantics*, vol. 11, pp. 96-111. https://doi.org/10.1016/j.websem.2011.08.005

Dutta, B, Toulet, A, Emonet, V, Jonquet, C, 2017. 'New Generation Metadata Vocabulary for Ontology Description and Publication', in: Garoufallou, E, Virkus, S, Siatri, R, Koutsomiha, D (eds.), *Metadata and Semantic Research, Communications in Computer and Information Science*. Springer International Publishing, Cham, pp. 173-185. https://doi.org/10.1007/978-3-319-70863-8_17

European Commission Expert Group on FAIR Data, 2018. *Turning FAIR into reality: final report and action plan from the European Commission expert group on FAIR data*. Directorate-General for Research and Innovation (European Commission), Brussels. https://doi.org/10.2777/1524

Goldfarb, D, Le Franc, Y. 2017. 'Enhancing the Discoverability and Interoperability of Multi-Disciplinary Semantic Repositories', in: *S4BioDiv@ISWC.* Presented at the *2nd International Workshop on Semantics for Biodiversity co-located with 16th International Semantic Web Conference (ISWC 2017*), Vienna.

Gruber, T.R., 1993. 'A Translation Approach to Portable Ontology Specifications', *Knowledge Acquisition*, vol. 5, pp. 199–220.

Guarino, N, Oberle, D, Staab, S, 2009. 'What Is an Ontology?', in: Staab, S, Studer, R (eds.), *Handbook on Ontologies, International Handbooks on Information Systems.* Springer, Berlin, Heidelberg, pp. 1–17. https://doi.org/10.1007/978-3-540-92673-3_0

Hartmann, J, Palma, R, Sure, Y, Suárez-Figueroa, MC, Haase, P, Gómez-Pérez, A, Studer, R, 2005. 'Ontology Metadata Vocabulary and Applications', in: Meersman, R, Tari, Z, Herrero, P (eds.), *On the Move to Meaningful Internet Systems 2005: OTM 2005 Workshops, Lecture Notes in Computer Science*. Springer, Berlin, Heidelberg, pp. 906–915. https://doi.org/10.1007/11575863_112

Hellström, M., Heughebaert, A, Kotarski, R, Manghi, P, Matthews, B, Ritz, R, Conrad, AS, Weigel, T, Wittenburg, P, 2019. *Initial Persistent Identifier (PID) policy for the European Open Science Cloud (EOSC)*. https://doi.org/10.5281/zenodo.3574203



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Hugo, W, Le Franc, Y, Coen, G, Bonino, L, & Parland-von Essen, J,. (2020, October 30). *Report on the FAIR Semantics workshop to discuss ontology design good practices and evaluate the first recommendations and the roadmap* (M2.6) (Version 1.0). Zenodo. http://doi.org/10.5281/zenodo.431430

Kulvatunyou, B, Wallace, EK, Kiritsis, D, Smith, B, Will, C, 2018. The Industrial Ontologies Foundry Proof-of-Concept Project.

Matentzoglu, N, Malone, J, Mungall, C, Stevens, R, 2018. 'MIRO: guidelines for minimum information for the reporting of an ontology.' *Journal of Biomedical Semantics*, vol. 9: 6. https://doi.org/10.1186/s13326-017-0172-7

Moore, Reagan, Stotzka, Rainer, Cacciari, Claudia, & Benedikt, Petr. (2015, February 26). Practical policy (Version 1.0). http://doi.org/10.15497/83E1B3F9-7E17-484A-A466-B3E5775121CC

Neuhaus, F, 2017. 'On the definition of 'ontology,'' in: *Proceedings of the Joint Ontology Workshops 2017.* Presented at the *Episode 3: The Tyrolean Autumn of Ontology, Sun SITE Central Europe* (CEUR), Bozen-Bolzano, Italy.

