

Project Title	Expanding FAIR solutions across EOSC
Project Acronym	FAIR-IMPACT
Grant Agreement No.	101057344
Start Date of Project	2022-06-01
Duration of Project	36 months
Project Website	https://fair-impact.eu/

[Version for community input]

M5.3 - Semantic artefact assessment methodology

Work Package	WP5 - Metrics, certification, and guidelines
Lead Author (Org)	Daniel Garijo, María Poveda-Villalón (UPM)
Contributing Author(s) (Org)	Pascal Flohr (KNAW-DANS), Alejandra Gonzalez-Beltran (STFC-UKRI), Yann le Franc (e-SDF), Maaike Verburg (KNAW-DANS)
Due Date	2023-08-31
Date	2023-08-31
Version	V1.0

Dissemination Level

<input checked="" type="checkbox"/>	PU: Public
<input type="checkbox"/>	PP: Restricted to other programme participants (including the Commission)
<input type="checkbox"/>	RE: Restricted to a group specified by the consortium (including the Commission)
<input type="checkbox"/>	CO: Confidential, only for members of the consortium (including the Commission)



Versioning and contribution history

Version 1.0 of this Milestone can be found on Zenodo at:

Garijo, D., Poveda-Villalón, M., Flohr, P., Gonzalez-Beltran, A., le Franc, Y., & Verburg, M. (2023). M5.3 Semantic artefact assessment methodology (Versie 1). Zenodo. <https://doi.org/10.5281/zenodo.8305173>

This version of the document is intended to invite community input on this presentation of the methodology for semantic artefact assessment. Please provide your input through suggestions or comments. The authors of the document will review the input to advance their work further. Your input is much appreciated!

Disclaimer

FAIR-IMPACT has received funding from the European Commission's Horizon Europe funding programme for research and innovation programme under the Grant Agreement no. 101057344. 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.

Table of Contents

Versioning and contribution history	1
Table of Contents	2
List of figures	2
List of tables	3
Terminology	4
1. Introduction	5
2. Description of the milestone	6
2.1 Role of the milestone	6
2.1.1 Means of verification	7
3. Process followed	8
4. FAIR assessment methodology	11
4.1 Methodology overview	11
4.2 Methodology steps in detail	15
4.2.1 Check implementation language	18
4.2.2 Check ontology elements metadata	19
4.2.3 Check existing vocabulary reuse	20
4.2.4 Check ontology metadata	21
4.2.4.1 Metadata recommendations in the proposed FAIR assessment methodology	22
4.2.5 Check whether the ontology is available in registries	24
4.2.6 Check online publication	25
4.2.7 Check community-based best practices	26
4.3 FAIR Scoring mechanism(s)	27
5. Conclusions and next steps	29
References	30
Appendices	32
Appendix 1: FOOPS! and O'FAIRE definitions	32
Appendix 2: Table A.2: Semantic Artefact landscape metadata analysis	36

List of figures

Figure 1 - Schematic overview of the LOT methodology workflow	12
Figure 2 - Overview of the proposed FAIR assessment methodology	13

List of tables

Table 1 - Overview of the methodology steps for assessing FAIR semantic artefacts	13
Table 2 - Mapping existing guidelines to the FAIR principles	14
Table 3 - Mapping between the tests performed by FOOPS! and O'FAIRE, and their relation to each FAIR principle for the "Findable" category	16
Table 4 - Mapping between the tests performed by FOOPS! and O'FAIRE, and their relation to each FAIR principle for the "Accessible" category	16
Table 5 - Mapping between the tests performed by FOOPS! and O'FAIRE, and their relation to each FAIR principle for the "Interoperable" category	17
Table 6 - Mapping between the tests performed by FOOPS! and O'FAIRE, and their relation to each FAIR principle for the "Reusable" category	17
Table 7 - Guidelines and existing tool support for checking the implementation language of a semantic artefact	19
Table 8 - Guidelines and existing tool support for checking the metadata of the elements included in an ontology	19
Table 9 - Guidelines and existing tool support for assessing vocabulary reuse within a semantic artefact	20
Table 10 - Guidelines and existing tool support for assessing ontology metadata to support FAIR	21
Table 11 - Guidelines and existing tool support for checking if an ontology exists in external registries or community repositories	24
Table 12 - Guidelines and existing tool support for checking whether an ontology is available on the Web	25
Table 13 - Guidelines and existing tool support for checking community-based best practices within an ontology	26
Table A1.1 - FOOPS! Tests per principle	32
Table A1.2 - O'FAIRe Tests per principle	33
Table A2.1 - Results of the semantic artefact landscape analysis	36

Terminology

Terminology /Acronym	Description
DOI	Digital Object Identifier. Persistent identifier for digital objects.
FOOPS!	Ontology Pitfall Scanner for FAIR
HTTP(S)	Hypertext Transfer Protocol (Secure). For (secure) exchange over a computer network like the Internet.
ID	Identifier
IRI	Internationalised Resource Identifier
IOF	Industry Ontology Foundry
JSON-LD	JavaScript Object Notation Linked Data. RDF serialisation
LOT	Linked Open Terms methodology
LOV	Linked Open Vocabularies
MIRO	Minimum Information for the Reporting of an Ontology. Guidelines
MOD	Metadata for Ontology Description
N-Triples (N3)	RDF Serialisation
OBO Foundry	Open Biological and Biomedical Ontology Foundry
O'FAIRE	Ontology FAIRness Evaluator
OWL	Web Ontology Language
PID	Persistent Identifier
PURL	Persistent URL
RDF	Resource Description Framework
RDF/XML	RDF Serialisation
RDFS	Resource Description Framework Schema
SKOS	Simple Knowledge Organization System
SPARQL	SPARQL Protocol and RDF Query Language. RDF query language
Turtle (TTL)	RDF Serialisation
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
W3C	World Wide Web Consortium
XML	Extensible Markup Language

1. Introduction

Semantic artefacts (i.e., ontologies, vocabularies and SKOS taxonomies, among others) define the structure, guide the construction of, and help validate many existing Knowledge Graphs. In the last years, a number of guidelines have been proposed (Poveda-Villalón et al. 2020; Garijo and Poveda-Villalón 2020; Hugo et al. 2020; Le Franc et al., 2022; Xu et al. 2023) to align semantic artefact best practices against the Findable, Accessible, Interoperable and Reusable principles (FAIR principles) (Wilkinson et al. 2016). Based on these guidelines, new validators and assistants have been developed (Garijo et al. 2021; Amdouni et al. 2022a; 2022b) in order to guide users assessing their own semantic artefacts against the FAIR principles. However, different tests are based on different interpretations of the FAIR principles, resulting in different scores and checks for semantic artefacts. To the best of our knowledge, there is no generic methodology grouping the types of tests to perform in semantic artefacts, in order to map existing assessment efforts in a consistent manner.

In this document, we propose such a methodology. We do so by taking an ontology development perspective, dividing semantic artefacts into smaller parts (their code, content, ontology metadata, etc.) that can be individually assessed at different stages of their development process. We build on the Linked Open Terms (LOT) methodology (Poveda-Villalón et al. 2022), adding a “FAIR assessment” module, and, for each activity, we validate our approach by mapping to two existing semantic artefact FAIR assessment validators: FOOPS!¹ (Garijo et al. 2021) and O’FAIRe² (Amdouni et al. 2022a; 2022b).

The rest of the document outlines our methodology, describes each step in detail, and maps it to existing FAIR principles and guidelines.

¹ <https://w3id.org/foops/>

² <https://github.com/agroportal/fairness>

2. Description of the milestone

In this milestone, we propose a methodology to quantitatively assess *semantic artefacts*, 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 formalisations, from loose sets of terms, taxonomies, thesauri to higher-order logics (Le Franc et al. 2022). Semantic artefacts are formalised using a variety of representation formats (mostly RDF(S)³ and OWL⁴), and are serialised using W3C standards such as Turtle,⁵ RDF/XML,⁶ N-Triples,⁷ and JSON-LD.⁸

Since semantic artefacts may be used to refer to a number of resources described semantically (e.g., knowledge graphs) or representing semantics for a given domain (e.g., ontologies, vocabularies, terminologies, knowledge graphs, shapes, mappings, etc.), here we prioritise semantic artefacts that aim to define hierarchies and structure, organising knowledge. That is, the work developed for and reported in this milestone focuses on **RDFS vocabularies**, **OWL ontologies**, and **SKOS vocabularies** as we build upon methodologies for ontology engineering.

Our approach takes an ontology development perspective. For this reason, we build on an existing ontology engineering methodology used for the industry domain, Linked Open Terms (LOT) (Poveda-Villalón et al. 2022), extending it with FAIR assessment activities. As a result, our methodology supports ontology developers that aim to assess their semantic resources at any stage of development. Our methodology also supports non-ontology engineers who aim to assess an existing semantic artefact, once the development process has produced a stable version.

Our methodology consists of two main *activities* (pre-assessment and assessment), depending on the development stage a semantic artefact is in. Each activity is itself divided in a series of steps, which are described in detail below, relating them to concrete guidelines, the FAIR principle these guidelines address, and concrete tests performed by existing assessment tools to verify them (if any).

