Coming to Terms with FAIR Ontologies

A position paper

María Poveda-Villalón[1][0000−0003−3587−0367], Paola Espinoza-Arias[1][0000−0002−3938−2064], Daniel Garijo[2][0000−0003−0454−7145], and Oscar Corcho[1][0000−0002−9260−0753]

[1] Ontology Engineering Group, Universidad Politécnica de Madrid
{mpoveda,pespinoza,ocorcho}@fi.upm.es
[2] Information Sciences Institute, University of Southern California
dgarijo@isi.edu

Abstract. Ontologies are widely used nowadays for many different purposes and in many different contexts, like industry and research, and in domains ranging from geosciences, biology, chemistry or medicine. When used for research, ontologies should be treated as other research artefacts, such as data, software, methods, etc.; following the same principles used to make them findable, accessible, interoperable and reusable (FAIR) to others. However, in comparison to the number of guides, indicators and recommendations available for making research data FAIR, not much attention has been paid so far on how to publish ontologies following the FAIR principles. This position paper reviews the technical and social needs required to define a roadmap for generating and publishing FAIR ontologies on the Web. We analyze four initiatives for ontology publication, aligning them in a common framework for comparison. The paper concludes by opening a discussion about existing, ongoing and required initiatives and instruments to facilitate FAIR ontology sharing on the Web.

Keywords: FAIR principles · Ontologies · Semantics.

1 Introduction

Since its inception in 2016, the FAIR (Findable, Accessible, Interoperable, Reusable) data principles [35] have gained an increasing importance in the context of research data management, and are being adopted by a large number of private and public organisations worldwide, including initiatives such as the European Open Science Cloud[3] (EOSC) or the Research Data alliance[4] (RDA).

Ontologies play a relevant role in some of the FAIR data principles, especially in relation to providing support for data "interoperability" and "reusability". The need for ontologies (also called vocabularies) is pointed out in the following principles: data and metadata should (I2)[5] use vocabularies that follow FAIR principles, (I1) use a formal, accessible, shared, and broadly applicable language for knowledge representation; (I3)

[3] https://www.eosc-portal.eu/
[4] https://www.rd-alliance.org/
[5] We point in parentheses to the principles numeration used in the original FAIR paper [35]

include qualified references to other (meta)data, and (R1.3) meet domain-relevant community standards. Ontologies are also relevant in terms of "findability", (F2) requiring to describe data with rich metadata.

The research community has already acknowledged the need for ontologies to follow the FAIR principles [7]. First, there is a clear movement towards expanding the application of the FAIR principles beyond research data, as described in the ongoing EOSC Interoperability Framework [8]. Since ontologies are often the result of research activities or fundamental components in many areas of research, the FAIR principles should be applied to them, independently of whether they are used to describe data or metadata. Second, ontologies are already identified as a relevant artefact in the principles (even though the term *vocabulary* is used more generally and there is a general preference to talk about semantic artefacts, including thesauri, glossaries, shared UML models, etc.). Therefore, we consider that it is critical for the community to discuss and analyse how the FAIR principles should be applied to these artefacts.

However, we do not start from scratch when it comes to making ontologies available to others. Before the appearance and general acceptance of FAIR principles in research, many approaches had already focused on how to publish ontologies on the Web following Linked Data principles, ensuring the existence of permanent identifiers and making them available through standardised protocols like HTTP [4, 18, 21]. Other approaches focused on making ontologies findable by creating metadata schemas and ontologies to describe them and register them in ontology catalogues and repositories [32, 34, 37, 9, 22, 16, 28].

Some initial studies and reports on how to make ontologies FAIR have recently appeared [26, 24]. For the time being they can be considered as initial proposals coming from working or interest groups under the umbrella of Open Science projects or initiatives (e.g., the FAIRsFAIR EU project,[6] the GO-FAIR implementation network GO-INTER,[7] the RDA Vocabulary Services Interest Group[8]). Other proposals like [11] focus mostly on the technical implementation of some of the FAIR principles. These initiatives are developing proposals and recommendations that may not necessarily fit the view of the Ontology Engineering community at large.

In this position paper we 1) argue that there is a need to open a broader and more open discussion of the technical and social consequences of adopting the FAIR principles for the publication and sharing of ontologies, and that such discussion should incorporate the views of the Ontology Engineering community; 2) analyze and compare existing approaches for making ontologies FAIR; and 3) describe the challenges ahead.

We start the paper with a general review of the FAIR and LOD principles (Sect. 2), moving then into how they have been already considered in seminal approaches that focus on the FAIRification of ontologies, providing both the description of such approaches (Sect. 3) and a comparative analysis of them (Sect. 4). We discuss previous work that may be reused in this context; provide concrete recommendations needed in order to make ontologies FAIR; and expose what we consider to be the next steps towards developing a community recommendation on how to make ontologies FAIR (Sect. 5).