Obrst, L, 2010. 'Ontological Architectures', in: Poli, R, Healy, M, Kameas, A (eds.), *Theory and Applications of Ontology: Computer Applications*. Springer Netherlands, Dordrecht, pp. 27–66. https://doi.org/10.1007/978-90-481-8847-5_2

Obrst, L, 2003. 'Ontologies for semantically interoperable systems', in: *CIKM '03: Proceedings of the Twelfth International Conference on Information and Knowledge Management.* Presented at the *The twelfth international conference on Information and knowledge management*, New Orleans, Louisiana, USA, pp. 366-369. https://doi.org/10.1145/956863.956932

Ong, E, Xiang, Z, Zhao, B, Liu, Y, Lin, Y, Zheng, J, Mungall, C, Courtot, M, Ruttenberg, A, He, Y, 2017. 'Ontobee: A linked ontology data server to support ontology term dereferencing, linkage, query and integration', *Nucleic Acids Research,* vol. 45, D347-D352. https://doi.org/10.1093/nar/gkw918

Rauber, A, Asmi, A, van Uytvanch, D, Proell, S, 2015. *Data Citation of Evolving Data: Recommendations of the Working Group on Data Citation (WGDC).* http://doi.org/10.15497/RDA00016

Schober, D, Smith, B, Lewis, SE, Kusnierczyk, W, Lomax, J, Mungall, C, Taylor, CF, Rocca-Serra, P, Sansone, S-A, 2009. 'Survey-based naming conventions for use in OBO Foundry ontology development', *BMC Bioinformatics*, vol. 10, 125. https://doi.org/10.1186/1471-2105-10-125

Seppälä, S, Ruttenberg, A, Smith, B, 2017. 'Guidelines for writing definitions in ontologies' *Ciência da*



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

*Informação*, vol. 46 (1), pp. 73-88 Available at: http://revista.ibict.br/ciinf/article/view/4015

Smith, B, Ashburner, M, Rosse, C, Bard, J, Bug, W, Ceusters, W, Goldberg, L.J, Eilbeck, K, Ireland, A, Mungall, CJ, Leontis, N, Rocca-Serra, P, Ruttenberg, A, Sansone, S-A, Scheuermann, RH, Shah, N, Whetzel, PL, Lewis, S, 2007. 'The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration' *Nature Biotechnology*, vol. 25, pp. 1251-1255. https://doi.org/10.1038/nbt1346

TFIR - European Commission Expert Group on FAIR Data, 2018. *Turning FAIR into reality: final report and action plan from the European Commission expert group on FAIR data*. Directorate-General for Research and Innovation (European Commission), Brussels. https://doi.org/10.2777/1524

Vandenbussche, P-Y, Atemezing, GA, Poveda-Villalón, M, Vatant, B, 2017. 'Linked Open Vocabularies (LOV): A gateway to reusable semantic vocabularies on the Web', *Semantic Web,* vol. 8, pp. 437-452. https://doi.org/10.3233/SW-160213

Whetzel, PL, Noy, NF, Shah, NH, Alexander, PR, Nyulas, C, Tudorache, T, Musen, MA, 2011. 'BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications.' *Nucleic Acids Research*, vol. 39, pp. W541-W545. https://doi.org/10.1093/nar/gkr469

Wilkinson, MD, Dumontier, M, Aalbersberg, IjJ, Appleton, G, Axton, M, Baak, A, Blomberg, N, Boiten, J-W, da Silva Santos, LB, Bourne, PE, Bouwman, J, Brookes, AJ, Clark, T, Crosas, M, Dillo, I, Dumon, O, Edmunds, S, Evelo, CT, Finkers, R, Gonzalez-Beltran, A, Gray, AJG, Groth, P, Goble, C, Grethe, JS, Heringa, J, 't Hoen, PAC, Hooft, R, Kuhn, T, Kok, R, Kok, J, Lusher, SJ, Martone, ME, Mons, A, Packer, AL, Persson, B, Rocca-Serra, P, Roos, M., van Schaik, R, Sansone, S-A, Schultes, E, Sengstag, T, Slater, T, Strawn, G, Swertz, MA, Thompson, M, van der Lei, J, van Mulligen, E, Velterop, J, Waagmeester, A, Wittenburg, P, Wolstencroft, K, Zhao, J, Mons, B, 2016. 'The FAIR Guiding Principles for scientific data management and stewardship.' *Scientific Data 3*, vol. 160018. https://doi.org/10.1038/sdata.2016.18

Wilkinson, MD, Dumontier, M, Sansone, S-A, Bonino da Silva Santos, LO, Prieto, M. Batista, D, McQuilton, P, Kuhn, T, Rocca-Serra, P, Crosas, M, Schultes, E, 2019. 'Evaluating FAIR maturity through a scalable, automated, community-governed framework ' *Scientific Data 6,* 174. https://doi.org/10.1038/s41597-019-0184-5



