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

D2.2 FAIR Semantics: First recommendations

Work Package	WP 2
Lead Author (Org)	Yann Le Franc (e-Science Data Factory)
Contributing Author(s) (Org)	Gerard Coen (DANS), Christine Staiger (DTL), Jessica Parland von Essen (CSC), Heikki Lehväslaiho (CSC), Luiz Bonino (LUMC - DTL)
Due Date	31.01.2020
Date	11.03.2020
Version	1.0, DRAFT NOT YET APPROVED BY THE EUROPEAN COMMISSION
DOI	https://doi.org/10.5281/zenodo.3707985

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 first iteration of recommendations for making semantic artefact FAIR. These recommendations result from initial discussions during a brainstorming workshop organised by FAIRsFAIR as co-located event with the 14th RDA Plenary meeting in Helsinki. We are proposing 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. These initial recommendations should not be considered as a gold standard but rather as a basis for discussion with the various stakeholders of the semantic community.

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
0.1	13.11.2019	Yann Le Franc (e-Science Data Factory)	Outline document
0.2	18.11.2019	Yann Le Franc (e-Science Data Factory)	Added text for introduction, methods and add recommendations generated during the Helsinki workshop.
0.3	21.11.2019	Gerard Coen (DANS)	Expanded recommendations sections and restructured the categories. Added information for stakeholders, implementation, adoption and quality of the recommendations.
0.35	2.12.2019	Jessica Parland-von Essen (CSC)	Added tables for recommendations with suggested recommendations for F,A, and I.
0.4	12.1.2020	Jessica Parland-von Essen (CSC)	Added suggested recommendations for R and cleaned sections F and R.
0.5	08.02.2020	Yann Le Franc (eSDF), Luiz Bonino (LUMC - DTL)	Evaluated individual recommendations, removed duplicates, separated FAIR related recommendations and "Best Practices"/Recommendations.
0.6	08.02.2020	Yann Le Franc (eSDF)	Write missing descriptions, reformulate erroneous descriptions, ...
0.7	08.02.2020	Yann Le Franc (eSDF)	Revise introduction, finalise recommendations, harmonise recommendation formatting, write conclusion, write executive summary, extend reference list, rework figure 1 and 2, create figure 3
0.8	09.02.2020	Yann Le Franc (eSDF)	Editorial changes
0.9	12.02.2020	Anusuriya Devaraju (UniHB), Ronald Siebes (DANS)	Feedback and comments as part of internal review process.
1.0	11/03/2020	Yann Le Franc (e-SDF), Luiz Bonino (LUMC - DTL)	Integration of comments and feedbacks from the reviewers, creation of the table in ANNEX A

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

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

ODRL	Open Digital Rights Language
OMV	Ontology Metadata Vocabulary
OWL	Web Ontology Language
PID	Persistent Identifier
PURL	Persistent Uniform Resource Locator
RDA	Research Data Alliance
RDF	Resource Description Framework
RDFS	Resource Description Framework Schema
Rec.	Recommendation
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

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, and thesauri, etc.) are key elements for the implementation of the FAIR principles as emphasized in the FAIR 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 usages of semantics in representing and describing information and knowledge. For this purpose, we are attempting to establish guidelines for practitioners and any related stakeholders. To ground these recommendations with 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 the outcomes of our internal work and our initial discussions with the community of semantic experts during the workshop entitled "Building the data landscape of the future: FAIR Semantics and FAIR Repositories", organised as a co-located event of the 14th RDA Plenary meeting in Helsinki (Oct 22, 2019). During the workshop, we gathered around 40 recommendations/requirements/requests for each of the FAIR aspects i.e. Findability, Accessibility, Interoperability and Reusability, as output from the brainstorming sessions with the participating experts. The main recommendations are summarized in Annex A.

We are presenting here two main outcomes resulting from our analysis of the discussions:

1. a set of preliminary recommendations (P-Rec) aligned with individual FAIR principles and
2. a list of best practices recommendations (BP-Rec) going beyond the 15 principles but contributing to improve the FAIRness of semantic artefacts.

For each recommendation, we are providing a short description offering some context and explanations to support further discussion. In addition, we are providing pointers to recommendations from existing initiatives and communities of practice that were known to us at the time of the writing. Our goal is to support existing efforts and investigate potential overlaps and gaps between these different initiatives to offer a harmonised framework for FAIR Semantics.
Finally, for each of the recommendations we are considering the stakeholder(s) they will impact.

This first version of the recommendations have not been developed to be a gold standard but rather to trigger further discussions with the semantic expert community. Therefore, these recommendations should be criticized, improved, refined and extended if necessary by the semantic community at large. These discussions should also include prioritization and definition of implementation strategies. To support the interaction with the community,
For this purpose, we will release the recommendation on another digital media allowing for large scale collaboration.

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

Our goal is to reach a mature version of these recommendations to be used to evaluate the FAIRness of semantic artefacts based on the approach developed by M. Wilkinson et al. (Wilkinson et al., 2019). These evaluations should be used as an indicator for improvement of resources and ideally guidelines provided to the practitioner to reach higher levels of FAIRness.

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

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

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

Glossary

Semantic artefact: 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 set of terms, taxonomies, thesauri to higher-order logics. Moreover, semantic artefacts are serialised using a variety of digital representation formats, e.g., RDF Turtle, OWL-RDF, XML, JSON-LD.

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 human and for machines.

Term/class/concept: a term/class/concept is an individual element with a unique semantic interpretation, represented with a unique identifier.

Glossary: a glossary is an alphabetical list of terms in a particular domain of knowledge with the definitions for those terms. - Source: Wikipedia[1]

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. - Source: newworldencyclopedia.org[2]

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).

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 that some or all terms have additional metadata such as scope notes (brief explanations of how the term should be used in indexing) or history notes.

Ontology: an ontology is a formal version of a thesaurus where relations are described using a formal system like Description Logic (DL) to mathematically classify individuals of classes and properties

[1] https://en.wikipedia.org/wiki/Glossary

[2] https://www.newworldencyclopedia.org/entry/Controlled_vocabulary

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

Introduction and scope

The FAIR principles are a set of technology-agnostic guidelines to make digital assets Findable, Accessible, Interoperable, and Reusable (Wilkinson et al., 2016). Semantic artefacts (i.e. controlled vocabularies, thesauri, ontologies, etc.) are machine readable models of knowledge.