[6] https://fairsfair.eu/

[7] https://www.go-fair.org/implementation-networks/overview/go-inter/

[8] https://www.rd-alliance.org/groups/vocabulary-services-interest-group.html

2 Background

The Linked Data principles[9] were proposed in 2006 as a set of guidelines for publishing and linking data on the Web [4]. The Linked Data principles may be summarized as: 1) use URIs for naming things, 2) use HTTP URIs to search things, 3) use standards (e.g., RDF) to provide useful information about URIs and 4) include links to other URIs. These principles were extended further in 2010, with the 5-star rating system for publishing Linked **Open** Data, which can be summarized as: make the data available in the Web with 1) an open licence, 2) in a machine readable manner, 3) in a non-proprietary format, 4) using RDF to identify and describe things and 5) linking to other data.

In 2016, the FAIR principles [35] were defined as a technology-agnostic and domain-independent guide to enhancing scientific data management and stewardship. Such principles are considered guidelines for those wishing to enhance the reusability of their data. In short, the four high-level FAIR principles stand that data must be easy to find, be accessible by standardized protocols, be machine-readable to enhance interoperability, and be well-described in order to be reusable for both humans and machines. The complete list of FAIR principles is provided in Annex A.

Despite both sets of principles having similar goals and definitions they also exhibit slight differences among them. Both approaches share the goal of using permanent identifiers to identify data (Uniform Resource Identifiers - URIs - for Linked Data, Persistent Identifiers - PIDs - for FAIR), and both promote using standards to provide further information about data, including references to other data. They also share the idea of using a standardized communication protocol to retrieve data (HTTP for Linked Data, and not specified for FAIR). Even though both approaches make explicit the need for licensing data, Linked Data principles are more restrictive in the sense than an open license is imposed while FAIR does not restrict any license permissions. However, unlike Linked Data, FAIR makes an explicit and strong focus on metadata management in order to enable resource findability and reusability. Finally, FAIR includes a set of principles to ease data and metadata findability, which are not covered by Linked Data principles. For further discussion about distinctions an overlaps among LOD and FAIR principles we refer readers to the analysis provided by Hasnain and Rebholz-Schuhmann in 2018 [17].

3 Applying Linked Data and FAIR principles for publishing semantic artefacts

Throughout this document, we use the term *semantic artefact* to refer to a specification of a conceptualization that may be represented by different levels of formalization [27] (including controlled lists, thesauri and ontologies - either lightweight or heavyweight). This section describes the most relevant proposals to address the FAIRness of semantic artefacts as a complement to the FAIR data principles. This includes the ongoing effort from the FAIRsFAIR EU project [24] and the recent guidelines for publishing FAIR ontologies [11], released by co-authors of the present position paper. We also describe existing recommendations for improving the publication of ontologies on the Web. Even

[9] `https://www.w3.org/DesignIssues/LinkedData.html`

though there is a large number of methodologies, guidelines and techniques that may be reused and considered to publish FAIR ontologies, we only consider in this work those initiatives adapting the LOD 5-star schema for ontologies.

A full analysis of the existing methods, guidelines, techniques and tools available for FAIR ontologies may be subject of a dedicated systematic review, beyond the scope of this position paper.

3.1 FAIRsFAIR Recommendations for Ontology Publication

The FAIRsFAIR project, started in 2019, is a European effort aiming to provide practical solutions for the use of the FAIR data principles throughout the research data lifecycle. This project is in close cooperation with other ongoing European projects and several stakeholders to work in an overall knowledge infrastructure on academic quality data management, procedures, standards, metrics, and related matters based on the FAIR principles for the research data providers and repositories. FAIRsFAIR's activities include a specific task dedicated to semantic interoperability, with the aim to support the creation of a federated semantic space. In 2020, this task released a deliverable [24] that provides a list of 17 preliminary recommendations related to the application of FAIR principles to improve the global FAIRness of semantic artefacts. Each recommendation and best practice is related to one or more FAIR principles and links to existing recommendations and related stakeholders (e.g: practitioners, repositories or the Semantic Web community). The list of recommendations [24] includes:

- **P-Rec1:** Use Globally Unique, Persistent and Resolvable Identifier for Semantic Artefacts, their content and their versions.
- **P-Rec2:** Use Globally Unique, Persistent and Resolvable Identifier for Semantic Artefact Metadata Record.
- **P-Rec3:** Use a common minimum metadata schema to describe semantic artefacts and their content.
- **P-Rec4:** Publish the Semantic Artefact and its content in a semantic repository.
- **P-Rec5:** 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.
- **P-Rec6:** Build semantic artefacts' search engines that operate across different semantic repositories.
- **P-Rec7:** Repositories should offer a secure protocol and user access control functionalities.
- **P-Rec8:** Define human and machine-readable persistency policies for semantic artefacts metadata.
- **P-Rec9:** Semantic artefacts should be represented using common serialization formats, e.g., Semantic Web and Linked Data standards.
- **P-Rec10:** Use a Foundational Ontology to align semantic artefacts.
- **P-Rec11:** Use a standardized language for describing semantic artefacts.
- **P-Rec12:** Semantic mappings between the different elements of semantic artefacts should use machine-readable formats based on W3C standards.

