FAIR or FAIRer? An integrated quantitative FAIRness assessment grid for semantic resources and ontologies

Emna Amdouni, Clement Jonquet

FAIR or FAIRer? An integrated quantitative FAIRness assessment grid for semantic resources and ontologies

Emna Amdouni[1 [0000-0002-2930-5938]] and Clement Jonquet[1,2 [0000-0002-2404-1582]]

[1] LIRMM, University of Montpellier & CNRS, France
[2] MISTEA, University of Montpellier, INRAE & Institut Agro, France
emna.amdouni@lirmm.fr and jonquet@lirmm.fr

Abstract. In open science, the expression "FAIRness assessment" refers to evaluating to which degree a digital object is Findable, Accessible, Interoperable, and Reusable. Standard vocabularies or ontologies are a key element to achieving a high level of FAIRness (FAIR Principle I2) but as with any other data, ontologies have themselves to be FAIR. Despite the recent interest in the open science and semantic Web communities for this question, we have not seen yet a quantitative evaluation method to assess and score the level of FAIRness of ontologies or semantic resources in general (e.g., vocabularies, terminologies, thesaurus). The main objective of this work is to provide such a method to guide semantic stakeholders in making their semantic resources FAIR. We present an integrated quantitative assessment grid for semantic resources and propose candidate metadata properties –taken from the MOD ontology metadata model– to be used to make a semantic resource FAIR. Aligned and nourished with relevant FAIRness assessment state-of-the-art initiatives, our grid distributes 478 credits to the 15 FAIR principles in a manner which integrates existing generic approaches for digital objects (i.e., FDMM, SHARC) and approaches dedicated to semantic resources (i.e., 5-stars V, MIRO, FAIRsFAIR, Poveda et al.). The credits of the grid can then be used for implementing FAIRness assessment methods and tools.

Keywords: FAIR data principles, FAIRness assessment, evaluation grid, semantic Web, ontologies, semantic resources / artefacts, metadata properties.

1 Introduction

In 2014, a group of researchers, research institutions, and publishers (called FORCE 11) defined fundamental guiding principles called FAIR (for Findable, Accessible, Interoperable, and Reusable) to make scientific data and their metadata interoperable, persistent, and understandable for both humans and machines [1]. The FAIR principles emphasize the importance of semantic technologies in making data interoperable and reusable. However, ontologies[1] –the backbone of semantic technologies– have themselves to be FAIR. Until recently, not much attention has been made to quantitatively evaluating ontologies using FAIR principles; all related work or state-of-the-art methods regarding ontologies are qualitative i.e., proposing

[1] In this paper, we will consider the terms ontologies, terminologies, thesaurus and vocabularies as a type of knowledge organization systems [2] or knowledge artefacts [3] or semantic resources [4]. For simplicity, we will sometimes use "ontology" as an overarching word.

recommendations and best practices without providing a scoring mechanism. It is clear that the development of FAIRness assessment methods –i.e., ways to measure to which level a digital object implements FAIR principles– remains challenging [5], including for ontologies and semantic resources. In fact, the complexity of FAIRness assessment is due to the fact that the FAIR principles are expressed at a very generic level and need to be expanded and projected to specific digital objects to be more explicit. Furthermore, some criteria are very hard to evaluate by a program and sometimes require subjective–human expertise.

For all these reasons, we believe it is essential to define a quantitative method i.e., a metric, for assessing and scoring to which degree a semantic resource is FAIR compliant –for example, to determine if a resource is "not FAIR", "FAIR" or even "FAIRer" than a certain threshold or another resource. The objective of this work is to provide a *grid* dispatching different values of *credits* to each FAIR principle, depending on its importance when assessing the FAIRness of ontologies. We talked about an *integrated grid*, to capture that our grid is aligned and nourished by existing generic approaches for digital objects in general (i.e., FDMM, SHARC) and approaches dedicated to semantic resources or artifacts (i.e., 5-stars V, MIRO, FAIRsFAIR, Poveda et al.). As a result, the proposed grid involves ***478 credits*** that can be used for implementing FAIRness assessment tools. Such tools will then guide semantic stakeholders in (i) making their semantic resources FAIR and (ii) selecting relevant FAIR semantic resources for use.

The rest of this paper is organized as follows: Section 2 presents related work in FAIRness assessment or alike. Section 3 describes the methodology followed to integrate the most prominent existing works into one schema and details the proposed FAIRness assessment grid. Section 4 presents candidate metadata properties –taken from the MOD 1.4 ontology metadata model– to be used to enable FAIRness assessment. Finally, Section 5 concludes and gives our perspective of developing a methodology to automatize FAIRness assessment.