They are used to facilitate the extraction and representation of knowledge within data sets using annotations or assertions. These annotations and assertions enable both integration and data retrieval. Both artefacts and services that support and offer them play an increasing role in the implementation of the FAIR principles (Principle I2 - (Meta)data use vocabularies that follow the FAIR principles) and in building FAIR Scientific Knowledge Graphs[3]. 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"

TFiR, pp. 42

According to this expert group, semantic artefacts and registries have been developed within almost all scientific disciplines (TFiR, pp. 41). However they have often been build with different formats (SKOS, XML, RDF, OWL), different levels of complexity/expressiveness (codelists, reference data, controlled vocabularies, taxonomies, thesauri, ontologies, formal ontologies,...) 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). However, the emergence of semantic registries such as BARTOC[4], FAIRsharing[5] and repositories such as Bioportal[6] (Whetzel et al., 2011), EBI-OLS[7] (Jupp and al., 2015), Ontobee[8] (Ong et al., 2017), Research Vocabulary Australia[9], the NERC Vocabulary Service[10], Linked Open Vocabulary[11] and others provide means to improve discoverability and enable reusability. The importance of such semantic registries/repositories and the issue of the

[3] A FAIR Scientific Knowledge Graph expresses and links scientific contributions and related artefacts in a semantically rich FAIR graphical model.

[4] BARTOC https://bartoc.org/

[5] FAIRsharing https://fairsharing.org/

[6] Bioportal https://bioportal.bioontology.org/

[7] Ontology Lookup Service https://www.ebi.ac.uk/ols/index

[8] Ontobee: http://www.ontobee.org/

[9] Research Vocabularies Australia https://ardc.edu.au/services/research-vocabularies-australia/

[10] NERC Vocabulary Server https://www.bodc.ac.uk/resources/products/web_services/vocab/

[11] 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

findability of semantic artefacts is already being worked on by various groups such as the DCMI/NKOS Interest Group[12] 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 knowledge engineers 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[13] for the biomedical domain or the recent Industry Ontology Foundry[14]. For this purpose, we are organising dedicated workshops to gather a large audience. These initial recommendations were proposed by experts during our first brainstorming session organised as a workshop co-located with RDA Plenary 14 in Helsinki. Following the publication of this report we will release these recommendations to the communities to gather more feedback.

[12] DCMI/NKOS Interest Group https://dublincore.org/groups/nkos/

[13] OBO Foundry http://www.obofoundry.org/

[14] 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

What do we mean with 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 practices. This ambiguity of the concept “Ontology” has been discussed 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 distanciate ourselves from this debate by proposing and using another 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. 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 are sharing 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

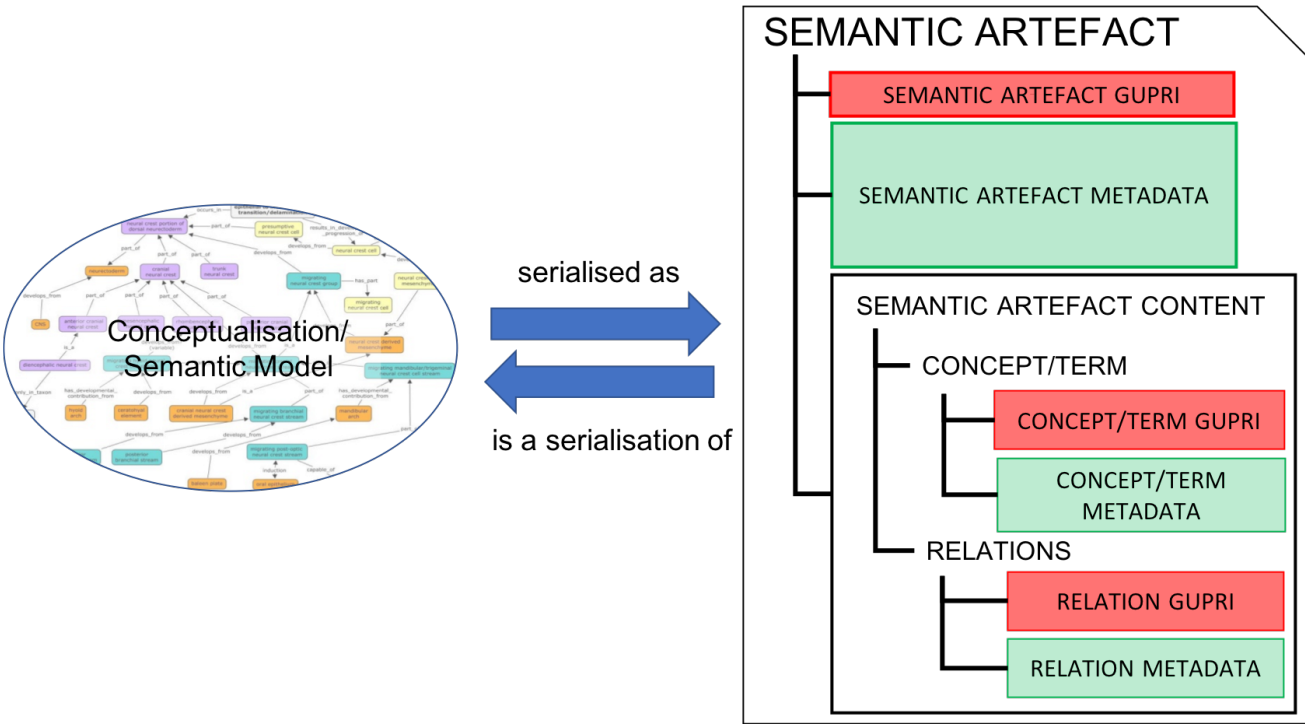

Figure 1: Common structure of semantic artefacts

Semantic artefacts are often structured text files. They have a common structure encapsulating a GUPRI 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 come with different formats and 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