P-Rec13: Crosswalks, mappings and bridging between semantic artefacts should be documented, published and curated.
P-Rec14: Use standard vocabularies to describe semantic artefacts.
P-Rec15: Make the references to the reused third-party semantic artefacts explicit.
P-Rec16: The semantic artefact should be clearly licensed for machines and humans.
P-Rec17: Provenance should be clear for both humans and machines.

The work proposed in [24] also identifies a list of 10 best practices (e.g use of naming conventions, use of ontology design patterns, workflows definition between formats, etc.) that go beyond the FAIR scope. Such practices are mostly inspired by the OBO foundry[10] and Industry Ontology Foundry principles[11] and are not necessarily related to any of the FAIR principles. Hence they fall out of scope of our analysis.

3.2 Best Practices for Implementing FAIR Vocabularies and Ontologies on the Web

A coetaneous effort with the FAIRsFAIR recommendation are the best practices for implementing vocabularies and ontologies on the Web [11]. In this work, specific practical guidelines are provided to help users in the following activities:

- Design of Accessible Ontology URIs
  1. Design ontology name and prefix
  2. Decide between hash or slash URIs
  3. Decide whether to use opaque URIs
  4. Define an ontology versioning strategy
  5. Use of permanent URIs
- Generate reusable ontology documentation
  6. Generate ontology metadata
  7. Generate HTML documentation
  8. Generate diagrams
- Publish an ontology on the Web
  9. Provide the ontology online in multiple formats (HTML and ontology serializations)
  10. Make the ontology findable on the Web

3.3 Initiatives for 5-star vocabularies

The 5-star schema for publishing Linked Open Data has been adapted to vocabularies by two different approaches. More precisely, the first approach of 5-star vocabularies was published by Bernard Vatant as a blog post[12] in 2012. The proposed 5-stars for vocabularies are defined as follows:

[10] http://www.obofoundry.org/principles/fp-000-summary.html
[11] https://www.industrialontologies.org/?page_id=87
[12] https://bvatant.blogspot.com/2012/02/is-your-linked-data-vocabulary-5-star_9588.html

1. ☆ Publish your vocabulary on the Web at a stable URI with a open license.[13]
2. ☆☆ Provide human-readable documentation and basic metadata such as creator, publisher, date of creation, last modification, version number.
3. ☆☆☆ Provide labels and descriptions, if possible in several languages, to make your vocabulary usable in multiple linguistic scopes.
4. ☆☆☆☆ Make your vocabulary available via its namespace URI, both as a formal file and human-readable documentation, using content negotiation.
5. ☆☆☆☆☆ Link to other vocabularies by re-using elements rather than re-inventing.

Later, in 2014, an editorial paper from the Semantic Web Journal [21] adapted the idea of 5-stars for vocabularies to the following schema:

1. ☆ There is dereferenceable human-readable information about the used vocabulary.
2. ☆☆ The information is available as machine-readable explicit axiomatization of the vocabulary.
3. ☆☆☆ The vocabulary is linked to other vocabularies.
4. ☆☆☆☆ Metadata about the vocabulary is available (in a dereferencable and machine-readable form).
5. ☆☆☆☆☆ The vocabulary is linked to by other vocabularies.

While these vocabulary-oriented 5-star schemes have not been widely adopted by the community so far, they are often referred to by reviewers when assessing ontology papers for journals and conferences, and ontology repositories are promoting their use. For example, in the Linked Open Vocabularies [32] registry, a vocabulary should 1) be written in RDF and be dereferenceable; 2) be parsed without errors; 3) provide `rdfs:label` for all of its terms; 4) refer to and reuse relevant existing vocabularies; and 5) provide some metadata. These constraints force authors to follow the stars 1,2,3 and 5 from 2012; although it does not force authors to provide human readable documentation with content negotiation nor an open license. Another case is the Smart Cities ontology catalogue [29], where quality indicators are established for ontologies taking into account: a) whether an ontology is available on the Web, in RDF and/or HTML: b) whether an ontology follows the W3C standards (e.g., RDF-S or OWL); and c) whether an ontology is available under an open license.

4 An Analysis Framework for FAIR ontologies

In this section we discuss and compare the initiatives described in Section 3, with the aim of providing insight and food for thought for the next wave of recommendations to be made for the FAIRification of semantic artefacts. We review each of the initiatives and align them with the FAIR principles. The results of our analysis are shown in Table 1, where FAIR principles are listed in columns and guidelines are listed in rows, grouped by initiatives. The numbering of the guidelines corresponds to the numbering provided

[13] Note that the "open license" is added to the first star as a comment by the author as a reaction to the feedback, but not shown in the original list.

in Section 3 and all FAIR principles are listed in Annex A.[14] The values provided for each cell are: "x" when a guideline (row) and a FAIR principle (column) have similar scope; "<" to indicate that the guideline is less strict than the principle; and ">" to indicate that the guideline is more strict than the principle.