2 Overview of related work for assessing FAIRness

We distinguish between two FAIRness assessment approaches: the first category concerns general schemes or generic tools applicable for any kind of digital object; the second category is specific for the description and assessment of ontologies or semantic resources. We review both of them chronologically.

2.1 Generic FAIRness assessment approaches

The Research Data Alliance (RDA) *SHARing Rewards and Credit* (SHARC) Interest Group, created in 2017, proposed a FAIRness assessment grid to enable researchers and other data stakeholders to evaluate FAIR implementations and provide the appropriate means for crediting and rewarding to facilitate data sharing [6]. The SHARC grid defines a set of 45 generic criteria with importance levels (essential, recommended, or desirable) evaluated by answering one of four values (Never/NA, If mandatory, Sometimes, Always) to a question; questions are sometimes dependent on one another as in a decision tree.

In 2018, the RDA *FAIR Data Maturity Model* (FDMM) Working Group recommended normalizing FAIRness assessment approaches and enabling comparison of their results [7]. It describes a set of 47 generic criteria derived from each FAIR principle with priorities (essential, important, or useful). Both the SHARC grid and the FDMM recommendation assumed that some FAIR principles were more important than others. We have kept this philosophy in our methodology and kept the SHARC and FDMM outputs as influences on our FAIRness assessment score.

Some FAIRness assessment tools recently appeared, such as FAIRdat tool[8], FAIR self-assessment tool [9], OzNome 5-star tool [10], FAIR Metrics [11], FAIR-Aware [12], F-UJI[2]. We cannot unfortunately detail them here. These tools are relevant but need to be improved in order to provide a clear methodology and a tool to assess any digital object quantitatively.

2.2 Specific FAIRness assessment approaches

Before the emergence of the FAIR principles in 2011, Berners-Lee presented the foundational principles for Linked Open Data (LOD) [13] for making data available, shareable, and interconnected on the Web. The FAIR principles have been proposed for similar reasons with a stronger emphasis on data reusability (consideration of license agreement and provenance information). The 5-stars LOD principles were specialized in 2014 for Linked data vocabularies [14] as five rules to follow for creating and publishing "good" vocabularies. Under this scheme, stars denote the quality, leading to better structure (i.e., use of W3C standards) and interoperability for reuse (i.e., metadata description, reuse of vocabularies, and alignment). The proposed 5-star rating system (later called *5-stars V*) for vocabularies is simple. However, no implementation tool was developed for making the assessment automatic, and the principles are not largely referenced today. A study of the degree to which the FAIR principles align, and extend the 5-star LOD principles was proposed first in [15] and later in [16]; we have incorporated this alignment in our methodology.

In 2017, the *Minimum Information for Reporting an Ontology* initiative published the MIRO guidelines for ontology developers when reporting an ontology in scientific reports [17]. The MIRO guidelines aim to improve the information content descriptions' quality and consistency, including development methodology, provenance, and context of reuse information. They define 34 information items (such as "ontology name", "ontology license", "ontology URL") and specify the level of importance "must", "should', "optional" for each. This work was significant, but there have been no studies on how the MIRO properties align with or extend the FAIR principles. However, the MOD 1.4 metadata model (see hereafter) aligned each MIRO guideline and the corresponding metadata properties in MOD. We, therefore, used this alignment in our methodology to influence the FAIRness assessment score with the MIRO guidelines.

Dutta et al. [18] reviewed and harmonized existing metadata vocabularies and proposed a unified ontology metadata model called MOD (for *Metadata for Ontology Description*) to facilitate manual and automatic ontology descriptions, identification, and selection. MOD is not another standard nor another metadata vocabulary, but more a set

[2] https://seprojects.marum.de/fuji/api/v1/ui/

of cataloged and regrouped properties one can use to describe a semantic resource. For instance, MOD does not require the use of a specific authorship property but rather encodes that `dc:creator`; `schema:author`, `foaf:maker`, or `pav:createdBy` can be used to say so. The MOD 1.2 model later extended in MOD1.4[3] was used in AgroPortal –a vocabulary and ontology repository for agronomy– to implement a richer, unified metadata model [19]. With this implementation, AgroPortal affirms to recognize 346 properties from 15 relevant metadata vocabularies (such as Dublin Core, Ontology Metadata Vocabulary, VoID, FOAF, Schema.org, PROV-O, DCAT, etc.) and map them to its unified model. Somehow, this previous work on a unified metadata model could be considered as the first step for enabling FAIRness assessment. For example, an ontology developer can focus on his/her responsibility of determining the license to use an ontology, while MOD offers means and recommendations to encode such information in a way machines can assess the level of FAIRness. Based on the MOD model, we produce in this article guidelines on how FAIR principles might be met and evaluated. Section 4 provides a clear alignment between the MOD properties and the FAIR principles. For instance, to assess F1, we rely on the existing MOD properties to encode the identifiers of an ontology (`omv:uri`) and (`dct:identifier`).