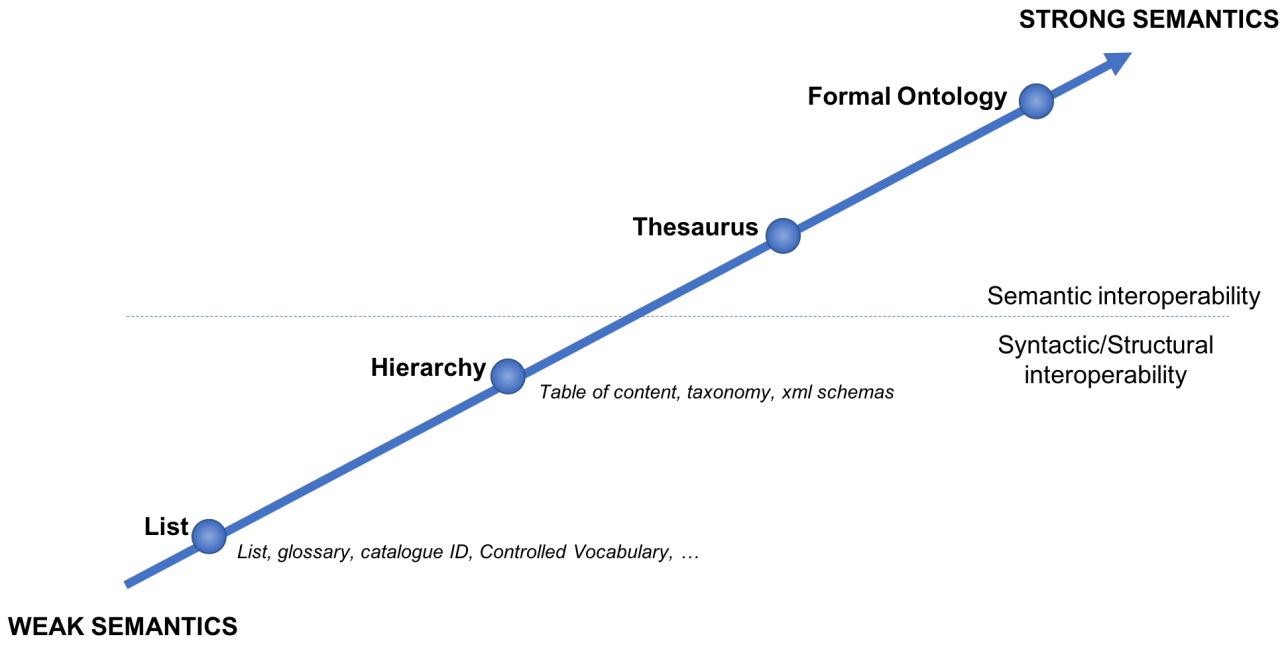

Figure 2: Semantic artefact spectrum. Derived from Leo Obrst, 2010

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

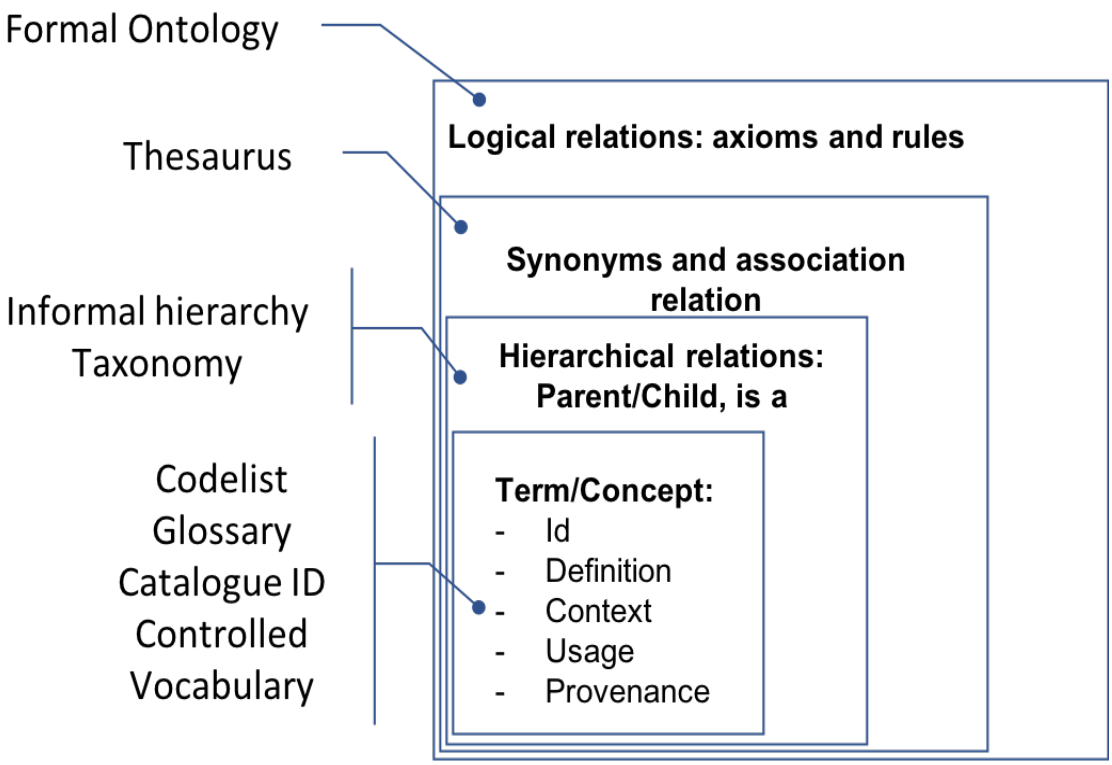

Figure 3: From list to formal ontology: a transformation path.