Since the FAIR principles focus on data (and its related metadata) and the analyzed initiatives target semantic artefacts (including ontologies), we have considered that a semantic artefact corresponds to the term 'data' in the principles.

Note that the table values for the FAIRsFAIR guidelines have been taken from the original draft publication [24]. The table includes question marks (highlighted in bold blue letters in Table 1) for matches that are not clear to the authors and that will be subject to further discussion below. For the rest of the initiatives, the cell values presented in this table reflect the agreement by the authors and incorporate external feedback and comments from other colleagues at the Ontology Engineering Group at UPM.

The following mismatches have been found between [24] and our understanding of the guidelines and the FAIR principles. It is worth noting that these mismatches, among others, haven been also reported and discussed with FAIRsFAIR representatives and will be reported publicly in the corresponding GitHub repositories when made available, as agreed with them. Indeed, we have included this discussion on mismatches in this paper since it shows the need for an extensive discussion on this topic. In our opinion:

- P-REC12, P-REC13 and P-REC14 relations with the FAIR principles may be revised. Our proposals are: **A)** P-REC12 may be related to I3 and I1 instead of R1.3. The reason is that P-REC12 describes the need for machine-readable descriptions of the mappings, what is more related to interoperability than to community standards. **B)** P-REC13 may be related to I3 instead of R1.3. The rationale behind this is that P-REC13 describes the need for documenting mappings and also mentions sharing such resources. This seems to be more related to interoperability than to community standards, as discussed for the case of P-REC12. **C)** P-REC14 may also be linked to R1.3, since the recommendation explicitly refers to relevant community standards to be used to describe semantic artefacts.
- P-REC10 is related to interoperability principles, emphasizing the need to align semantic artefacts to foundational ontologies, such as DOLCE [5] or UFO [15]. While we acknowledge the benefits that foundational ontologies may bring into ontology development, first, we consider this as a very strong requirement at this stage, considering that many domain ontology developers may have difficulties to understand how to align their semantic artefacts to these ontologies, as shown by the small amount of published ontologies that are currently aligned to them. Second, we think that the definition of foundational ontology could be broaden so that it includes reference ontologies that are well-adopted within some communities, such as the case of schema.org [13], Wikidata, etc. Taking this into account, the description of P-REC10 may be relaxed to emphasize the benefits of linking to foundational ontologies rather than the need to do it, that is stating it as a possibility rather than an obligation.

[14] To ease the reading of the rest of the paper we recommend to have the FAIR principles list (Annex A) and Section 3 at hand.

Table 1. Relationship between initiatives for FAIR semantic artefacts and FAIR principles. In the guidelines the row numbering corresponds to the numbering provided in Section 3 and the FAIR principles column numbering corresponds to the list provided in Annex A.

		FAIR Principles														
Guidelines ↓		F1	F2	F3	F4	A1	A1.1	A1.2	A2	I1	I2	I3	R1	R1.1	R1.2	R1.3
FAIRsFAIR	P-Rec1	x														
	P-Rec2	x		x												
	P-Rec3		x										x	x	x	x
	P-Rec4				x											
	P-Rec5				x	x	x									
	P-Rec6				x											
	P-Rec7							x								
	P-Rec8								x							
	P-Rec9									x						
	P-Rec10									x	x	x				
	P-Rec11									x						
	P-Rec12									x		x				**x?**
	P-Rec13											**?**			x	**x?**
	P-Rec14										x					**?**
	P-Rec15											x			x	
	P-Rec16													x		
	P-Rec17														x	
FAIR ontologies	1	x														
	2	x														
	3	x														
	4												x			
	5	x														
	6		x							x			x	x	x	x
	7					x						x	x			
	8												x			
	9					x				x						
	10				x											x
5-stars 2012	1	<					>							>		
	2												x			
	3		x										x			
	4	<				x	x			x						
	5											x				
5-stars 2014	1					x										
	2					x				x						
	3											x				
	4					x							x			
	5															

– P-REC9 and P-REC11 present some inaccuracies when analysed from a Semantic Web perspective. First, P-REC11 is entitled "use a standardised language for describing semantic artefacts" pointing to SHACL [23], SWRL[15] and OntoUML [14]. SHACL is the only official recommendation from a standardisation body, while RDF(S) [6] and OWL [3, 19] are mentioned in P-REC9. Second, P-REC9 mentions that semantic artefacts should be represented using common serializations formats, however from the Semantic Web perspective the different serializations of an ontology or dataset are just different ways of implementing them in a particular format and syntax, but the semantics are equivalent and are defined by the ontology language, not the serialization. The rationale behind P-REC9 seems to promote the use of standardised ontology implementation languages for defining semantic artefacts and for P-REC11 to extend them with more complex languages when the former are

[15] `https://www.w3.org/Submission/SWRL/`

not enough. Hence our proposal would be to merge both recommendations into one proposing the use of standardized languages like RDF(S) and OWL for implementing ontologies, extending them with SHACL for constraint definitions if applicable, and using SKOS [2] for the implementation of thesauri. Some mentions may then be included to other initiatives, not yet standardized, like SWRL, or SheX.[16]

Furthermore, we have additional comments related to some other principles.