In March 2020, the FAIRsFAIR H2020 project delivered the first version of a list of 17 recommendations and 10 best practices recommendations for making semantic artefacts FAIR [3] (later revised in Dec. 2020 in a new deliverable [19]). For each recommendation, the authors provided a detailed description, listed its related semantic Web technologies, and outlined potential technical solutions in some cases. Similarly, best practices are introduced as recommendations that are not directly related to a FAIR principle but contribute to the overall evaluation of a semantic resource. This proposal is currently being discussed in the context of the RDA Vocabulary and Semantic Services Interest Group (VSSIG). The recommendations are also publicly available for comments on GitHub.[4]

Later, in September 2020, Poveda et al. considered some previous works and produced "guidelines and best practices for creating accessible, understandable and reusable ontologies on the Web" [16]. In another position paper [20], they complete a qualitative analysis of how four ontology publication initiatives cover the foundational FAIR principles. They propose some recommendations on making ontologies FAIR and list some open issues that might be addressed by the semantic Web community in the future. These two publications are very relevant for our methodology; our work is a step further. It completes this work and proposes a concrete metric necessary for further work on automatic FAIRness assessment.

Other recent related works on FAIR principles for semantic resources include a list of functional metrics and recommendations for Linked Open Data Knowledge Organization Systems (LOD KOS) products proposed in 2020 [21], a list of ten simple rules for making a vocabulary FAIR [22]. Finally, the DBPedia Archivo tool [23], an ontology archive also released at the end of 2020 that aims to help developers and consumers in "implementing FAIR ontologies on the Web."

[3] https://github.com/sifrproject/MOD-Ontology
[4] https://github.com/FAIRsFAIR-Project/FAIRSemantics/issues/

To design our FAIRness assessment methodology, we analyzed and merged relevant related approaches namely FDMM version v0.04, SHARC version v1.1, LOD 5-stars V, MIRO, FAIRsFAIR recommendations, and Poveda et al.'s guidelines. We consider both generic and specific approaches to provide a specialized solution for ontologies but are still influenced by more general concerns, as ontologies are a kind of digital object. The integration was not straightforward because none of the approaches used is simply and strictly aligned with the 15 sub-principles (e.g., FDMM provides 47 criteria). Two of them (i.e., MIRO and 5-stars V) were totally disconnected from the FAIR prism. Table 1 gives a summary of our selection. We classify approaches into three groups: (A) for generic approaches which set priorities for each FAIR principle or sub-principle; for specific approaches for semantic resources which: (B) includes FAIRsFAIR and Poveda et al. as these guidelines do not set priorities; (C) includes LOD 5-stars and MIRO as they are not aligned to the FAIR principles. In the next section, we explain how we proceeded to integrate all these methodologies into the proposed grid.

Table 1. Summary of related works on FAIRness assessment integrated into our approach.

Category	Generic (A)		Specific (B, C)			
Format	grid		principles	scheme	recommendations	
Approach	SHARC	FDMM	5-stars V	MIRO	FAIRsFAIR	Poveda et al.
Reference	[6]	[7]	[14]	[17]	[3]	[20]
Year	2017	2018	2011	2017	2020	2020
FAIR principles	after	after	before	before	after	after
Priorities	yes	yes	n/a	n/a	no	no

3 Integrated quantitative FAIRness assessment grid for ontologies

3.1 Methodology

In what follows, we explain how we quantify each approach against the FAIR principles, then for each category (A) and (B-C), and finally determine a set of final FAIR credits that could be used in evaluating any semantic resource.

We chose to provide numerical credits {0;1;2;3} to respectively represent the degree of priorities/qualification of each *indicator* (other names for sub-principle e.g., F1, F2, F3 et F4 for F) within an approach {e.g., “none”, “may”, “should”, “must”}. A “must” represents an essential principle, a “should” means that a principle is important except under some particular circumstance, “may” is an optional requirement, and “none” a non-revealed/specified qualification of a principle. Table 2 lists the correspondences between the five approaches, their priorities, and their attributed credits.