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[15], RDFs[16], OWL[17], SKOS[18]...) and serialization format (XML[19], XML Schema[20], JSON[21], RDF/XML[22], OWL/XML[23], Manchester Syntax[24], JSON-LD[25], Turtle[26], N-Triples[27]…) have been proposed by W3C . For simple models, common formats such as XML, XML Schema and JSON are usually used. 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

[15] RDF https://www.w3.org/TR/rdf11-concepts/
[16] RDFs https://www.w3.org/TR/2014/REC-rdf-schema-20140225/
[17] OWL https://www.w3.org/OWL/
[18] SKOS https://www.w3.org/TR/skos-primer/
[19] XML https://www.w3.org/TR/xml/
[20] XML Schema https://www.w3.org/TR/xmlschema-0/
[21] JSON https://tools.ietf.org/html/rfc8259
[22] RDF/XML https://www.w3.org/TR/rdf-syntax-grammar/
[23] OWL/XML https://www.w3.org/TR/2012/REC-owl2-xml-serialization-20121211/
[24] OWL Manchester Syntax https://www.w3.org/TR/2012/NOTE-owl2-manchester-syntax-20121211/
[25] JSON-LD https://www.w3.org/TR/json-ld11/
[26] Turtle https://www.w3.org/TR/turtle/
[27] 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

describing information as a very simple graph-oriented data schema. Based on 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)[28]. This standard, less formal and constrained that RDF and OWL, is used to build knowledge organisation systems (KOS). 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 are 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
Hierarchical list	XML schema, RDF, SKOS
Thesaurus	RDF/RDFs, SKOS
Formal ontology	OWL, OntoUML, FOL, Modal logic

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

[28] 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

What do we mean with 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 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 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 mappings 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 included the following recommendations:

- OBO Foundry (Smith et al., 2007)
- Industry Ontology Foundry (Kulvatunyou et al., 2018)
- Agrisemantic Working Group recommendations: 39 Hints to Facilitate the Use of Semantics for Data on Agriculture and Nutrition[29].
- MOD (Dutta et al., 2017)
- OMV (Hartmann et al., 2005)
- MIRO (Matentzoglu et al., 2018)
- MIREOT (Courtot et al., 2009)
- Linked Open Vocabulary (Vandenbussche et al., 2017)

Community experts are invited to contribute to extend our view of community practices by adding any missing recommendation. This report documents the alignment between community specific recommendations, which will be further improved in the later iteration of the document.

[29] 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

Who should be involved and impacted by these recommendations?

Our end goal 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 toward 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 model for semantic interoperability.

Therefore, for this initial phase of the work, we are considering three main stakeholders:

1. Practitioners i.e. knowledge engineers, ontologists who are building and maintaining semantic artefacts
2. Repositories i.e. development team and curators of community specific semantic repositories
3. Semantic Web Community i.e. all the stakeholders from all the domains.

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

These recommendations are preliminary. We are inviting all the interested stakeholders to join our effort and contribute to establish common guidelines for harmonizing the semantic landscape.

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

Recommendations to make FAIR Semantic artefacts

In this section, we are providing a plain list of individual preliminary recommendations (P-Rec; X). These recommendations are 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. We have come up with the 17 preliminary principles presented below. They are each enriched with a description providing some context and whenever possible existing recommendations. In addition, we have been considering the FAIR principle(s) fulfilled by such recommendation and we also consider who was impacted/responsible for such recommendation. The recommendations that could not be directly aligned to the FAIR principles were then aggregated into a set of suggested best practices presented in the next section. These Best Practice Recommendations are not directly linked to a particular principle but contribute to improve the overall FAIRness.

Individual recommendations are represented as tables in which the top row contains the ID of the recommendation, the recommendation and the associated FAIR principles in a dedicated cell.

The other rows contains as follows:

1. Description
2. Existing Recommendations/ Existing technologies
3. Stakeholder

P-Rec. 1: Use Globally Unique, Persistent and Resolvable Identifier for Semantic Artefacts, their content and their versions	F1
Description Semantic artefacts are typically structured text files. They are *de facto* digital objects 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). As shown in fig. 1, semantic artefacts are composite digital objects 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). Finally, semantic artefacts are living digital objects by nature, evolving over time. Another specific GUPRI should be added to track the different versions of semantic artefacts allowing to get the latest version but also to have access to previous version in use in existing information systems.	

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 Web-based documents, semantic artefacts are usually identified by globally unique (i.e. two different files cannot have the same identifier) and resolvable identifiers. In the scope of WWW, usually semantic artefacts are represented by two key URIs: the URI pointing to the file and the URI namespace of the semantic artefact. As an example, consider a semantic artefact hosted on github which has a local namespace that points to the content of the artefact. This goes against the principle of uniqueness of the identifier. To solve this issue, the namespace and the file URI can be joined through HTTP redirects. This doesn't address the issue of persistence. To cope with these issues, the Web community developed the concept of Persistent URLs[30] and implemented dedicated servers guaranteeing the persistence of the URL and any associated necessary HTTP redirects. The value of this approach has been identified in the Biomedical domain by the OBO foundry which explicitly recommends the usage of PURL for identifying semantic artefacts within its ID policy (see Existing recommendations). However it has demonstrated a limitation when in 2016 the central PURL server has been stopped due to lack of funding. Fortunately the system has been integrated into a more perenne organisation, the Internet Archive[31]**.** Finally, the Industrial Ontology Foundry recommends using IRI (enabling the use of Unicode for defining web addresses) that are registered in their system.

Another alternative to implement GUPRIs is the use of Persistent IDs based on the handle system[32]. The handle is a Web-based identification system using a prefix which identifies a "naming authority" and a suffix which gives the "local name" of a resource that can be resolved through a handle server which will provide direct access to the associated metadata through a redirect to the landing page corresponding to the record for human consumption. This approach is currently being investigated and promoted through the scientific data community (RDA, EOSC, …). A particular kind of handle i.e. the DOI could be used to identify a particular which should support citations (see Rec. 17). However, a limitation of the DOI is that it only refers to the landing page which represents a dead-end for machines. One of the limitations of the PIDs compared to URL/URI is the lack of control of the practitioners. PID are attributed by international organisations which require you to pay a fee for minting new PIDs. In a sense this business model allows to foster the perennity of the Ids. However, it does require to use a dedicated service to mint and affect new PIDs.

As discussed as introduction, these identification systems should apply to the semantic artefact but also to its content. Indeed, semantic artefacts can be considered as datasets of concepts and relations. Therefore, in this context, each element of the semantic artefact should also have an associated GUPRI. Both OBO Foundry and Industry Ontology Foundry are proposing to use special conventions to define URI based identifiers (see BP-Rec. 1 and BP-Rec 2).

Finally, a unified identifier schema should 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.