2.1 Role of the milestone

This milestone has been driven by three main objectives: 1) Guiding semantic artefact developers in the steps that need to be followed in order to assess the FAIRness of their results; 2) Providing a common framework for comparing existing semantic artefact

³ <https://www.w3.org/TR/rdf-schema/>

⁴ <https://www.w3.org/OWL/>; <https://www.w3.org/TR/owl2-overview/>

⁵ <https://www.w3.org/TR/turtle/>

⁶ <https://www.w3.org/TR/rdf-syntax-grammar/>

⁷ <https://www.w3.org/TR/n-triples/>

⁸ <https://json-ld.org/>; <https://www.w3.org/TR/json-ld11/>

assessment tools; and 3) Detecting gaps in current assessment tools, as well as promoting the discussion on the interpretation of different FAIR principles by existing guidelines.

The first version of the methodology presented in the milestone will be put into practice during the remainder of the project. In addition, it will be used during an initial gap analysis to inform the development of FAIR assessment tools during the duration of the project. Throughout the rest of the project, this output may be subject to updates and new versions to reflect new developments and insights.

2.1.1 Means of verification

The required means of verification for this Milestone is to have the methodology available. This document is the verification of the methodology being publicly available online.



3. Process followed

In order to define the FAIR assessment method for semantic artefacts, four lines of work have been followed:

1. Ontology development methodologies analysis

Existing methodologies and common practices were analysed to understand how ontology development activities are currently carried out and how existing methodologies may be extended to incorporate a FAIR assessment activity along their lifecycle.

On the one hand, according to the H2020 project OntoCommons deliverable “D4.2 - Methodological framework for ontology management” (Fernández-Izquierdo et al. 2021), the only ontology development methodology partly taking into account FAIR principles is the LOT methodology (Poveda-Villalón et al. 2022; see also Section 2). This is due to the fact that most of the methodologies for building ontologies were developed prior to the publication of the FAIR principles. On the other hand, it has been observed that the assessment of the FAIRness level of semantic artefacts is usually carried out once the artefact is published or indexed in a registry. In both cases, it means that the FAIRness is checked only after developing the artefact. This may be a consequence of the current capabilities of semantic artefact FAIR checkers, as they require either a URI or the registration of the ontology code in an ontology registry.

Having selected the LOT methodology as basis for ontology development and taking into account the current tool limitations, two new activities for semantic artefact FAIR assessment are proposed to be added to the LOT methodology. These activities are “FAIR ontology pre-assessment” and “FAIR ontology assessment” and are explained in detail in Section 4.

2. Common metadata analysis in existing ontologies

1961 unique OWL and RDFS ontologies and 587 SKOS vocabularies have been analysed in order to define a harmonised set of common metadata fields. The process was carried out in several subsequent steps, described in this section.

First, a search was conducted for ontologies and vocabularies in commonly used registries and repositories, namely: Linked Open Vocabularies (1495 ontology and vocabulary versions were found),⁹ Archivo (1750),¹⁰ Bioportal (976),¹¹ EcoPortal (23),¹² IndustryPortal (45),¹³ and MedPortal (54).¹⁴ These were downloaded together with their metadata. Additional

⁹ <https://lov.linkeddata.es/dataset/lov/>

¹⁰ <https://archivo.dbpedia.org>

¹¹ <https://bioportal.bioontology.org/>

¹² <https://ecportal.lifewatch.eu/>

¹³ <http://industryportal.enit.fr/>

¹⁴ <https://medportal.bmicc.cn/>

ontologies were found, and subsequently downloaded, by searching in w3id.org (687)¹⁵ and OnToology (160).¹⁶ All sites and APIs were accessed in January 2023. Since some ontologies did not resolve from their URIs, we retrieved their versioned contents from repositories when possible. As many ontology files were present in more than one registry (e.g., Archivo incorporates many LOV ontologies), the final set consisted of 2784 files (with their corresponding metadata).

Next, we extracted from each ontology only its identifier (URI) and main descriptions, and aggregated all the results in a single Knowledge Graph. This step removed duplicate URIs, conflating redundant vocabularies and resulted in **1961 ontologies and 587 SKOS vocabularies**. The results, scripts and Knowledge Graph are available online.¹⁷

Next, we counted the number of occurrences of a metadata term per vocabulary. For example, if an ontology defines three authors, the “author” property would only be counted once, in order to avoid over-representation. Subsequently, the occurrences of the metadata properties in the total Knowledge Graph were counted as well as the number of times they occur in each vocabulary or ontology. Three different annotators then manually aligned each property to MOD2.1 (Metadata for Ontology Description and Publication),¹⁸ linking them when possible to all the existing MOD categories (Garijo et al. 2023). This mapping is summarised in Table A2.1, grouping similar properties by MOD category and showing their overall support in the metadata analysis. Table A2.1 allows determining which metadata properties are currently supported in existing vocabularies and ontologies.

3. FAIR semantic artefacts best practices analysis

The FAIR principles have been adapted for semantic artefacts by the scientific community. In particular, we have assessed four different guidelines (Garijo and Poveda-Villalón 2020; Cox et al. 2021; Le Franc et al. 2022; Xu et al. 2023) together with the corresponding authors (Section 4, Table 2). Each step of our methodology maps to a FAIR principle, making it easy to find a correspondence for each best practice. This is detailed in Section 4.2.

4. FAIR semantic artefact assessment tool analysis

In order to illustrate the methodology steps with example implementations from existing tools for assessing ontology FAIRness (namely, FOOPS! and O’FAIRe), we have 1) compared the tests from each tool to assess potential alignments between them, and 2) indicated how each tool test addresses each step in our methodology.

This activity has been carried out within a collaboration between FOOPS! developers at the Universidad Politécnica de Madrid (UPM)¹⁹ and O’FAIRe developers at INRAE (the French

¹⁵ <https://github.com/perma-id/w3id.org>

¹⁶ <https://ontoology.linkeddata.es/>

¹⁷ https://github.com/dgarijo/ontology_metadata_landscape_analysis/tree/main/analysis_results

¹⁸ <https://w3id.org/mod/2.0> (GitHub: <https://github.com/FAIR-IMPACT/MOD>)

¹⁹ <https://www.upm.es/>

National Research Institute for Agriculture, Food and Environment).²⁰ The tests were then aligned with the recommendations given by the proposed methodology. In this way, users may look up which test from which tool they can use for each of the proposed recommendations, or whether there is no tool available and the validation should be done manually. Moreover, the guidelines of our methodology are linked to existing proposals and best practices adapting the FAIR principles to semantic artefacts (Garijo et al. 2020; Hugo et al. 2020; Le Franc et al. 2020; Poveda-Villalón et al. 2020; Cox et al., 2021; Le Franc et al. 2022;), described in Section 4.1.

²⁰ <https://www.inrae.fr/en>



4. FAIR assessment methodology

This section presents the proposed extension of the LOT methodology in order to include a semantic artefact FAIR assessment. Details about the guidelines proposed for each step are also provided.

4.1 Methodology overview

Figure 1 shows an overview of the LOT methodology with the suggested FAIR assessment extensions (green boxes). To extend the LOT methodology with the necessary activities to assess the FAIRness level of the semantic artefacts, we analysed what input is needed to check each FAIR principle and at which stage of the methodology that input is already available. Up until now, ontology developers analysed the FAIRness level once the ontology was publicly available in the Web, i.e., after the “Online publication” activity of the “Ontology publication” phase in Figure 1. The output of the latter activity is making the ontology available online and possibly indexed in ontology registries or repositories. For this reason, the new proposed activity “FAIR ontology assessment” is placed after the online publication activity. However, the input needed to check the FAIR principles may be available at earlier stages of the process. For this reason, a “FAIR ontology pre-assessment” activity is proposed to be carried out directly after the “ontology evaluation activity” during the implementation phase, as shown in Figure 1. The FAIR assessment can still be accomplished at the “FAIR ontology assessment” activity, while some parts can be carried out in advance during the “FAIR ontology pre-assessment” (every time a semantic artefact ready for evaluation is available). This is depicted in Figure 2, illustrating the steps that may be carried out in each activity. This proposed methodology is driven by the common ontology development processes and the resources generated during each activity. The proposed steps are defined according to the inputs needed to assess each FAIR principle and the aspects involved in that principle.

Different FAIR principles need different parts of the ontology, depending on what aspect is being assessed. After analysing all principles, the following required inputs have been identified:

- **Ontology code:** file(s) containing the ontology implementation including concepts, properties, individuals, and axioms (ontology content), and the ontology metadata.
 - **Ontology content:** the part of the ontology code containing the concepts, properties, individuals, and axioms.
 - **Ontology metadata:** the part of the ontology code containing the ontology metadata. This information is usually included in the ontology file, but it could be provided as an external resource.
- **Online ontology:** the ontology is published online through an accessible ontology URI. The ontology should follow content negotiation and be available in one or more RDF serialisations.