- F3 encourages making clear and explicit references from the metadata to the data. This is poorly addressed by the guidelines, being absent from the Semantic Web oriented guidelines (FAIR ontology, and the 5-stars schemas). This may be a consequence of the fact that in the Semantic Web, ontology metadata is commonly embedded in the ontology itself and not as a first-class citizen, and would be retrieved by looking up the ontology URIs, therefore there is no clear need for this link.
- A1.2 and A2 are also lacking guidelines. On the one hand, A1.2 is not described in the 5-stars2012 because it is assumed that the vocabulary will be open (star 1). In addition, all the Semantic Web oriented guidelines assume HTTP and HTTPS as protocols to share the semantic artefacts. On the other hand, the absence of A2 is related to the fact that usually ontologies themselves contain their metadata together in a unique artefact, as discussed above.

Finally, it should be mentioned that the fifth star from the schema proposed by [21] is not related to any of the principles. The reason is that the star states that the vocabulary should be linked to by other vocabularies and this is a measure of the success of the vocabulary after being published rather than recommendation or an action to be taken by the developers or publishers. That is, even though it is related to interoperability, it is not related to any principle in particular as there is no equivalent principle stating that the data should be linked back from other data.

5 Towards FAIR Ontology Engineering Practices

This Section aims at providing a summary of the items that we consider that should be further discussed by the Ontology Engineering (OE) and Open Science (OS) communities, so as to propose our contributions towards a unified recommendation on how to make ontologies FAIR.

To be Findable

The F1 principle refers to using globally unique and persistent identifiers. In the OE community URIs are already used to refer to one ontology or SKOS schema, and sometimes for their elements as well. This practice complies with the "unique" definition of FAIR, which means that an identifier refers to only one entity. It is worth noting that the use of "unique" in the FAIR principles is different from (and compatible with) the meaning of "unique" in the non-unique naming assumption used in OWL, which means

[16] `http://shex.io/shex-primer/`

that one entity may be identified by more than one name. Regarding persistence, even though there are good practices and services (w3id or purl) for generating permanent URIs, no strict rules are defined to ensure persistence and no mechanisms as the use of DOIs are established to persist URIs. The Semantic Web community background on the Web of documents has modelled the practitioners to understand and work with the Web as a living ecosystem, where resources may disappear, in contrast to other communities that are more oriented to archiving and preservation practices. In order to align this principle to the publication of semantic artefacts, the following questions should be subject to discussion: Should the Semantic Web community establish mechanisms and authorities to coin persistent identifiers (PIDs) for semantic artefacts? Should these PIDs refer only to semantic artefacts as a whole or also to each of their components (e.g., specific concepts or properties, specific SKOS concepts)?

The F2 principle refers to describing data with rich metadata. As documented in [20] F2 refers to metadata to allow for data findability in contrast to metadata to improve its reusability, which is mentioned in principle R1. In this sense the OE community should agree on a minimum set of metadata that semantic artefacts should have. This does not imply imposing a specific vocabulary, but defining which attributes (e.g., license, title, creators, etc.) the community considers as crucial for ensuring findability of a semantic artefact. For example, the WIDOCO Best Practices[17] recommend stating the creator(s) of an ontology, which can be identified by using `dcterms:creator`, `dc:creator`, `schema:creator`, `prov:wasAttributedTo` or `pav:createdBy`. In this sense, DCAT or Dublin Core should be considered as reference vocabularies for providing metadata, however some communities might use their own common vocabularies. Finally, it is also needed to provide more practical guidelines for declaring metadata, for example generating templates, of how these annotations are implemented in each case and defining clearly what is embedded in semantic artefacts, for example in OWL ontologies.

Nowadays, the F3 principle is not applicable to ontologies because in practice they contain the metadata that describes them, both as a resource and for each ontology element defined. Therefore, the question here is in which cases metadata should be provided as a separate object? This principle might not be applicable from the Semantic Web perspective unless we refer to metadata assets managed by third-party applications like ontology indexes and registries rather than the metadata provided by ontology publishers.