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

[30] PURL https://www.oclc.org/research/themes/data-science/purl.html

[31] Internet Archive https://archive.org/

[32] Handle System https://en.wikipedia.org/wiki/Handle_System

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

Related recommendations

- W3C Data on the Web - Best Practice 9: Use persistent URIs as identifiers of datasets namespace[33]
- OBO Foundry - Principle 3[34]
- OBO Foundry - Identifier Policy[35]
- OBO Foundry - Principle 4[36]
- 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)

Stakeholders: Practitioner and Repository

P-Rec. 2: Use Globally Unique, Persistent and Resolvable Identifier for Semantic Artefact Metadata Record	F1, F3

Description

Semantic artefacts are 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 should be parsed to access the content including the metadata. These practices contribute to the lack of findability by hiding the metadata to machines. For this purpose, it is necessary to consider publishing the ontology metadata separately allowing potential users to find it.
This metadata record should also have a Globally Unique, Persistent and Resolvable Identifier (FAIR principle F1) and an explicit reference to the semantic artefact it describes (FAIR principle F3). 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 and Repository

[33] W3C Data on the Web - Best Practice 9 https://www.w3.org/TR/dwbp/#DataIdentifiers
[34] OBO Foundry principle 3 http://www.obofoundry.org/principles/fp-003-uris.html
[35] OBO Foundry ID policy http://www.obofoundry.org/id-policy
[36] 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

P-Rec. 3: Use a common minimum metadata schema to describe semantic artefacts and their content	F2, R1.1, R1.2 and R1.3

Description

As any type of data, semantic artefact should be described by different levels of metadata to allow users to retrieve them and to understand their content. In particular, 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.

This metadata should be available in different formats and should be accessible for harvesting by search engines and metadata aggregators. Unfortunately, there is currently no consensus on a common set of metadata elements to describe semantic artefact. Several initiatives are proposing their recommendations such as OBO Foundry and IOF and several metadata schemata have been developed such as LOV (Vandenbussche et al., 2017), Ontology Metadata Vocabulary (OMV)[37], Metadata for Ontology Description and Publication Ontology (MOD),... (see list of related recommendations below). However, the heterogeneity of these metadata schema hampers indexing, retrieval as well as reuse of the semantic artefacts.

As for semantic artefact themselves, the concept/term 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, …

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-Rec. 8).

This recommendation emphasizes the need for the semantic web community to define a common minimal metadata schema that practitioners could use to describe semantic artefacts.

Existing recommendations:

- OBO Foundry - Principle 8 Documentation[38]
- OBO Foundry - Principle 5 Scope[39]
- OBO Foundry - Principle 6 Textual definition[40]

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

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

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

[40] OBO Foundry principle 6 http://www.obofoundry.org/principles/fp-006-textual-definitions.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

- Industry Ontology Foundry - Requirement 9 Documentation[41]
- Industry Ontology Foundry - Requirement 5 Scope[42]
- LOV - DCAT based metadata schema
- VOAF[43]
- Ontology Metadata Vocabulary[44]
- Metadata for Ontology Description and Publication Ontology[45]
- W3C Data on the web best practices - BP1, BP2 and BP3[46]
- Networked Knowledge Organization Systems Dublin Core Application Profile (NKOS AP)[47]

Stakeholders: Practitioner, Repository and Community

P-Rec. 4: Publish the Semantic Artefact and its content in a semantic repository	F4

Description

Semantic artefact can be made accessible on various supports such as github or even web pages. Unfortunately, most of the time semantic artefacts need to be downloaded and parsed in order to have access to its content i.e. 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). These repositories act as an archive, should provide GUPRIs, publish metadata making the semantic artefact findable for human through a dedicated User Interface and for machines through an API. 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] or Research Vocabulary Australia.

[41] IOF Technical Principles https://www.industrialontologies.org/?page_id=87
[42] IOF Technical Principles https://www.industrialontologies.org/?page_id=87
[43] VOAF https://lov.linkeddata.es/vocommons/voaf/
[44] Ontology Metadata Vocabulary http://omv2.sourceforge.net/
[45] MOD-Ontology https://github.com/sifrproject/MOD-Ontology
[46] Data on the Web Best Practices https://www.w3.org/TR/dwbp/
[47] NKOS AP https://nkos.slis.kent.edu/nkos-ap.html
[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/

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

This recommendation does not aim to support any particular technology but emphasize the necessity to share/publish semantic artefact with one of them for improving both findability and reuse.
Existing technologies: • SKOSMOS[52] • Bioportal[53] • Research Vocabularies Australia[54]
Stakeholder: Practitioner

P-Rec. 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 any search engines	F4, A1, A1.1

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 schemata 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 based on a common minimum set of metadata for describing semantic artefacts (see P-Rec. 3).

Enabling automated indexing across repositories will require agents to access a machine readable description of the API and the description of information that can be accessed. Therefore, repositories should consider publishing at least the description of their API using OpenAPI specifications which will provide first a human readable API documentation in a machine readable format. A recent extension of the OpenAPI specification, called smartAPI[55], has been proposed to provide semantically annotated API description to make 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

[52] SKOSMOS https://www.kansalliskirjasto.fi/en/services/system-platform-services/skosmos

[53] Bioportal https://bioportal.bioontology.org/;

[54] Research Vocabularies Australia https://vocabs.ands.org.au/

[55] 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

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. This recommendation does not aim to support any particular solution but rather emphasize the need for the community of semantic repositories to agree on a common solution.
Existing recommendations: ● The FAIR Data point specification[56] ● Registry Interchange Format - Collections and Services (RIF-CS) (ISO 2146)[57]. ● Linked Data Platform[58] ● OpenAPI[59] ● smartAPI[60]
Stakeholder: Repository

P- Rec. 6: Build semantic artefacts' search engines that operate across different semantic repositories	F4
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, we need semantic artefact search engines that can operate across different semantic repositories. These search engines will enable federated queries across the semantic artefacts and provide means to gather analytics across all ontologies (overlaps, mappings, reuse) and largescale automated mappings to resolve semantic ambiguity. The indexes supporting the search engines could also be directly integrated within the tooling to provide access to the existing resource at the time of the creation of a new semantic artefact for instance. This recommendation emphasizes that such service is an important element of the infrastructure to support FAIR data and FAIR Semantics.	
Existing recommendations: N/A	
Stakeholder: Community	