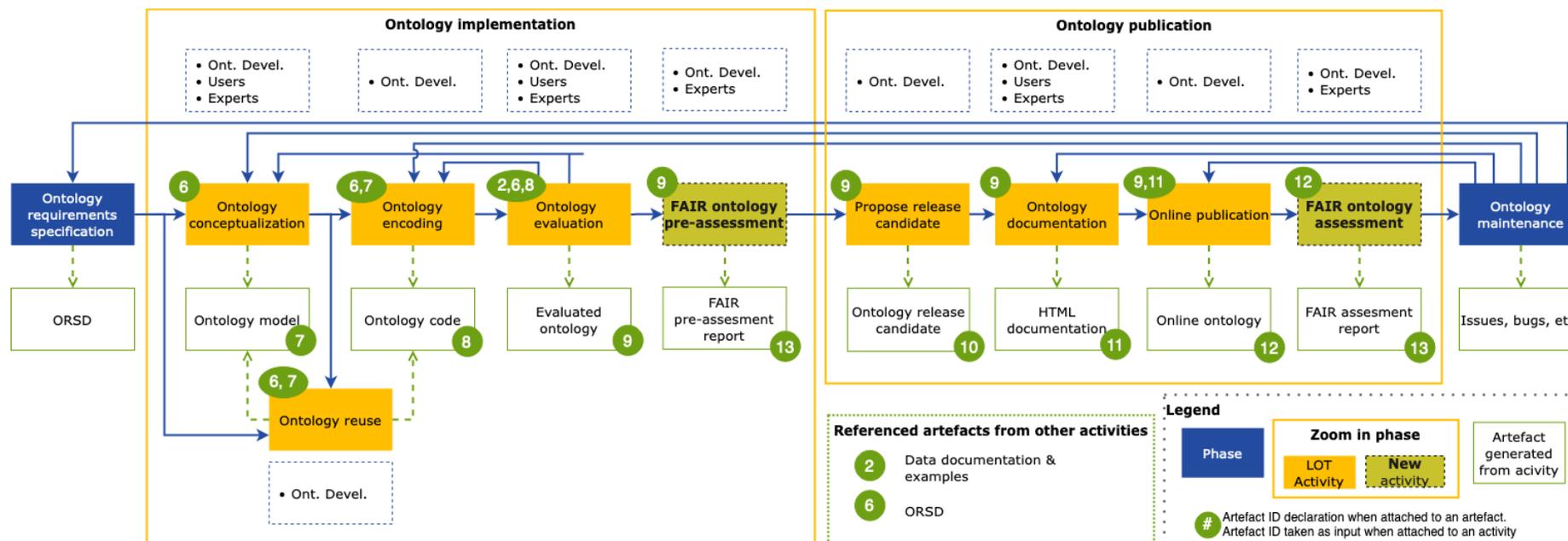


Figure 1 - Schematic overview of the LOT methodology workflow for the ontology implementation and ontology publication activities (blue and yellow boxes) with new FAIR assessment steps inserted (green boxes).

The proposed steps included in the methodology for FAIR assessment are depicted in Figure 2, including the input needed to carry out each step and the associated FAIR principle. This information is also summarised in Table 1. The step “Check community-based practices” is applicable to all the steps, as each community might have different common practices implementing ontologies (ontology code), or specific metadata fields (ontology metadata), or particular publication mechanisms (online ontology).

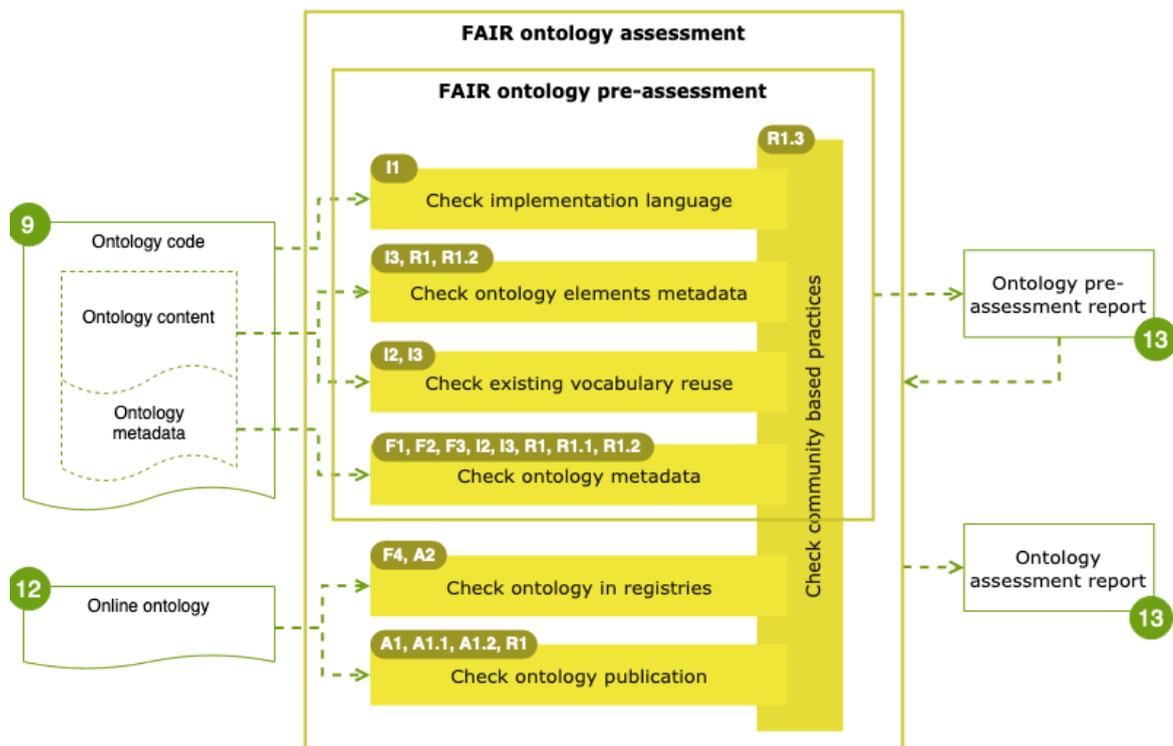


Figure 2 - Overview of the proposed FAIR assessment methodology. There are six main steps to be carried out for FAIR assessment when developing the ontology, with a transversal step depending on the best practices issued by specific communities. Four of the proposed activities may be carried out without the ontology being published online, and hence they are part of the pre-assessment activity. The two remaining activities (i.e., looking for an ontology in existing registries and checking the ontology publication) can only be addressed once a release of the ontology is available online (some registries require online availability to store an ontology).

Table 1 - Overview of the methodology steps for assessing FAIR semantic artefacts

Step	Input	Activity	FAIR principles
Check implementation language	Ontology code	<ul style="list-style-type: none"> ● pre-assessment ● assessment 	I1
Check ontology elements metadata	Ontology content	<ul style="list-style-type: none"> ● pre-assessment ● assessment 	I3, R1, R1.2
Check existing vocabulary reuse	Ontology content	<ul style="list-style-type: none"> ● pre-assessment 	I2, I3, R1.3

		<ul style="list-style-type: none"> assessment 	
Check ontology metadata	Ontology metadata	<ul style="list-style-type: none"> pre-assessment assessment 	F1, F2, F3, I2, I3 R1, R1.1, R1.2, R1.3
Check ontology in registries	Online ontology	<ul style="list-style-type: none"> assessment 	F4, A2
Check ontology publication	Online ontology	<ul style="list-style-type: none"> assessment 	A1, A1.1, A1.2, R1
Check community-based practices	Ontology code Ontology content Online ontology	<ul style="list-style-type: none"> pre-assessment assessment 	R1.3

Table 2 complements Table 1 by showing how existing guidelines for enabling FAIR semantic artefacts map to each FAIR principle. These guidelines have informed the assessment tools included to illustrate our methodology. The next section describes the proposed guidelines for each step in detail.

Table 2 - Mapping existing guidelines to the FAIR principles (guidelines: Garijo and Poveda-Villalón 2020, Poveda-Villalón et al. 2020 (Garijo & Poveda); Le Franc et al. 2022 (FAIRsFAIR); Cox et al. 2021 (10SimpleRules); Xu et al. 2023 'Features of a FAIR Vocabulary' (FVF)).