F4 suggests that data and metadata are indexed in searchable resources. While there are general ontology registries and community or domain oriented ones, a federation model for ontologies should be defined. Regarding repositories and search engines that would be needed to find semantic artefacts, P-REC6 proposes to build search engines to operate across distributed and heterogeneous repositories. However no existing recommendations are listed for this. For doing this some federation models existing for data as for example the European Data Portal,[18] based on DCAT, or the JoinUp initiative, based on ADMS, may be considered as examples. For the semantic artefacts case the DCAT2 vocabulary[19] may be used for the federation system. This federation mecha-

[17] `https://w3id.org/widoco/bestPractices`
[18] European Data Portal `https://www.europeandataportal.eu/en`
[19] `https://www.w3.org/TR/vocab-dcat-2`

nism would be closely related to the F2 principle regarding the agreement on metadata for findability. Other practice to be taken into account is the inclusion of metadata in the form of JSON-LD [30] snippets within the HTML describing ontologies in order to be indexed by web search engines, as it is currently done by WIDOCO [10] and Agroportal [22]. Finally, any of these federation approaches may be combined with the idea of de-centralized web exposed in [33] in which each semantic artefact owner will store and manage the data about the published artefact to be integrated by third party registries or applications. Standard definitions of SAODs (Semantic Artefact Online Data[20]) should be created as well as SAODs discovery approaches.

To be Accessible

Ontologies published following the Semantic Web technologies and best practices use HTTP URIs as identifiers and are shared under HTTP or HTTPS protocols; complying with A1, A1.1, A1.2 principles. These already existing technologies and protocols are suggested to be adopted by FAIR implementations.

The A2 principle requires keeping metadata accessible even when the data is no longer available. This principle clashes with the (Semantic) Web aspect where resources as ontologies may become unavailable at any moment, as it happens for websites. Complying to this principle would involve developing registries or infrastructures to act as ontology libraries, to preserve the metadata. From the Semantic Web perspective, having preservation policies (for example how long a semantic artefact will be preserved, what version will be retained, what serialization formats will be stored, etc.) for publishing resources may be a good practice to adopt [1].

To be Interoperable

To be compliant with the I1 principle, semantic artefacts should use knowledge representation languages proposed by a standardization body, such as W3C. To this end, as commented in Section 3 in regards with P-REC9 and P-REC11, well-known W3C recommendations like RDF(S) and OWL are used for implementing ontologies, and SKOS [2] for thesauri. In addition, SHACL may be used to extend ontologies with additional data constraints definitions.

The I2 principle states that (meta)data should use vocabularies that follow FAIR principles. An attempt to translate this principle to ontologies would be recommending the reuse of FAIR semantic artefacts to the extent possible, in addition to the common practice about reusing ontologies that follow best practices and Linked Data principles. This also applies to the reuse of other ontologies for annotating ontology metadata. This leads us to the need of indicators that describe compliance with FAIR principles in order to decide whether an ontology is FAIR, such as the ones proposed by the RDA maturity model [12]. Therefore, validators should be developed to automatically compute these indicators, such as proposed in [36]. However, this principle should not force to reuse only (and at least one) FAIR vocabulary, as circular references would appear, that is,

[20] Acronym adapted from the PODs defined in [33] as Personal Online Data

if a vocabulary should (re)use other FAIR vocabularies, how would be the first FAIR vocabulary be considered as such?

In order to comply with I3, ontologies should include qualified references to other ontologies. The Semantic Web technologies already provide a number of mechanisms to refer to other ontologies. When referring to another ontology element URI the reference is explicit and in addition the relations could be explicit by using `owl:equivalentClass`, `owl:equivalentProperty` or the different relations for SKOS concepts. Finally, the `owl:import` construct also allows for referring to (and importing) other ontologies explicitly and in a machine readable way.

To be Reusable

The minimum set of metadata mentioned in F2 should also contain the minimum attributes to assess whether a semantic artefact is appropriate for reuse as required by R1. For example, provenance, term detailed descriptions (usually included in the ontologies by using `rdfs:comment` annotations), rationales behind the inclusion of terms, examples of use, etc. In addition, the community should suggest vocabularies that could be used to represent such fields and the mappings between such vocabularies. Ontologies should rely on the human oriented complementary documentation such as examples of use and diagrams of the conceptualizations to ease the task of understanding the model represented in the code to potential users. Therefore, there is a need for research towards best practices to document and communicate ontologies.

Taking into account that FAIR advocates for the reuse of data as much as possible, it is advisable to provide minimum information about the permissions and conditions included in the licenses of semantic artefacts to be considered FAIR compliant with R1.1. Also, such license descriptions should be linked from the resources and provided in RDF. This could be done in two ways. The simplest way would be providing a link to the applicable license URI, which in the best case scenario would be described in RDF. A more complete way would be providing the RDF description of the license (what it is allowed, or not, and under which conditions) using vocabularies as the Creative Commons vocabulary[21] or ODRL [31].

To comply with principle R1.2 the W3C already provides the PROV-O ontology and standard specification [25] that should be adopted.

Meeting domain-relevant standards, as defined in R1.3, might refer to technological ones like the use of RDF(S) and OWL to describe ontologies as already proposed in I1. However, standards may involve another aspects which will depend on the communities. For example, in the OBO community there is a standard way of naming ontology elements while in the Semantic Web community the rule is to keep the naming convention, whichever is chosen, consistent. This principle is also related to the minimum set of metadata already defined in several communities [20]. Therefore, there is a need here for each community to agree on common standards and best practices to follow in regard to ontology engineering.