[56] FAIR Data point specification https://github.com/FAIRDataTeam/FAIRDataPoint-Spec
[57] RIF-CS https://vocabs.ands.org.au/viewById/1
[58] Linked Data Platform https://www.w3.org/TR/ldp/
[59] OpenAPI https://www.openapis.org/
[60] 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

P-Rec. 7: Repositories should offer a secure protocol and user access control functionalities	A1.2
Description Semantic artefacts should be openly shared to support reuse for avoiding concept redundancy and semantic ambiguities. However, semantic artefact might be developed under specific copyrights with paywalls (e.g. Dewey Decimal Classification) preventing direct access for use. Although the metadata describing the artefact should be made available, the access should be restricted with user access control functionalities and secured access to the content should be provided through a secure protocol such as HTTPS for machines.	
Existing recommendations: N/A	
Stakeholder: Repository	

P-Rec. 8: Define human and machine-readable persistency policies for semantic artefacts *metadata*	A2
Description Once published in a semantic repository, semantic artefacts are hopefully being reused by others to build their information system. In the eventuality where the semantic artefact, a concept/term or a relation is deprecated or simply replaced, the repository should offer persistency policies for the metadata (duration of archiving, …). These policies should be both humans and machines readable. Machine readable policies will allow building services which could automatically detect the change to either warn the user or to directly integrate the change whenever it is possible. For humans, repositories could use the classical tombstone page with redirect to the new page when the semantic artefact or the element has been replaced.	
Existing recommendations: N/A	
Stakeholder: Repository	

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

P-Rec. 9: Semantic artefacts should be represented using common serialization formats, e.g., Semantic Web and Linked Data standards	I1

Description

Semantic artefacts should be serialized in the formats developed in the context of the Semantic Web and Linked Data i.e. OWL, OBO, RDF, SKOS. Semantic repositories should provide a Linked Data compliant API to enable the creation of a semantic graph for analysis and reuse. 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 to define a common serialization to describe complex logical relations.

Existing recommendations:

- RDF and RDFS[61]
- W3C-OWL (OWL 2 / Full; OWL 2 / EL; OWL 2 / QL; OWL 2 / RL) stack[62]
- SKOS[63]
- The OBO Foundry Principle 2 "Common Format" requires the OWL file in RDF-XML format. Legacy formats are automatically converted to OWL.[64]
- Industry Ontology Foundry - requirement 3[65]
- OBO
- Linked Data Platform

Stakeholder: Practitioner and Repository

P-Rec. 10: Use a Foundational Ontology to align semantic artefacts	I1, I2, I3

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 of domain specific semantic artefacts acting as

[61] RDF and RDFS https://www.w3.org/TR/rdf11-primer/

[62] W3C-OWL https://www.w3.org/TR/owl2-profiles/

[63] SKOS https://www.w3.org/TR/skos-primer/

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

[65] 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 language bridging them. Several foundational ontologies exist such as UFO[66], BFO[67], DOLCE[68], EMMO[69]. This recommendation does not make any claim regarding which foundational ontology to use but emphasizes the necessity to be aligned with one.
Existing recommendations: ● Industry Ontology Foundry - requirement 8[70]
Stakeholder: Practitioner

P-Rec. 11: Use a standardized language for describing semantic artefacts	I1
Description 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 coverage impedes interoperability as complex semantic artefact have to be simplified which corresponds to a large loss of information and expressivity. To address this issue, the semantic web community should define an additional common language for high expressivity model representation as per FAIR principle I1.	
Existing recommendations: ● SHACL[71] ● SWRL[72] ● OntoUML[73]	
Stakeholder: Community	

[66] UFO https://ontouml.readthedocs.io/en/latest/intro/ufo.html
[67] BFO https://basic-formal-ontology.org/
[68] DOLCE http://www.loa.istc.cnr.it/dolce/overview.html
[69] EMMO https://emmc.info/taxonda/emmo-european-materials-modelling-ontology/
[70] IOF Technical Principles https://www.industrialontologies.org/?page_id=87
[71] SHACL https://www.w3.org/TR/shacl/
[72] SWRL https://www.w3.org/Submission/SWRL/
[73] OntoUML https://ontouml.readthedocs.io/en/latest/intro/ontouml.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

P-Rec. 12: Semantic mappings between the different elements of semantic artefacts should use machine-readable formats based on W3C standards	I1, I3, R1.3
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 duplication is often due to the lack of knowledge regarding existing semantic artefact. In order to aggregate distributed resources aligned with these different models, it is necessary to create mapping relations between the different elements of different semantic artefacts. In many cases, these mapping are based on existing relations such as sameAs from SKOS. However, mappings can become complex especially when considering the logical relations or to represent content drift is a potential risk 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. This recommendation aims at highlighting this gap and to emphasize the need for a machine readable description of mapping which fosters interoperability.	
Existing recommendations: ● **DOOR**[74]	
Stakeholder: Practitioner	

P-Rec. 13: Crosswalks, mappings and bridging between semantic artefacts should be documented, published and curated	R1.2, R1.3
Description Mappings, crosswalks and semantic bridges discussed in the previous recommendations (Rec. 13) should be made publicly available to allow the reuse by others. Mappings can be considered as semantic artefacts and therefore should be shared and published in semantic repositories. Sharing these resources in a standardised way will improve interoperability.	
Existing recommendations: N/A	

[74] DOOR http://oro.open.ac.uk/24326/1/keod9.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

Stakeholder: Practitioner and Repository

P-Rec. 14: Use standard vocabularies to describe semantic artefacts	I2
Description As stated in Rec. 3, the semantic artefact metadata is important for both findability and reusability. Standard vocabularies such as Dublin Core and FOAF for example, should be used to describe the semantic artefact. Agreeing on a common set of standards 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[75] ● FDP metadata scheme[76] ● OBO Foundry	
Stakeholder: Practitioner and Community	

P-Rec. 15: Make the references to the reused third-party semantic artefacts explicit	I3, R1.2
Description New semantic artefacts can be built upon existing artefacts. In many cases the reuse is made by copying and pasting bits and pieces of semantic artefacts together. This does not allow accessing automatically the reused elements and their semantic artefact. In order to be able track the element reused from other semantic artefacts, reference to third party semantic artefact should be made explicit. For this, external semantic artefact (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 reference, it becomes possible to extract this information automatically from the artefacts. This reference can be made just by providing a direct link to the used resource 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 even by considering the requirement of the Minimum Information to Reference an External OnTology, MIREOT (Courtot et al., 2009).	