FAIR principle	Garijo & Poveda	FAIRsFAIR	10SimpleRules	FVF
F1. (meta)data are assigned a globally unique and persistent identifier	✓	✓	✓	✓
F2 data are described with rich metadata	✓	✓	✓	✓
F3 metadata clearly and explicitly include the identifier of the data it describes	✓	✓		
F4 (meta)data are registered or indexed in a searchable resource	✓	✓	✓	✓
A1 (meta)data are retrievable by their identifier using a standardized communications protocol	✓	✓	✓	✓
A1.1 the protocol is open, free, and universally implementable	✓	✓		✓
A1.2 the protocol allows for an authentication and authorization procedure, where necessary		✓		✓
A2 metadata are accessible, even when the data are no longer available	✓	✓		✓
I1 (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.	✓	✓	✓	✓

I2 (meta)data use vocabularies that follow FAIR principles		✓		
I3 (meta)data include qualified references to other (meta)data		✓	✓	✓
R1 meta(data) are richly described with a plurality of accurate and relevant attributes	✓	✓	✓	✓
R1.1 (meta)data are released with a clear and accessible data usage licence	✓	✓	✓	✓
R1.2 (meta)data are associated with detailed provenance	✓	✓	✓	✓
R1.3 (meta)data meet domain-relevant community standards		✓		✓

4.2 Methodology steps in detail

In this section we provide more insight into each of the proposed steps of our methodology for the pre-assessment and full assessment activities depicted in Figure 2. For each activity, our methodology proposes a series of guidelines, which map to each of the FAIR principles (and therefore to existing FAIR semantic artefact guidelines, as shown in Table 2).

We illustrate each guideline by mapping the tests defined by two FAIR semantic assessment tools: FOOPS! and O’FAIRE. Our rationale for choosing these tools is that, to the best of our knowledge, these are the most developed efforts focused on semantic artefact assessment to date.

While both tools perform semantic artefact assessment, they have been developed with different use cases in mind. FOOPS! takes as input an ontology URL (i.e., it assumes the ontology is publicly available somewhere in the Web), while O’FAIRE requires an ontology to be registered in a registry such as AgroPortal. Both tools interpret the FAIR principles with slight variations, which we have mapped in Table 3 to Table 6. Some of the O’FAIRE questions and FOOPS! tests have direct correspondence (e.g., the ontology URI is declared and is resolvable), while others partially overlap (e.g., minimum metadata fields, provenance metadata fields, etc.). A list with the definitions of each test/question is available in Appendices A1.1 and A1.2. Since O’FAIRE does not provide an identifier (ID) for each question, we have labelled them with the FAIR principle that is addressed and the respective question number.

Table 3 - Mapping between the tests performed by FOOPS! and O’FAIRE, and their relation to each FAIR principle for the “Findable” category.

Principle	O’FAIRE question	FOOPS! test
F1	F1.Q1	PURL1
	F1.Q2	
	F1.Q3	URI2
	F1.Q4	VER1
		URI1
		VER2
F2	F2.Q1	OM1
	F2.Q2	
	F2.Q3	OM1
F3	F3.Q1	
	F3.Q2	
	F3.Q3	
		FIND1
F4	F4.Q1	FIND2, FIND3
	F4.Q2	FIND2, FIND3
	F4.Q3	

Table 4 - Mapping between the tests performed by FOOPS! and O’FAIRE, and their relation to each FAIR principle for the “Accessible” category.

Principle	O’FAIRE question	FOOPS! test
A1	A1.Q1	CN1
	A1.Q2	
	A1.Q3	CN1
	A1.Q4	
A1.1	A1.1.Q1	HTTP1
	A1.1..Q2	HTTP1
	A1.1.Q3	
A1.2	A1.2.Q1	
	A1.2	
A2	A2.Q1	FIND_3_BIS
	A2.Q2	

Principle	O'FAIRe question	FOOPS! test
	A2.Q3	
	A2.Q4	

Table 5 - Mapping between the tests performed by FOOPS! and O'FAIRE, and their relation to each FAIR principle for the "Interoperable" category.

Principle	O'FAIRe question	FOOPS! test
I1	I1.Q1	RDF1
	I1.Q2	RDF1
	I1.Q3	
	I1.Q4	
	I1.Q5	RDF1
I2	I2.Q1	VOC2
	I2.Q2	VOC1, VOC2
	I2.Q3	
	I2.Q4	
	I2.Q5	
	I2.Q6	VOC2
	I2.Q7	VOC1
I3	I3.Q1	
	I3.Q2	
	I3.Q3	

Table 6 - Mapping between the tests performed by FOOPS! and O'FAIRE, and their relation to each FAIR principle for the "Reusable" category.

Principle	O'FAIRe question	FOOPS! test
R1	R1.Q1	VOC3, VOC4
	R1.Q2	
	R1.Q3	VOC3
	R1.Q4	VOC4
	R1.Q5	
	R1.Q6	

Principle	O'FAIRe question	FOOPS! test
		DOC1
		OM2
		OM3
R1.1	R1.1.Q1	OM4.1, OM4.2
	R1.1.Q2	
	R1.1.Q3	
R1.2	R1.2.Q1	OM5.1
	R1.2.Q2	OM5.2
	R1.2.Q3	
	R1.2.Q4	OM5.2, OM2, OM3
	R1.2.Q5	
	R1.2.Q6	
	R1.2.Q7	
	R1.2.Q8	
R1.3	R1.3.Q1	
	R1.3.Q2	
	R1.3.Q3	

4.2.1 Check implementation language

Table 7 describes the guidelines for checking the implementation language of a semantic artefact. These guidelines include assessing whether the checked resource is available in one of the common RDF serialisations (Turtle,²¹ RDF/XML,²² N-Triples,²³ JSON-LD²⁴). FOOPS! addresses this test by attempting to read the semantic artefact file through its URI (RDF1). O'FAIRe instead looks at the declared metadata in the repository, looking for the representation language used (I1Q1), whether the representation language is a W3C specification (I1Q2) and whether the ontology syntax is declared (I1Q3). Additionally, O'FAIRe checks if the semantic artefact contains explicit links to the supported serialisations in their metadata (I1Q5) and their formality level (I1Q4). The table associates the guideline with the relevant FAIR principle, and how it is supported in the tools.

²¹ <https://www.w3.org/TR/turtle/>

²² <https://www.w3.org/TR/rdf-syntax-grammar/>

²³ <https://www.w3.org/TR/n-triples/>

²⁴ <https://json-ld.org/>; <https://www.w3.org/TR/json-ld11/>

Table 7 - Guidelines and existing tool support for checking the implementation language of a semantic artefact.

Step	Check implementation language		
Input	Ontology code		
Rationale	Ontologies may be available in multiple implementation languages. In order to check for interoperability, this activity inspects whether the ontology is available in a standardised language. This step is part of the pre-assessment method, as the language an ontology is in may be verified locally even if the ontology has not been made publicly available in the Web.		
Guidelines	FAIR principle	Tool support	
		FOOPS!	O'FAIRe
The semantic artefact has a commonly used RDF serialisation (Turtle, RDF/XML, N-Triples, JSON-LD)	I1	RDF1	I1Q1, I1Q2, I1Q3
The semantic artefact metadata indicates the serialisations it is available in	I1	-	I1Q3, I1Q4, I1Q5

4.2.2 Check ontology elements metadata

Table 8 describes the two guidelines related to the assessment of the metadata associated with ontology elements. We divide them into two main categories. The first one is the assessment of the metadata elements describing terms themselves, such as labels (VOC3 in FOOPS!, R1Q3 in O'FAIRe), definitions (VOC4 in FOOPS!, R1Q4 in O'FAIRe), or equivalences (R1Q5). The second one are those metadata elements that indicate where a term may have been derived from, e.g., properties like author or source (R1Q6), as well as establishing links to existent URIs (I3Q3).

Those tests that are accompanied by '(ni)' in the table indicate that they are defined but not yet implemented within the tool (e.g., R1Q3 (ni)).

Table 8 - Guidelines and existing tool support for checking the metadata of the elements included in an ontology.

Step	Check ontology elements metadata
Input	Ontology content

Rationale	A key aspect for ontology reusability is whether all its elements are properly defined and described with metadata. This includes provenance statements detailing the rationale for the addition of terms in the ontology. This activity proposes an assessment in this direction.		
Guidelines	FAIR principle	Tool support	
		FOOPS!	O'FAIRe
Ontology terms are defined with definition and rationale metadata	R1	VOC3, VOC4	R1Q3 (ni), R1Q4 (ni), R1Q5 (ni)
Ontology terms include provenance information	R1.2, I3	-	R1Q6(ni), I3Q3

(ni): not implemented

4.2.3 Check existing vocabulary reuse

Table 9 introduces the current guidelines for assessing vocabulary reuse in an ontology or vocabulary. These may be produced at two main levels: a semantic artefact may reuse existing vocabularies for defining its own metadata, or it may reuse vocabularies for defining its main concepts and relationships.

The first level is assessed by principle VOC1 in FOOPS! and I2Q2 in O'FAIRe, as both tools look into the properties used in the metadata of an ontology to search for commonly used vocabularies. The second level is assessed by VOC2 in FOOPS! and (mainly) by I2Q1 in O'FAIRe, as both tests look into the imported and reused ontologies within a given semantic artefact. O'FAIRe also defines an assessment on how well the alignment against existing vocabularies is documented within the ontology, and whether the semantic artefact refers to other external resources such as databases (I3Q1, I3Q2), but these are not yet currently supported by the tool.

Finally, best practices suggest that reused vocabularies should follow the FAIR principles (Le Franc et al. 2022). This is addressed by I2Q4 and I2Q7 in O'FAIRE, which explore if the authors declare such information in the ontology metadata.