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

# 7. Appendix A: Initial list of recommendations compiled from internal and from workshop discussions

This appendix compiles the outcomes of the workshop discussion (Espoo) that were presented during the summary session. These recommendations were aggregated with additional comments written on paper board during the brainstorming session and led to 47 recommendations, after removing duplicates.

| Recommendations/Requests/Requirements | FAIR Principles |
|---|---|
| Persistent identifiers for terms and terminologies (not metainfo about artifacts) (ensure unique namespace across terminologies) | F1 |
| Schema covering the core set of metadata for representing terminologies<br>i. Check the schema used by Taxonda<br>ii. Related Paper by Clement Jonquet et al. (Metadata for ontology description), schema.org implementation to enable findability by search engines | F2 |
| Business model for private ontologies - access to metadata only which describes access information | F2 |
| Multilingual ontologies - (a) term/concept - the same label represented in multiple languages, (b) terminology descriptions | F2 |
| Schema elements should support values based on persistent identifiers (ontology creator (orcid), organization (grid, ror,..) | F2 |
| (FAIR+Q) ranking/recommending terminologies based on<br>i. Usage information<br>ii. Consistency checks<br>iii. Coverage | F2 |
| Concepts/Terms alignment between terminologies | F2 |
| Metadata about an ontology should also include the identifier of the ontology | F3 |
| Federated search across ontologies repositories for both humans and machines, through shared ontology API (versioning info included in search results, compare versions) | F4 |
| General and domain-specific registry for ontologies, and registry for ontologies repositories | F4 |



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

| | |
|---|---|
| FAIR ontology repositories | F4 |
| What should an identifier of a terminology should resolve to? Ontology file, ontology landing page, sparql query | F4 |
| schema.org implementation to enable findability by search engines | F4 |
| Open API to access terminologies, in various serializations, e.g., bioportals Resolve terms URIs to their description pages | A1.1 |
| Support accessibility of terminologies through HTTPS secure communication | A1.2. |
| User access control | A1.2 |
| Make the landing page of a terminology available if the source file is not accessible (e.g., commercially developed terminology) | A2 |
| Should ontologies be aligned with upper-ontology (e.g. BFO, DOLCE) as part of the FAIR maturity indicators | I1 |
| Implement ontology alignment | I1 |
| Involve domain expertise in alignment validation | I1 |
| Recommend the use of ontology design patterns and other shared practices for ontology development | I1,I2 |
| Identify the version of ontologies using unique permanent identifiers | I2 |
| Use reification to overcome ontology mismatch (e.g. when searching across many datasets described using different conflicting ontologies | I1,I3 |
| Need to evaluate the quality of the ontology | N/A |
| Granularity: ontology to be extended in granularity depending on use-case. Reuser should adhere to the same design principles | N/A |
| Same practice used in software development should apply to ontologies | N/A |
| Should be maintained by community in wikidata style | N/A |
| Versioning and governance | N/A |
| Implement mechanisms to warn users when a new version is out | N/A |
| Training people to create and maintain semantic artefacts | N/A |



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

| | |
|---|---|
| Need to provide information regarding cardinality and value type when applicable | R1 |
| Need proper definition to be able to evaluate the difference between similar classes from different ontologies | R1 |
| Should have a contact point and feedback mechanisms | R1 |
| Two levels of metadata description: ontology for retrieving and evaluating the ontology and then the metadata about the content: should look into/use existing metadata schemata to describe ontologies (OMV, MOD, …) | R1 |
| Strive to choose semantic artefacts with highest precision (e.g. existence of information cardinality and value) | R1 |
| Strive to choose semantics which have most well defined documentation | R1 |
| Clarify the roles and responsibilities for the Ontology Engineer, Domain (community) expert, data steward, researcher, the need for common tools, the possibilities for automation and manual checking | R1.3 |