[21] `https://creativecommons.org/ns`

Summarizing, to pave the path for FAIR semantics publishing, understanding and exploitation, the OE community needs to:

- Agree on a minimum set of metadata suggesting vocabularies to represent it and provide more technical guidelines for its declaration.
- Define a federation model for ontologies that may be combined with standard definitions of SAODs as well as SAODs discovery approaches.
- Define and adopt preservation policies for publishing resources together with mechanisms to determine whether this preservation is fulfilled.
- Use knowledge representation languages from standardization bodies.
- Define FAIR indicators for semantic artefacts.
- Define best practices to document and communicate ontologies.

Finally, the following questions remain open for discussion: 1) should the Semantic Web community establish mechanisms and authorities to coin persistent identifiers (PIDs) for semantic artefacts? and 2) in which cases metadata should be provided as a separate object and whether to define third party certification agencies is needed?

Acknowledgments

This work has been supported by a Predoctoral grant from the I+D+i program of the Universidad Politécnica de Madrid and the Spanish project DATOS 4.0: RETOS Y SOLUCIONES (TIN2016-78011-C4-4-R). Authors would like to thank Yann Le Franc for his clarifications and explanations about the FAIRsFAIR recommendations content and their development process and OEG, especially Victor Rodríguez Doncel, for all the valuable comments.

References

1. Baker, T., Vandenbussche, P.Y., Vatant, B.: Requirements for vocabulary preservation and governance. Library Hi Tech **31**(4), 657–668 (2013)
2. Bechhofer, S., Miles, A.: SKOS simple knowledge organization system reference. W3C recommendation, W3C (2009)
3. Bechhofer, S., Van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D.L., Patel-Schneider, P.F., Stein, L.A., et al.: OWL web ontology language reference. W3C recommendation **10**(02) (2004)
4. Bizer, C., Heath, T., Berners-Lee, T.: Linked data: The story so far. In: Semantic services, interoperability and web applications: emerging concepts, pp. 205–227. IGI Global (2011)
5. Borgo, S., Masolo, C.: Foundational choices in DOLCE. In: Handbook on ontologies, pp. 361–381. Springer (2009)
6. Brickley, D., Guha, R.V., McBride, B.: RDF Schema 1.1. W3C recommendation **25** (2014)
7. Collins, S., Genova, F., Harrower, N., Hodson, S., Jones, S., Laaksonen, L., Mietchen, D., Petrauskaitė, R., Wittenburg, P.: Turning FAIR into reality: Final report and action plan from the European Commission expert group on FAIR data (Nov 2018). https://doi.org/10.2777/54599
8. Corcho, O., Eriksson, M., Kurowski, K., Ojsteršek, M., Choirat, C., van de Sanden, M., Coppens, F.: EOSC interoperability framework (May 2020), `https://www.eoscsecretariat.eu/sites/default/files/eosc-interoperability-framework-v1.0.pdf`