Table 9 - Guidelines and existing tool support for assessing vocabulary reuse within a semantic artefact.

Step	Check existing vocabulary reuse
Input	Ontology content
Rationale	Ontologies are created by extending other vocabularies. This activity assesses whether an ontology reuses or extends other vocabularies, as well as their compliance against the FAIR principles.

Guidelines	FAIR principle	Tool support	
		FOOPS!	O'FAIRe
Ontology reuses common metadata annotation properties from other vocabularies	I2	VOC1	I2Q2
Ontology extends or imports other vocabularies	I2, I3	VOC2	I2Q1, I2Q3, I2Q4, I2Q5 (ni), I2Q6, I3Q1 (ni), I3Q2 (ni), I3Q3
Reused vocabularies follow the FAIR principles	I2	-	I2Q4, I2Q7

4.2.4 Check ontology metadata

Most of the FAIR principles are highly associated with quality metadata of a resource. Hence, this step contains most of the guidelines in our methodology. Table 10 summarises each guideline and corresponding test by FOOPS! and O'FAIRe. Our guidelines include checking that a persistent identifier (PID) has been used as the semantic artefact main URI (PURL1 and F1Q1, F1Q2 for FOOPS! and O'FAIRe respectively), that an ontology provides version information and that the version is consistent with the ontology URI (VER1, URI2 in FOOPS!, F1Q3, F1Q4 in O'FAIRe), that the ontology reuses common metadata properties from existing vocabularies, and that the ontology contains a valid licence as well as provenance information. Several tests from FOOPS! and O'FAIRe correspond to the last two guidelines, as both tools check for the existence of a set of licence- and provenance-related metadata properties within a semantic artefact.

One additional guideline also checks that the metadata of the ontology refers to the ontology itself. This is useful as in many semantic artefacts' metadata is stored along with the ontology itself. However, external registries may contain additional information of a given resource (tests F3Q2 and F3Q3 from O'FAIRe explore this direction).

As for the minimum (required, mandatory), recommended, and optional metadata properties, both FOOPS! and O'FAIRe have a series of tests to assess their existence (OM1, OM2 and OM3 in FOOPS! and F2Q1, F2Q2, F3Q3, I3Q3, R1Q1 and Q1Q2 in O'FAIRe). These sets of metadata data are defined in Garijo et al. 2020 for FOOPS! and in MIRO (Matentzoglu et al. 2018) for O'FAIRe. These overlap significantly but are not equal to our proposal, discussed in Section 4.2.4.1 based on expert evaluation (Le Franc et al. 2022) and an extensive quantitative metadata analysis of semantic artefacts.

Table 10 - Guidelines and existing tool support for assessing ontology metadata to support FAIR.

Step	Check ontology metadata
-------------	-------------------------

Input	Ontology metadata		
Rationale	Metadata is often crucial for assessing different FAIR principles. This activity focuses on the different metadata categories that may be used to describe ontologies.		
Guidelines	FAIR principle	Tool support	
		FOOPS!	OFAIRE
Ontology URI has a persistent identifier (e.g., using a persistent identifier service)	F1	PURL1	F1Q1, F1Q2
Ontology provides version information	F1, R1.2	VER1, URI2	F1Q3, F1Q4
Ontology includes minimum metadata	F2, R1, I3	OM1	F2Q1, I3Q3
Ontology includes recommended metadata	F2, R1, I3	OM2, OM3	F2Q2, I3Q3
Ontology includes optional metadata	F2, R1, I3	OM2, OM3	F2Q2, F2Q3, I3Q3, R1Q1, R1Q2
Ontology metadata refers to the ontology it describes	F3	OM1, URI2, FIND1	F2Q1, F3Q1, F3Q2, F3Q3
Ontology reuses common metadata annotation properties from other vocabularies	I2	VOC1	I2Q2, I2Q7
Ontology includes licence information	R1.1	OM4.1, OM4.2	R1.1Q1, R1.1Q2, R1.1Q3
Ontology defines provenance information	R1.2	OM5.1, OM5.2, OM2, OM3	R1.2Q1 to R1.2Q8

4.2.4.1 Metadata recommendations in the proposed FAIR assessment methodology

Table A2.1 in Appendix 1 shows the support for MOD categories in the over 2500 analysed different semantic artefacts (nearly 2000 ontologies and over 500 SKOS vocabularies). Since different semantic artefacts use different metadata properties, we have mapped each of them following the most recent MOD specification.²⁵ The column “Total #properties grouped” summarises how many properties have been manually clustered to a given MOD category. The column “support by percentage” details the percentage of vocabularies including a particular MOD category, while the column “ranking according to Jonquet et al 2023” compares our results against a cross-community minimum metadata schema for FAIR semantic artefacts developed in the FAIRsFAIR project (Le Franc et al. 2022; Jonquet et al. 2023), based on DCAT and MOD. That work presents an orthogonal perspective to our approach, as it specifies the agreement for each metadata category by a set of community experts. Finally, the column “supports FAIR principle” summarises the FAIR principles addressed (at least partially) by specifying that metadata category.

²⁵ MOD2.1, see <https://github.com/FAIR-IMPACT/MOD>

Based on the support of each metadata category for each FAIR principle and the expert agreement shown in (Jonquet et al. 2023), as well as metadata relevance for FAIR, we suggest minimum, recommended, and optional metadata for semantic artefacts. Our metadata analysis indicates the support and expert consensus agreement for many of the metadata categories. The minimum, recommended and optional metadata can be seen below:

Minimum metadata: Metadata contributing to address the FAIR principles, supported by at least **20%** of existing semantic artefacts, and having expert support (Table A2.1). These are: *Title, Description, Creator, Creation date, Modification date, License, Version IRI*.

Since ontologies may not have *contributors* or *imports*, we consider those categories as important to be included, but only if such information is available.

Version information, Preferred Namespace URI, and Preferred and Namespace Prefix are not listed directly within the metadata list in (Jonquet et al. 2023), but hold significant support in the analysis, and hence we recommend them. Alternatively, *Acronym, Keywords, and Contact* are deemed mandatory by experts, but have not been used by the community. This may be due to the fact that ontology developers usually represent the acronym with the namespace prefix, and the point of contact of an ontology is usually one of its creators. However, the presence of keywords is key for findability, and we recommend their inclusion in ontologies.

Recommended metadata: Metadata contributing to address the FAIR principles, supported by at least 5% of existing semantic artefacts, and recommended by Jonquet et al. (2023) (Table A2.1). These categories are *Access rights, Submission date, and Status*.

Source and *Prior version* have support among vocabularies and are relevant for identifying the provenance of a semantic artefact. Hence, we include them among our recommendations. Finally, *Notes or comments* are commonly used among vocabularies to clarify usage aspects about semantic artefacts (hence related to R1) but are less important for FAIR in general.

Root resources are typically found among SKOS vocabularies, and hence we recommend describing them too. A *publisher* may not always be associated with an ontology, but when it is, we consider this property recommended, as it may include an authoritative source important for reuse. *Bibliographic reference* is only found in 1-5% of the current semantic artefacts (Table A2.1; see also below), but since it is important to include this in order to make a semantic artefact FAIR, we recommend its use nonetheless.

Optional metadata: Occurring in between 5% and 1% of existing analysed semantic artefacts and at least optional in Jonquet et al. (2023) (Table A2.1). These properties include useful metadata and pointers to related resources such as documentation, related papers and activities that may help understand an ontology. The related MOD categories are: *Natural language, Has format, Subject, Other identifier, Homepage, Relies on, Generic Type,*



Specializes, Bibliographic reference, Was generated by, Changes, Is part of (view of), Representation Language, and Rights Holder.

MOD categories that have less than 1% of support in the analysis (Table A2.1) have not been included in the recommendations of the methodology, except if considered in Jonquet et al. (2023). For example, the category *Indexed or Included in catalog or repository* shows no current adoption by any of the analysed vocabularies, but may be key to indicate whether an ontology is available in an existing registry (very relevant for the A2 principle).

As shown in Table 10, different tools provide adoption for different metadata properties according to the interpretation of the authors and the community. The proposed categorization here takes into account the current support for metadata properties in the vocabularies developed by the community and the advice of experts, in order to balance the need for metadata and the effort needed from developers and authors.

4.2.5 Check whether the ontology is available in registries

Once an ontology has been made publicly available on the web, it is highly recommended to share it in registries, so it can be easily found by others. Our guidelines assess in Table 11 whether an ontology and its metadata are available in existing registries: LOV²⁶ (Vandenbussche et al. 2017) and prefix.cc²⁷ for FOOPS!, and the Agroportal instance where O’FAIRE is installed, respectively. This is supported by FIND2, FIND3 and FIND3_BIS from FOOPS!, and tests F4Q1, F4Q2 from O’FAIRE. Tests A2Q2 and A2Q3 instead look at whether the different versions of a semantic artefact are stored within a given resource.

Table 11 - Guidelines and existing tool support for checking if an ontology exists in external registries or community repositories.