[75] LOV https://lov.linkeddata.es/dataset/lov/

[76] FAIR Data point specification https://github.com/FAIRDataTeam/FAIRDataPoint/wiki/FAIR-Data-Point-Specification

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

This recommendation does not pretend to enforce a particular solution but rather emphasizes the need for the community of semantic web to define a common way of importing external semantic artefacts.
Existing recommendations: • LOV[77] • VoID[78] • MIREOT[79]
Stakeholder: Practitioner

P-Rec. 16: The semantic artefact should be clearly licenced for machines and humans	R1.1
Description Proper reuse of digital objects requires a licence. Well documented legal interoperability is a prerequisite for automatic distributed search and use of both the semantic artefacts and the individual terms/concepts and relations. Although we are encouraging Open Source types of 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.	
Existing recommendations: • Creative Commons licences[80] • ODRL[81]	
Stakeholder: Practitioner and Repository	

[77] LOV https://lov.linkeddata.es/dataset/lov/
[78] VoID https://www.w3.org/TR/void/
[79] MIREOT http://precedings.nature.com/documents/3574/version/1
[80] Creative Commons licences https://creativecommons.org in rdf https://github.com/creativecommons/cc.licenserdf
[81] ODRL https://www.w3.org/TR/odrl-model/

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

P-Rec. 17: Provenance should be clear for both humans and machines	R1.2

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 potential external users. This information can be thus used to evaluate the semantic artefact and understand the release cycle.

Provenance should be documented at an appropriate level of granularity to enable reuse of semantic artefacts and its constituting elements (class/term and relation). Provenance should 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 pid (data object) and creator pid) and the provenance should provide dates and lifecycle events. The provenance information should be described using an appropriate standard model such as PROV[82]. 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 the backward interoperability.

Existing recommendations:

- MIRO (Matentzoglu et al., 2018)
- OBO foundry - Principle 4[83]
- OBO Foundry - Principle 8[84]
- PROV

Stakeholder: Practitioner

Recommendations beyond the FAIR principles: best practices for semantic artefacts

In this section, we are listing recommendations that do not apply to a particular FAIR principle but contribute to the improvement of Findability, Accessibility, Interoperability and Reusability.
These recommendations should be used as a set of best practices.

[82] PROV data model https://www.w3.org/TR/prov-primer/

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

[84] 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

BP-Rec. 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 chain. This chain 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. 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)[85] ● Industry Ontology Foundry - requirement 11[86]
Stakeholder: Practitioner and Community

BP-Rec. 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 global consensus on the naming of semantic artefact and the use of acronyms/prefix 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:

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

[86] 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

- OBO Foundry - Principle 3[87]
- Industry Ontology Foundry[88]

Stakeholder: Community

BP-Rec. 3: Use defined ontology design patterns

Description

To foster interoperability, semantic artefacts should be designed with defined design patterns whenever relevant and possible. These patterns should be documented and published as a resource for practitioners following the example of OntologyDesignPatterns.org[89] that focuses on OWL design patterns.

Existing recommendations: N/A

Stakeholder: Practitioner

BP-Rec. 4: Create mappings validated 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 be generated by knowledge experts and validated by domain experts. As stated in P-Rec. 12 & P-Rec. 13, these mappings should be serialised with standard formats and be published.

Existing recommendations: N/A

Stakeholder: Practitioner

[87] OBO Foundry principle 3 http://www.obofoundry.org/principles/fp-003-uris.html
[88] IOF Technical Principles https://www.industrialontologies.org/?page_id=87
[89] 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

BP-Rec. 5: Define workflows between different formats
Description Semantic artefact can be serialized in various formats (SKOS, RDF, OWL, XML, …). This diversity of format makes it complicated to integrate and work with heterogeneous semantic artefacts. Practitioners should describe the particular workflow 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
Stakeholder: Practitioner, Community

BP-Rec. 6: Harmonize the methodologies used to develop semantic artefacts
Description Semantic artefacts can be built with very different methodologies depending on the available resources and the expertise of the practitioner. These methodologies should be documented to allow external experts willing to reuse the artefact to assess the quality of the semantic artefact. To support interoperability and prevent ill-formed ontologies, the community of practice should work toward harmonizing and documenting the various methodologies which could be used as a training resource for newcomers and guidelines for expert practitioners.
Existing recommendations: ● OBO Foundry - Principle 8[90] ● Industry Ontology Foundry - Principle 9[91]
Stakeholder: Community

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

[91] 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-Rec. 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 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 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[92]
- OBO Foundry - Principle 11 Authority[93]
- OBO Foundry - Principle 16 Maintenance[94]
- IOF Principle 16[95]
- IOF Principle 15
- IOF Principle 14

Stakeholder: Practitioner

BP-Rec. 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.

[92] OBO Foundry principle 9 http://www.obofoundry.org/principles/fp-009-users.html
[93] OBO Foundry principle 11 http://www.obofoundry.org/principles/fp-011-locus-of-authority.html
[94] OBO Foundry principle 16 http://www.obofoundry.org/principles/fp-016-maintenance.html
[95] 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

There are already quite few recommendations providing guidelines for writing human readable definition as OBO Foundry principle 6 and the IOF principle 10. A set of 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[96].

Existing recommendations:

- OBO Foundry - Principle 6[97]
- IOF - Principle 10[98]
- Guidelines for writing definition in ontologies[99]

Stakeholder: Practitioner

BP-Rec. 9: The underlying logic of semantic artefacts should be grounded on the domain it intends to describe

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.
Whenever the semantic artefact is reused, it should be extended in granularity depending on use-case. This implies that the re-user should adhere to the same design principle.

Existing recommendations: N/A

Stakeholder: Practitioner

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

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

[98] IOF Technical Principles https://www.industrialontologies.org/?page_id=87

[99] 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

BP-Rec. 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
Stakeholder: Practitioner and Repository

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

Summary

In this section, we are summarizing the various recommendations into a unique table and we are evaluating the coverage of our recommendations with the FAIR principles.