9. Côté, R.G., Jones, P., Apweiler, R., Hermjakob, H.: The Ontology Lookup Service, a lightweight cross-platform tool for controlled vocabulary queries. BMC bioinformatics **7**(1), 97 (2006)
10. Garijo, D.: Widoco: a wizard for documenting ontologies. In: International Semantic Web Conference. pp. 94–102 (2017)
11. Garijo, D., Poveda-Villalón, M.: Best practices for implementing FAIR vocabularies and ontologies on the Web (Mar 2020), `https://arxiv.org/abs/2003.13084`
12. Group, F.D.M.M.W.: FAIR Data Maturity Model: specification and guidelines (Apr 2020). https://doi.org/10.15497/RDA00045, `https://doi.org/10.15497/RDA00045`
13. Guha, R.V., Brickley, D., Macbeth, S.: Schema. org: evolution of structured data on the web. Communications of the ACM **59**(2), 44–51 (2016)
14. Guizzardi, G.: Ontological foundations for structural conceptual models (2005)
15. Guizzardi, G., Wagner, G.: Towards ontological foundations for agent modelling concepts using the unified fundational ontology (UFO). In: International Bi-Conference Workshop on Agent-Oriented Information Systems. pp. 110–124. Springer (2004)
16. Hartmann, J., Sure, Y., Haase, P., Palma, R., Suarez-Figueroa, M.: Omv–ontology metadata vocabulary. In: ISWC. vol. 3729 (2005)
17. Hasnain, A., Rebholz-Schuhmann, D.: Assessing FAIR data principles against the 5-star Open Data principles. In: European Semantic Web Conference. pp. 469–477. Springer (2018)
18. Heath, T., Bizer, C.: Linked data: Evolving the web into a global data space. Synthesis lectures on the semantic web: theory and technology **1**(1), 1–136 (2011)
19. Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P.F., Rudolph, S., et al.: OWL 2 Web Ontology Language Primer. W3C recommendation **27**(1), 123 (2009)
20. Jacobsen, A., de Miranda Azevedo, R., Juty, N., Batista, D., Coles, S., Cornet, R., Courtot, M., Crosas, M., Dumontier, et al.: FAIR Principles: Interpretations and Implementation Considerations. Data Intelligence **2**(1-2), 10–29 (2020). https://doi.org/10.1162/dint_r_00024
21. Janowicz, K., Hitzler, P., Adams, B., Kolas, D., Vardeman, I., et al.: Five stars of linked data vocabulary use. Semantic Web **5**(3), 173–176 (2014)
22. Jonquet, C., Toulet, A., Arnaud, E., Aubin, S., Yeumo, E.D., Emonet, V., Graybeal, J., Laporte, M.A., Musen, M.A., Pesce, V., et al.: Agroportal: A vocabulary and ontology repository for agronomy. Computers and Electronics in Agriculture **144**, 126–143 (2018)
23. Knublauch, H., Kontokostas, D.: Shapes constraint language (SHACL). World Wide Web Consortium recommendation (2017)
24. Le Franc, Y., Parland-von Essen, J., Bonino, L., Lehväslaiho, H., Coen, G., Staiger, C.: D2.2 FAIR semantics: First recommendations (Mar 2020). https://doi.org/10.5281/zenodo.3707985
25. Lebo, T., Sahoo, S., McGuinness, D., Belhajjame, K., Cheney, J., Corsar, D., Garijo, D., Soiland-Reyes, S., Zednik, S., Zhao, J.: Prov-o: The prov ontology. W3C recommendation (2013)
26. Lehväslaiho, H., Parland-von Essen, J., Behnke, C., Laine, H., Riungu-Kalliosaari, L., Le Franc, Y., Staiger, C.: D2.1 Report on FAIR requirements for persistence and interoperability 2019 (Nov 2019). https://doi.org/10.5281/zenodo.3557381
27. McGuinness, D.L.: Ontologies come of age. In: Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential, pp. 171–194. MIT press Cambridge, MA (2002)
28. Palma, R., Haase, P.: Oyster–sharing and re-using ontologies in a peer-to-peer community. In: International Semantic Web Conference. pp. 1059–1062. Springer (2005)
29. Poveda-Villalón, M., García-Castro, R., Gómez-Pérez, A.: Building an ontology catalogue for smart cities. In: Proceedings of the 10th European Conference on Product and Process Modelling, ECPPM 2014. pp. 1–8 (2014)
30. Sporny, M., Longley, D., Kellogg, G., Lanthaler, M., Lindström, N.: JSON-LD 1.0: a JSON-based serialization for linked data. World Wide Web Consortium recommendation (2014)

31. Steidl, M., Iannella, R., Myles, S., Rodríguez-Doncel, V.: ODRL vocabulary & expression 2.2. W3C recommendation, W3C (Feb 2018)
32. Vandenbussche, P.Y., Atemezing, G.A., Poveda-Villalón, M., Vatant, B.: Linked Open Vocabularies (LOV): A gateway to reusable semantic vocabularies on the Web. Semantic Web **8**(3), 437–452 (Jan 2017)
33. Verborgh, R.: Re-decentralizing the Web, for good this time. In: Seneviratne, O., Hendler, J. (eds.) Linking the World's Information: Tim Berners-Lee's Invention of the World Wide Web. ACM (2020), `https://ruben.verborgh.org/articles/redecentralizing-the-web/`
34. Whetzel, P.L., Noy, N.F., Shah, N.H., Alexander, P.R., Nyulas, C., Tudorache, T., Musen, M.A.: Bioportal: enhanced functionality via new web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. Nucleic acids research **39**(suppl_2), W541–W545 (2011)
35. Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E., et al.: The FAIR Guiding Principles for scientific data management and stewardship. Scientific data **3** (2016)
36. Wilkinson, M.D., Dumontier, M., Sansone, S.A., da Silva Santos, L.O.B., Prieto, M., Batista, D., McQuilton, P., Kuhn, T., Rocca-Serra, P., et al.: Evaluating FAIR maturity through a scalable, automated, community-governed framework. Scientific data **6**(1), 1–12 (2019)
37. Xiang, Z., Mungall, C., Ruttenberg, A., He, Y.: Ontobee: A linked data server and browser for ontology terms. In: ICBO (2011)

A Annex: FAIR principles

The list of FAIR guiding principles defined in [35] is:

- To be Findable
  - F1. (meta)data are assigned a globally unique and persistent identifier
  - F2. data are described with rich metadata (defined by R1 below)
  - F3. metadata clearly and explicitly include the identifier of the data it describes
  - F4. (meta)data are registered or indexed in a searchable resource
- To be Accesible
  - 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
- To be Interoperable
  - 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
- To be Reusable
  - 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 license
  - R1.2. (meta)data are associated with detailed provenance
  - R1.3. (meta)data meet domain-relevant community standards