Step	Check ontology is available in registries		
Input	Online ontology (URI)		
Rationales	Ontologies may be self-contained artefacts, which contain both metadata and axioms about a domain. However, once an ontology is available online, it may also be stored in existing registries, thereby easing finding existing ontology terms for reuse.		
Guideline	FAIR principle	Tool support	
		FOOPS!	OFAIRE
Ontology (and its metadata) is available in external registries	F4, A2	FIND2, FIND3, FIND3_(BIS)	F4Q1, F4Q2, F4Q3 (ni), A2Q1, A2Q4

²⁶ See also <https://lov.linkeddata.es/dataset/lov/>

²⁷ <http://prefix.cc>





Ontology versions are available in external repositories	A2	-	A2Q2, A2Q3
--	----	---	------------

4.2.6 Check online publication

Table 12 shows an overview of the different tests we propose to assess the online publication of an ontology. First, an RDF serialisation should be retrievable from the ontology URI (CN1 in FOOPS!, A1Q1 and A1Q3 in O’FAIRE) and implementing content negotiation (CN1 in FOOPS!, A1Q1 and A1Q3 in O’FAIRE). Second, a human-readable version of the ontology documentation should be available online (DOC2). Next, the ontology should declare a version IRI (as proposed in section 4.2.4), and that version IRI should resolve to the corresponding ontology version (URI1, VER2 in FOOPS! and A1Q1, A1Q2 in O’FAIRE).

Finally, ontologies should be served through an open protocol (typically HTTP(S)), and the ontology metadata should be accessible through its URI, as typically metadata can be found in the ontology file itself (HTTP1, CN1 in FOOPS!, A1.1Q1 - A1.1Q3, A1Q4 and A1Q2 in O’FAIREe). O’FAIRE also checks if the repository where an ontology has been stored supports authentication and authorization (A1.2Q1, A1.2Q2).

Table 12 - Guidelines and existing tool support for checking whether an ontology is available on the Web.

Step	Check online publication		
Input	Online Ontology (URI)		
Rationale	Once an ontology finishes a development iteration, it should be made available online following the Linked Data principles for other researchers to reuse. Typically, ontologies will be published under a persistent URI, which resolves to the corresponding RDF representation when requested (if available).		
Guidelines	FAIR principle	Tool support	
		FOOPS!	OFAIRE
Ontology RDF serialisation is available online	A1	CN1	A1Q1, A1Q3
Ontology documentation is available online	A1, R1	DOC1	-
Ontology declares a version IRI, which resolves	F1, A1	URI1, VER2	A1Q1, A1Q2
Ontology implements content negotiation	A1	CN1	A1Q1, A1Q3
Ontology is accessible through an open protocol (e.g., HTTPS)	A1.1	HTTP1	A1.1Q1, A1.1Q2, A1.1Q3, A1Q4

²⁸ <https://www.w3.org/TR/swbp-vocab-pub/>



Ontology metadata is accessible through the ontology URI	A1, A1.1	CN1	A1Q2
Protocol supports authentication and authorization if an ontology has access restrictions	A1.2	-	A1.2Q1, A1.2Q2

4.2.7 Check community-based best practices

Table 13 describes our proposed guidelines to check community-based best practices. Since best practices differ between different communities, the guidelines here only present a generic guideline to support community-based best practices (e.g., OBO Foundry,²⁹ MIRO,³⁰ IOF,³¹ etc.). When adapting this methodology to a new domain, developers are encouraged to add new guidelines in this table in order to adapt to that particular domain. As shown in Figure 2 this step is transversal to all the other steps. That is, some communities might implement language standards while other communities may implement standard metadata vocabularies or publishing practices. For this reason, this step is depicted overlapping with the rest of the steps in the figure.

As for existing tool support for these guidelines, FOOPS! does not define any tests, leaving them open to developers to expand on them. O’FAIRe considers informing on the projects of a particular community that reuse a given ontology (R1.3Q1), as well as indicating whether an ontology is open and available (R1.3Q3), which is a FOOPS! assumption. Finally, O’FAIRe also checks whether the metadata of an ontology indicates belonging to a community set or group, such as the OBO library³² (R1.3Q2).

Table 13 - Guidelines and existing tool support for checking community-based best practices within an ontology.

Step	Check community-based best practices		
Input	Online Ontology (URI), Ontology content, Ontology metadata		
Rationale			
Guidelines	FAIR principle	Tool support	
		FOOPS!	OFAIRE
Ontology follows criteria to align to different community standards or best practices. For example, alignment with OBO	R1.3	-	R1.3Q1, R1.3Q2, R1.3Q3

²⁹ <http://obofoundry.org/>

³⁰ Minimum Information for Reporting an Ontology, see Matentzoglou et al. 2018

³¹ Industry Ontology Foundry, see <https://industrialontologies.org/> or <https://ontocommons.eu/initiatives/industry-ontology-foundry>

³² <http://obofoundry.org/>

terms, inclusion in a community registry, etc.			
--	--	--	--

4.3 FAIR Scoring mechanism(s)

The methodology presented here proposes a series of tests to assess the FAIRness of a semantic artefact, grouped in a series of common steps. These steps may be performed before a version of an ontology is available (pre-assessment), or once it has been published online (assessment). In order to obtain an interpretation of the number of tests passed, both FOOPS! and O’FAIRe present a final score, summarising the status of a semantic artefact regarding its compliance with FAIR. However, even though both FOOPS! and O’FAIRe have an overlap between many tests and questions, the scores given by each of them for the same semantic artefact may diverge. Aligning scores is challenging for three main reasons:

- 1) **Many tests are non-binary:** For example, we propose a set of minimum, recommended, and optional metadata. However, a semantic artefact may only have a percentage of the recommended or minimum metadata, in which case it would not be appropriate to state that the test is completely unfulfilled. Hence, scores must support partial test compliance.
- 2) **Tests may have different weights:** Different assessment tool developers may give different weights to certain tests, based on their expertise or domain. For example, I1Q1 in O’FAIRe may score up to 18 points if an ontology and metadata is available in multiple languages such as PDF, XML, TXT, or OWL. Others may state that having a licence is key for their domain, hence biasing their FAIR analysis in that regard.
- 3) **Grouping scores:** Another design decision that may be made by developers is to group assessment scores according to a mechanism (e.g., grouping per FAIR (sub)principle). On the one hand, if developers decide to group all tests by FAIR principle, then the FAIR principles with more tests (Findable, Reusable) become less relevant in the final FAIR assessment. Alternatively, the scoring may depend on the number of tests passed out of the total number of tests. This is the direction taken into account by FOOPS! and O’FAIRe (and the recommendations made in this document), although the latter includes weights to bias the score towards the most important tests deemed by the tool developers.

In summary, there is no one-size-fits-all scoring function, as there are small subtleties depending on particular use cases and design decisions by tool developers. For example, on the one hand, O’FAIRe may prioritise the availability of mappings against existing resources in a registry, as the tool currently is supported by registries like AgroPortal. On the other hand, FOOPS! assesses the resource as a standalone ontology, where links to other resources are important, but not so much if those resources can also be found in registries.

While here we do not propose a specific scoring function, thanks to this effort we can now correlate similar tests, making it possible to create a consistent output even if the final



scores for a given semantic artefact are different (i.e., scores grow and decrease consistently for the different semantic artefacts when assessed by both FOOPS! and O'FAIRe). Discussions towards harmonising tests and scores are currently taking place in the community for different kinds of research outputs (Verburg et al. 2023).



5. Conclusions and next steps

In this document we propose a methodology to assess the FAIRness of semantic artefacts. We do so by extending LOT, an existing ontology engineering methodology already used in developing many ontologies. Our methodology's guidelines map to the tests of two existing assessment tools for semantic artefacts: FOOPS! and O'FAIRe, mapping their tests together against the FAIR principles and to existing guidelines and best practices for ontologies and vocabularies.

As part of our methodology, we have identified a series of gaps in the FOOPS! and O'FAIRe assessment tools. First, neither of the tools supports a pre-assessment of a semantic artefact, since both assume that the ontology or vocabulary to assess is already available on the Web or in a public registry. Enabling developers with early pre-assessment of their artefacts may help them to align against the FAIR principles before their ontologies are officially released. Second, our ontology metadata landscape and alignment analysis has uncovered metadata recommendations that differ slightly between both tools. We are currently in discussions to enhance tool support for all the metadata properties proposed here, in order to align the outputs of both FOOPS! and O'FAIRe. A series of issues have been opened in the FOOPS! development repository to guide the future development of the tool.³³ The minimum, recommended, and optional metadata properties will be made available in a machine-readable manner, in order to help developers check compliance with them.

During this work we have established different collaborations. Firstly, we have carried out the alignments between FOOPS! and O'FAIRe in collaboration with INRAE members and ex-members. In addition, there is an ongoing collaboration in the FAIR-IMPACT project to align different strands of work relating to the methodologies and outputs of FAIR assessment tools (for different research outputs and scientific domains). The goal of this collaboration is to ensure homogeneous test reporting in other types of FAIR products, such as data and software.