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

# 8. Appendix B: Stakeholder views

## 8.1. Practitioner view

| No | Preliminary Recommendation | Related FAIR Principle |
|---|---|---|
| 1 | Use Globally Unique, Persistent and Resolvable Identifier for Semantic Artefacts and their content | F1 |
| 2 | Use Globally Unique, Persistent and Resolvable Identifier for Semantic Artefact Metadata Record | F1 |
| 3 | Use a common minimum metadata schema to describe semantic artefacts and their content | F2 |
| 4 | Publish the Semantic Artefact and its content in a semantic repository | A1 |
| 9 | Semantic artefacts should be compliant with Semantic Web and Linked Data standards | I1 |
| 10 | Use a Foundational Ontology to align semantic artefacts | I |
| 12 | Semantic mappings should use machine-readable formats based on W3C standards | I & R |
| 13 | Crosswalks, mappings and bridging between semantic artefacts should be documented, published and curated | I & R |
| 14 | Use standard vocabularies to describe semantic artefacts | I2 |
| 15 | Make the references to the reused third-party semantic artefacts explicit | I3 |
| 16 | The semantic artefact should be clearly licenced for machines and humans | R1.1 |
| 17 | Provenance should be clear for both humans and machines | R.2 |
| | **Suggested Best Practice** | |
| 1 | Use a unique naming convention for concept/class and relations | |
| 3 | Use defined ontology design patterns | |
| 4 | Create mappings validated by domain experts | |
| 5 | Define workflows between different formats | |
| 7 | Interact with the designated community and manage user-centric development | |
| 8 | Provide a structured definition for each concept | |



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

| | | |
|---|---|---|
| 9 | The semantics of semantic artefacts should be accurately grounded on the domain it intends to describe | |
| 10 | Define a set of governance policies for the semantic artefacts | |
| 11 | Use TRUSTed and FAIR compliant repositories to persist Semantic Artefacts | |
| 12 | Semantic Artefacts, if developed with public funding, should be published with open licenses unless one of the overriding conditions discussed below are true. | |
| 13 | A standard architecture for semantic artefact management, services for content and metadata, protocols, and serialisations/ content negotiation should be proposed, and semantic repositories are urged to use this architecture as a design pattern. | |

## 8.2. Repository view

| No | Preliminary Recommendation | Related FAIR Principle |
|---|---|---|
| 1 | Rec. 1: Use Globally Unique, Persistent and Resolvable Identifier for Semantic Artefacts and their content | F1 |
| 2 | Use Globally Unique, Persistent and Resolvable Identifier for Semantic Artefact Metadata Record | F1 |
| 3 | Use a common minimum metadata schema to describe semantic artefacts and their content | F2 |
| 5 | Semantic repositories should offer a common API to access Semantic Artefacts and their content in various serializations for both use/reuse and indexation by search engines | A1, A1.1 |
| 7 | Repositories should offer a secure protocol and user access control functionalities | A1.2 |
| 8 | Define human and machine-readable persistency policies for metadata | A2 |
| 9 | Semantic artefacts should be compliant with Semantic Web and Linked Data standards | I1 |
| 13 | Crosswalks, mappings and bridging between semantic artefacts should be documented, published and curated | I & R |
| 16 | The semantic artefact should be clearly licenced for machines and humans | R1.1 |
| | **Suggested Best Practice** | |
| 10 | Define a set of governance policies for the semantic artefacts | |



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

| | | |
|---|---|---|
| 11 | Use TRUSTed and FAIR compliant repositories to persist Semantic Artefacts | |
| 12 | Semantic Artefacts, if developed with public funding, should be published with open licenses unless one of the overriding conditions discussed below are true. | |
| 13 | A standard architecture for semantic artefact management, services for content and metadata, protocols, and serialisations/ content negotiation should be proposed, and semantic repositories are urged to use this architecture as a design pattern. | |
| 14 | User registration and authentication should be as simple as possible and adhere to legislative requirements | |

## 8.3. Community view

| No | Preliminary Recommendation | Related FAIR Principle |
|---|---|---|
| 3 | Use a common minimum metadata schema to describe semantic artefacts and their content | F2 |
| 6 | Build semantic artefacts search engines that operate across different semantic repositories | F4 |
| 11 | Use a standardized description for complex logical relations | I |
| 14 | Use standard vocabularies to describe semantic artefacts | I2 |
| | **Suggested Best Practice** | |
| 1 | Use a unique naming convention for concept/class and relations | |
| 2 | Use an Ontology Naming Convention | |
| 5 | Define workflows between different formats | |
| 6 | Harmonize the methodologies used to develop semantic artefacts | |
| 11 | Use TRUSTed and FAIR compliant repositories to persist Semantic Artefacts | |
| 12 | Semantic Artefacts, if developed with public funding, should be published with open licenses unless one of the overriding conditions discussed below are true. | |
| 13 | A standard architecture for semantic artefact management, services for content and metadata, protocols, and serialisations/ content negotiation should be proposed, and semantic repositories are urged to use this architecture as a design pattern. | |



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

| | | |
|---|---|---|
| 14 | User registration and authentication should be as simple as possible and adhere to legislative requirements | |



FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558