Recommendations summary

Preliminary Recommendations		
Rec. #	**Recommendation**	**FAIR Principles**
P-Rec. 1	Use Globally Unique, Persistent and Resolvable Identifier for Semantic Artefacts, their content and their versions	F1
P-Rec. 2	Use Globally Unique, Persistent and Resolvable Identifier for Semantic Artefact Metadata Record	F1, F3
P-Rec. 3	Use a common minimum metadata schema to describe semantic artefacts and their content	F2, R1.1, R1.2 and R1.3
P-Rec. 4	Publish the Semantic Artefact and its content in a semantic repository	F4
P-Rec.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	F4, A1, A1.1
P- Rec. 6	Build semantic artefacts' search engines that operate across different semantic repositories	F4
P-Rec. 7	Repositories should offer a secure protocol and user access control functionalities	A1.2
P-Rec. 8	Define human and machine-readable persistency policies for metadata	A2
P-Rec. 9	Semantic artefacts should be compliant with Semantic Web and Linked Data standards	I1
P-Rec. 10	Use a Foundational Ontology to align semantic artefacts	I1, I2, I3
P-Rec. 11	Use a standardized description for complex logical relations	I1
P-Rec. 12	Semantic mappings should use machine-readable formats based on W3C standards	I1, I3, R1.3
P-Rec. 13	Crosswalks, mappings and bridging between semantic artefacts should be documented, published and curated	R1.2, 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

P-Rec. 14	Use standard vocabularies to describe semantic artefacts	I2
P-Rec. 15	Make the references to the reused third-party semantic artefacts explicit	I3, R1.2
P-Rec. 16	The semantic artefact should be clearly licenced for machines and humans	R1.1
P-Rec. 17	Provenance should be clear for both humans and machines	R1.2
Best Practices Recommendations		
BP-Rec. 1	Use a unique naming convention for concept/class and relations	
BP-Rec. 2	Use an Ontology Naming Convention	
BP-Rec. 3	Use defined ontology design patterns	
BP-Rec. 4	Create mappings validated by domain experts	
BP-Rec. 5	Define workflows between different formats	
BP-Rec. 6	Harmonize the methodologies used to develop semantic artefacts	
BP-Rec. 7	Interact with the designated community and manage user-centric development	
BP-Rec. 8	Provide a structured definition for each concept	
BP-Rec. 9	The underlying logic of semantic artefacts should be accurately grounded on the domain it intends to describe	
BP-Rec. 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

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

Conclusions

This document presents the first set of the FAIR Semantics preliminary recommendations and good practices.

An approach proposing a monolithic and static set of recommendations would clearly fail to capture the diversity of the semantic usages. Therefore, in this project we are considering using a more agile approach which will allow the modification, extension and update of the recommendations based on expert feedback. Recommendations should be benchmarked in a different context and should be updated accordingly.

This document is the result of our first iteration with the community experts and it can be definitely improved. This first version of the recommendations have been built as a basis for discussion with the semantic community at large.

To improve and evaluate these recommendations, we have three main objectives :

1. gather and document use-cases which would allow us to test the recommendations,
2. adapt and use existing tools (such as FAIR Metrics[100]) for evaluating the FAIRness of a large number of semantic artefacts to establish a baseline,
3. align our work with the outcomes of the RDA FAIR Maturity Working Group (FAIR Maturity Indicators).

For this purpose, we will initiate collaborations with stakeholders by:

1. Organising at least 2 workshops/Hackathons in 2020,
2. Starting an open consultation to the international community of semantic practitioners by publishing the recommendations on a web-based platform allowing collaborative work and
3. Initiating discussions within existing international and pluri-disciplinary networks/initiatives such as the Vocabulary Services & Semantics Interest Group (RDA), the GO INTER implementation network[101] (GO FAIR)

From these interactions, we expect to establish a framework to evaluate and test our recommendations, to test the FAIRness of semantic artefacts and to extend the list of recommendations and improve the existing ones.

[100] FAIR Metrics https://github.com/FAIRMetrics/Metrics

[101] 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

References

Courtot, M., Gibson, F., Lister, A.L., Malone, J., Schober, D., Brinkman, R.R., Ruttenberg, A., 2009. MIREOT: the Minimum Information to Reference an External Ontology Term. Nat. Preced. https://doi.org/10.1038/npre.2009.3574.1r

d'Aquin, M., Noy, N.F., 2012. Where to publish and find ontologies? A survey of ontology libraries. J. Web Semant. 11, 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.

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. Knowl. Acquis. 5, 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, M.C., 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, A.S., Weigel, T., Wittenburg, P., 2019. Initial Persistent Identifier (PID) policy for the European Open Science Cloud (EOSC). https://doi.org/10.5281/zenodo.3574203

Kulvatunyou, B., Wallace, E.K., 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. J. Biomed. Semant. 9, 6. https://doi.org/10.1186/s13326-017-0172-7

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

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

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 Res. 45, D347–D352. https://doi.org/10.1093/nar/gkw918

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

Seppälä, S., Ruttenberg, A., Smith, B., 2017. Guidelines for writing definitions in ontologies. Ciênc. Informação 46.

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

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

Whetzel, P.L., Noy, N.F., Shah, N.H., Alexander, P.R., Nyulas, C., Tudorache, T., Musen, M.A., 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 Res. 39, W541–W545. https://doi.org/10.1093/nar/gkr469

Wilkinson, M.D., Dumontier, M., Aalbersberg, Ij.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., 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. Sci. Data 3, 160018. https://doi.org/10.1038/sdata.2016.18

Wilkinson, M.D., Dumontier, M., Sansone, S.-A., Bonino da Silva Santos, L.O., 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. Sci. 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

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 lead to about 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 i. Check the schema used by Taxonda 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 i. Usage information ii. Consistency checks iii. Coverage	F2
Concepts/Terms alignment between terminologies	F2
Metadata about an ontology should also include the identifier of the ontology	F3

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

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
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

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

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
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

ANNEX B: Stakeholder views

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	

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	Interact with the designated community and manage user-centric development	
8	Provide a structured definition for each concept	
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	

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

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	

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