Finally, work is also ongoing in building a semantic artefact benchmark to ease the comparison of FAIR assessment scores between FOOPS! and O'FAIRe. The benchmark will aid in evaluating the impact of different semantic artefacts in our methodology and both assessment tools.

The work of this milestone will be summarised in a research paper, to be published during the lifetime of the project.

³³ https://github.com/oeg-upm/fair_ontologies/issues?q=is%3Aissue+is%3Aopen+label%3A%22FAIRIMPACT+enhancement%22

References

- Amdouni, E., Bouazzouni, S. and Jonquet, C. (2022a) 'O'FAIRe makes you an offer: metadata-based automatic FAIRness assessment for ontologies and semantic resources', *International Journal of Metadata, Semantics and Ontologies*, 16(1), pp. 16–46. Available at: <https://doi.org/10.1504/IJMSO.2022.131133>.
- Amdouni, E., Bouazzouni, S. and Jonquet, C. (2022b) 'O'FAIRe: Ontology FAIRness Evaluator in the AgroPortal semantic resource repository', in. *ESWC 2022 - 19th Extended Semantic Web Conference, Poster and demonstration*, p. 89. Available at: https://doi.org/10.1007/978-3-031-11609-4_17.
- Cox, S.J.D. et al. (2021) 'Ten simple rules for making a vocabulary FAIR', *PLOS Computational Biology*, 17(6), p. e1009041. Available at: <https://doi.org/10.1371/journal.pcbi.1009041>.
- Garijo, D. and Poveda-Villalón, M. (2020) 'Best Practices for Implementing FAIR Vocabularies and Ontologies on the Web', in G. Cota, M. Daquino, and G.L. Pozzato (eds) *Studies on the Semantic Web*. IOS Press. Available at: <https://doi.org/10.3233/SSW200034>.
- Garijo, D., Corcho, Ó. and Poveda-Villalón, M. (2021) 'FOOPS!: An Ontology Pitfall Scanner for the FAIR principles', in *International Semantic Web Conference (ISWC) 2021: Posters, Demos, and Industry Tracks*. International Workshop on the Semantic Web, CEUR-WS.org. Available at: <http://ceur-ws.org/Vol-2980/paper321.pdf>.
- Garijo, D., Beltran, A., & Wilson, A. (2023). 'Ontology metadata categorization (0.0.1)' [Data set], Zenodo. Available at: <https://doi.org/10.5281/zenodo.8299390>.
- Hugo, W., Le Franc, Y., Coen, G., Parland-von Essen, J., & Bonino, L. (2020). 'D2.5 FAIR Semantics Recommendations Second Iteration (1.0)', FAIRsFAIR, Zenodo. Available at: <https://doi.org/10.5281/zenodo.5362010>.
- Jonquet, C., Dutta, B., Bonino da Silva Santos, L., Pergl, R., Le Franc, Y. (2023) 'Common Minimum Metadata for FAIR Semantic Artefacts', in C. Trojahn, L. O. Bonino da Silva Santos, G. Guizzardi, C. Jonquet (eds) *2nd Workshop on Ontologies for FAIR and FAIR Ontologies (Onto4FAIR)*. Sherbrooke, Canada. Available at: <https://hal.science/hal-04106533v2>.
- Le Franc, Y.; Parland-von Essen, J.; Bonino, L.; Lehtväslaiho, H.; Coen, G. & Staiger, C.. (2020). 'D2.2 FAIR Semantics: First recommendations (1.0)', FAIRsFAIR, Zenodo. Available at: <https://doi.org/10.5281/zenodo.5361930>.
- Le Franc, Y., Bonino, L., Koivula, H., Parland-von Essen, J. & Pergl, R. (2022). 'D2.8 FAIR Semantics Recommendations Third Iteration (V1.0)', FAIRsFAIR, Zenodo. Available at: <https://doi.org/10.5281/zenodo.6675295>.
- Matentzoglou, N., Malone, J., Mungall, C. et al. (2018). 'MIRO: guidelines for minimum information for the reporting of an ontology', *Journal of Biomedical Semantics*, 9(6). Available at: <https://doi.org/10.1186/s13326-017-0172-7>.



Poveda-Villalón, M. et al. (2020) 'Coming to Terms with FAIR Ontologies', in C.M. Keet and M. Dumontier (eds) Knowledge Engineering and Knowledge Management. Cham: Springer International Publishing (Lecture Notes in Computer Science), pp. 255–270. Available at: https://doi.org/10.1007/978-3-030-61244-3_18.

Poveda-Villalón, M. et al. (2022) 'LOT: An industrial oriented ontology engineering framework', Engineering Applications of Artificial Intelligence, 111, p. 104755. Available at: <https://doi.org/10.1016/j.engappai.2022.104755>.

Vandenbussche, P. Y., Ateazing, G. A., Poveda-Villalón, M., & Vatant, B. (2017). 'Linked Open Vocabularies (LOV): a gateway to reusable semantic vocabularies on the Web', Semantic Web, 8(3), pp. 437-452. Available at: <https://doi.org/10.3233/SW-160213>.

Verburg, M. L., Huber, R., Jonquet, C., & Garijo, D. (2023). 'FAIR-IMPACT project response to "FAIR Assessment Tools: Towards an "Apples to Apples" Comparisons"', FAIR-IMPACT, Zenodo. Available at: <https://doi.org/10.5281/zenodo.7848101>.

Wilkinson, M.D. et al. (2016) 'The FAIR Guiding Principles for scientific data management and stewardship', Scientific Data, 3(1), p. 160018. Available at: <https://doi.org/10.1038/sdata.2016.18>.

Xu, F. et al. (2023) 'Features of a FAIR vocabulary', Journal of Biomedical Semantics, 14(1), p. 6. Available at: <https://doi.org/10.1186/s13326-023-00286-8>.



Appendices

Appendix 1: FOOPS! and O'FAIRE definitions

Table A1.1 - FOOPS! Tests per principle

FAIR principle		FOOPS! test	Definition
F1		PURL1	Persistent URL: This check verifies if the ontology has a persistent URL (w3id, PURL, DOI, or a W3C URL)
		URI1	Ontology URI is resolvable: This check verifies if the ontology URI found within the ontology document is resolvable
		VER1	Version IRI: This check verifies if there is an id for this ontology version, and whether the id is unique (i.e., different from the ontology URI)
		VER2	Version IRI resolves: This check verifies if the version IRI resolves
		URI2	Consistent ontology ids: This check verifies if the ontology URI is equal to the ontology ID
F2		OM1	Minimum metadata: This check verifies if the The following minimum metadata [title, description, licence, version IRI, creator, creationDate, namespace URI] are present in the ontology
F3		FIND1	Ontology prefix: This check verifies if an ontology prefix is available
F4		FIND2	Prefix is in a registry. This check verifies if the ontology prefix can be found in prefix.cc or LOV registries. This check also verifies if the prefix resolves to the same namespaceprefix found in the ontology.
		FIND3	Ontology in metadata registry. This check verifies if the ontology can be found in a public registry (LOV)
A1		CN1	Content negotiation for RDF and HTML. This check verifies of the ontology URI is published following the right content negotiation for RDF and HTML
A2		FIND_3_BIS	Metadata are accessible, even when the data are no longer available. Metadata are accessible even when the ontology is no longer available. Since the metadata is usually included in the ontology, this check verifies whether the ontology is registered in a public metadata registry (LOV)
A1.1		HTTP!	Open protocol. This check verifies if the ontology uses an open protocol (HTTP or HTTPS)

I1		RDF1	RDF Availability. This check verifies if the ontology has an RDF serialisation (Turtle, N3, RDF/XML, JSON-LD)
I2		VOC1	Vocabulary reuse (metadata). This check verifies if the ontology reuses other vocabularies for declaring metadata terms
		VOC2	Vocabulary reuse. This check verifies if the ontology imports/extends other vocabularies (besides RDF, OWL and RDFS)
R1		DOC1	HTML availability. This check verifies if the ontology has an HTML documentation
		OM2	Recommended metadata. This check verifies if the following recommended metadata [NS Prefix, version info, creation date, citation] are present in the ontology. It also checks if [contributor] is present, but with no penalty (as no all ontologies may have a contributor)
		OM3	Detailed metadata. This check verifies if the following detailed metadata [DOI, publisher, logo, status, source, issued date] are present in the ontology. It also checks if [previous version, backward compatibility, modified] are present, but with no penalty (as no all ontologies may have, e.g., a previous version)
		VOC3	Documentation labels. This check verifies the extent to which all ontology terms have labels (rdfs:label in OWL vocabularies, skos:prefLabel in SKOS vocabularies)
		VOC4	Documentation definitions. This check verifies whether all ontology terms have descriptions (rdfs:comment in OWL vocabularies, skos:definition in SKOS vocabularies)
R1.1		OM4.1	Licence availability. This check verifies if a licence associated with the ontology
		OM4.2	Licence is resolvable. This check verifies if the ontology licence is resolvable
R1.2		OM5.1	Basic provenance metadata: This check verifies if basic provenance is available for the ontology: [author, creation date]. This check also verifies whether [contributor, previous version] are present, but with no penalty (as no all ontologies may have a previous version or a contributor)
		OM5.2	Detailed provenance metadata. This check verifies if detailed provenance information is available for the ontology: [issued date, publisher]

Table A1.2 - O'FAIRe Tests per principle

FAIR principle	O'FAIRe id	O'FAIRe definition
----------------	------------	--------------------



F1	F1Q1	Does the ontology have a "local" identifier, i.e., a globally unique and potentially identifier assigned by the developer (or developing organisation)?
	F1Q2	Does the ontology provide an additional "external" identifier, i.e., a guarantee globally unique and persistent identifier assigned by an accredited body? If yes, is the external identifier a DOI?
	F1Q3	Are the ontology metadata clearly identified either by the same identifier than the ontology (if included in the ontology file) or with its own globally unique and persistent identifier?
	F1Q4	Does the ontology provide a version-specific URI, and is this URI resolvable?
F2	F2Q1	Is the ontology described with additional 'MIRO must' metadata properties?
	F2Q2	Is the ontology described with additional 'MIRO should' or 'optional' metadata properties?
	F2Q3	Is the ontology described with another metadata property with no explicit corresponding MIRO requirement?
F3	F3Q1	Are the ontology metadata included and maintained in the ontology file?
	F3Q2	If not, are the ontology metadata described in an external file?
	F3Q3	Does that external file explicitly link to the ontology and vice-versa?
F4	F4Q1	Is the ontology registered in multiple ontology 'libraries'?
	F4Q2	Is the ontology registered in multiple open ontology 'repositories'?
	F4Q3	Are the ontology 'libraries' or 'repositories' properly indexed by Web search engines?
A1	A1Q1	Do the ontology URI and other identifiers, if they exist, resolve to the ontology?
	A1Q2	Does the ontology URI (if metadata are included in the ontology file) or the external metadata URI resolve to the metadata record?
	A1Q3	Do the ontology URI and the external metadata URI (if the metadata are not included in the ontology file), support content negotiation?
	A1Q4	Are the ontology and its metadata accessible through another standard protocol such as SPARQL?
A1.1	A1.1Q1	Is the ontology relying on HTTP/URIs for its identification and access mechanisms?
	A1.1Q2	Is the ontology access protocol open, free, and universally implementable?
	A1.1Q3	If the ontology and metadata are accessible through another protocol, is that protocol open, free, and universally implementable?
A1.2	A1.2Q1	Is the ontology accessible through a protocol that supports authentication and authorization?



	A1.2Q2	Are the ontology metadata accessible through a protocol that supports authentication and authorization?
A2	A2Q1	Is the ontology accessible in a repository that supports versioning?
	A2Q2	Are the ontology metadata of each version available?
	A2Q3	Are the ontology metadata accessible even if no more versions of the ontology are available?
	A2Q4	Is the status of the ontology clearly informed?
I1	I1Q1	What is the representation language used for the ontology and ontology metadata?
	I1Q2	Is the representation language used in a W3C Recommendation?
	I1Q3	Is the syntax of the ontology informed?
	I1Q4	Is the formality level of the ontology informed?
	I1Q5	Is the availability of other syntaxes/formats informed?
I2	I2Q1	Does the ontology import other FAIR vocabularies?
	I2Q2	Does the ontology reuse terms from other FAIR vocabularies (URIs)?
	I2Q3	If yes, does it include the minimum information for those terms?
	I2Q4	Is the ontology aligned to other FAIR vocabularies?
	I2Q5	If yes, are those alignments well represented and to unambiguous entities? If yes, are those alignments curated?
	I2Q6	Does the ontology provide information about the relation to or influence of other FAIR vocabularies?
	I2Q7	Does the ontology reuse standard and FAIR metadata vocabularies to describe its metadata?
I3	I3Q1	Does the ontology provide qualified cross-references to external resources/databases?
	I3Q2	If yes, are those cross-references well represented and to unambiguous entities?
	I3Q3	Does the ontology use valid URIs to encode some metadata values?
R1	R1Q1	Does the ontology provide information about how classes or concepts are defined?
	R1Q2	Does the ontology provide metadata information about its hierarchy?
	R1Q3	How many of the ontology objects are described with labels?
	R1Q4	How many of the ontology objects are defined using a text description?
	R1Q5	How many ontology objects are defined using a property restriction or an equivalent class?
	R1Q6	How many ontology objects provide provenance information with annotation properties (e.g., author, date)?

R1.1	R1.1Q1	Is the ontology licence clearly specified, with an URI that is resolvable and supports content negotiation?
	R1.1Q2	Are the ontology access rights specified and permissions documented?
	R1.1Q3	Are the ontology usage guidelines and copyright holder documented?
R1.2	R1.2Q1	Does the ontology provide information about the actors involved in its development?
	R1.2Q2	Does the ontology provide information about its general provenance?
	R1.2Q3	Are the accrual methods and policy of the ontology documented?
	R1.2Q4	Is the ontology clearly versioned with version information and links to previous versions?
	R1.2Q5	Are the ontology latest changes documented?
	R1.2Q6	Are the methodology and tools used to build the ontology documented?
	R1.2Q7	Is the ontology rationale documented?
	R1.2Q8	Does the ontology inform about its funding organisation?
R1.3	R3Q1	Does the ontology provide information about projects using or organisations endorsing?
	R3Q2	Is the ontology included in a specific community set or group?
	R3Q1	Is the ontology openly and freely available?

Appendix 2: Table A.2: Semantic Artefact landscape metadata analysis

The table below shows the results from a metadata landscape analysis over 2500 semantic artefacts (nearly 2000 ontologies and 500 SKOS vocabularies). Each of the metadata properties used have been mapped to a MOD metadata category and quantified (Garijo et al. 2023).

Table A2.1 - Results of the semantic artefact landscape analysis, after mapping against existing properties to MOD categories and ranking them by their support. A ranking based on community expert votes (in Jonquet et al. 2023 (Table 1)) is added for comparison (a question mark indicates that the metadata category was not considered). The FAIR principle supported by each metadata category is available in the last column.

MOD Category name	Total #properties grouped	Support by % (0-1)	Ranking based on support	Ranking according to Jonquet et al 2023 (Table 1)	Supports FAIR principle
-------------------	---------------------------	--------------------	--------------------------	---	-------------------------



Title	26	1	Minimum (20-100%)	Minimum ³⁴	F2, R1
Description	15	0,9450158228	Minimum	Minimum	F2, R1
Creator	32	0,519778481	Minimum	Minimum	R1.2
Version information	19	0,4853639241	Minimum	?	F1, R1.2
Licence	23	0,3844936709	Minimum	Minimum	R1.1
Imports	5	0,3338607595	Minimum	?	R1.2
Version IRI	4	0,284414557	Minimum	Minimum	F1, F3, R1.2
Modification date	14	0,2788765823	Minimum	Minimum	R1.2
Creation date	24	0,2638449367	Minimum	Minimum	R1.2
Preferred Namespace URI	8	0,2598892405	Minimum	?	F1, A1
Preferred Namespace Prefix	8	0,2405063291	Minimum	?	F1, R1
Contributor	12	0,2136075949	Minimum	Recommended	R1.2
Access rights	5	0,198971519	Recommended (5-20%)	Minimum	R1.1
Submission date	7	0,184335443	Recommended	Minimum	R1.2
Publisher	8	0,1653481013	Recommended (but may be optional)	Recommended	R1.2
Source	9	0,1428006329	Recommended	?	R1.2
Notes or comments	14	0,116693038	N/A	?	R1
Prior version	6	0,08662974684	Recommended	?	R1.2
Status	12	0,07238924051	Recommended	Recommended	R1
Root resources	3	0,05261075949	Optional, as it is for SKOS vocabularies	?	R1
Natural language	7	0,04707278481	Optional (1-5%)	Recommended	R1
Has format	2	0,04390822785	Optional	Minimum	I1
Subject	9	0,04311708861	Optional	Minimum	R1
Other identifier	5	0,04232594937	Optional	Minimum	I3

³⁴ Note that (Jonquet et al, 2023) uses “mandatory” instead of “minimum”.

Homepage	4	0,0332278481	Optional	Recommended	R1
Relies on	3	0,03125	Optional	Optional	R1.2
Generic Type	5	0,01977848101	Optional	Minimum	R1
Specializes	4	0,01819620253	Optional	?	I2, R1.2
Bibliographic reference	8	0,01621835443	Optional	?	R1
Was generated by	2	0,01542721519	Optional	Optional	R1.2
Changes	2	0,01344936709	Optional	?	R1.2
Is part of (view of)	5	0,01305379747	Optional	?	R1
Representation Language	5	0,01147151899	Optional	?	I2
Rights holder	2	0,01147151899	Optional	?	R1.1
Acronym	6	0,00514240506		Minimum	F1, R1
Keywords	5	0,00276898734		Minimum	F1, R1
Contact	2	0,00079113924		Minimum	R1.2
Indexed or Included in catalog or repository	0	0		Optional	R1.3, F4, A2