Table 2 Alignment between priorities in related work approaches and credits.

Group	**Approach**	**none (0 credit)**	**may (1 credit)**	**should (2 credits)**	**must (3 credits)**
A	**FDMM**	n/a	useful	important	essential
	SHARC	n/a	desirable	recommended	essential
B	**FAIRsFAIR**	n/a	1 Rec	2 Rec	3 Rec or >
	Poveda et al.	n/a	1 Rec	2 Rec	3 Rec or >
C	**MIRO**	n/a	optional	should	must

	5-stars V	n/a	1 star	2 stars	3 stars or >

To determine the FAIR principle credits for each group, we follow certain rules:

- Group A: we calculate the approximate average value of credits per indicator (FDMM) or principle (SHARC). For SHARC, we divide the obtained average value by the number of indicators associated with a principle.
- Group B: we count the number of recommendations to determine the credits.
- Group C: we count the number "must", "should" and "optional" properties for MIRO and the number of the principles for 5-stars V.

The final credits for each sub-principle are the sum of all obtained credits per methodology. An example is provided hereafter:

Example 1: We illustrate how we determine for each group (i.e., A, B and C) the credits of F1 (noted $Credits_{F1}$):

Group A:

- FDMM defines 4 "essential" indicators (F1-01M, F1-01D, F1-02M, and F1-02D). Thus, $Credits_{F1,FDMM} = 3 * 4 = 12$.
- SHARC defines 12 sub-indicators (8 essential, 4 recommended) for F. Thus the approximative $Credits_{F1,SHARC} = (8 * 3 + 4 * 2) \div 4 = 8$.

Group B:

- FAIRsFAIR defines 2 recommendations (P-Rec 1 and P-Rec 2) related to F1 thus, $Credits_{F1,FsF}$=2.
- Poveda et al. define 4 recommendations related to F1 (Rec 1, Rec 2, Rec 3 and Rec 5) thus $Credits_{F1,Poveda\ et\ al.} = 3$.

Group C:

- MIRO refers to 2 "must" properties ("A" category- basics) for F1 sub-principle: ontology version (A.4) and IRI version (A.4). Thus, $Credits_{F1,MIRO}$=6.
- LOD 5-stars V does not especially cover Findability; thus, $Credits_{F1,5-starsV}$=0.

3.2 Results

From a semantic Web perspective, the results obtained emphasize the need for the establishment of agreement about a set of core metadata ontology descriptions, a federation model for ontologies regarding repositories and search engines, clear ontology and metadata ontology perseveration strategies within endpoints, mechanisms for references qualification, and best practices to document and communicate ontologies. Fig. 1 (c) provides final integrated FAIR credits per indicator; it shows how both generic approaches and semantic resources specific approaches address FAIRness and scores each FAIR indicator.
For example, $Credits_F = 113$. Which is the sum of 41 (F1)+27 (F2)+ 21(F3)+24(F4).

Fig. 1 (c) illustrates the importance of each indicator in our integrated method. When doing the final sums, we have chosen a baseline value fixed to 10, to represent the fact that originally, as suggested by the FORCE 11 group, the FAIR principles were not ordered by importance; they were supposed to all contribute equally. The final credits are presented in our integrated FAIRness assessment grid (Table 3); the 478 credits of the grid, dispatched by each sub-principle, can be used for the assessment of any semantic resource or ontologies.

Table 3. Integrated FAIRness assessment grid for semantic resources and ontologies.

Principle		Baseline	SHARC	FDMM	5-stars V	MIRO	FAIRs FAIR	Poveda et al.	Credits	
F	**F1**	10	8	12	0	6	2	3	**41**	**113**
	F2	10	8	2	0	5	1	1	**27**	
	F3	10	8	2	0	0	1	0	**21**	
	F4	10	8	3	0	0	2	1	**24**	
A	**A1**	10	6	18	3	3	1	2	**43**	**113**
	A1.1	10	6	11	0	0	1	0	**28**	
	A1.2	10	6	5	0	0	1	0	**22**	
	A2	10	6	3	0	0	1	0	**20**	
I	**I1**	10	4	12	1	12	3	2	**44**	**109**
	I2	10	4	7	0	9	2	0	**32**	
	I3	10	4	12	1	3	2	1	**33**	
R	**R1**	10	9	3	1	6	0	3	**32**	**143**
	R1.1	10	9	12	0	3	2	1	**37**	
	R1.2	10	9	3	0	12	3	1	**38**	
	R1.3	10	9	12	0	0	3	2	**36**	
Total credits										**478**

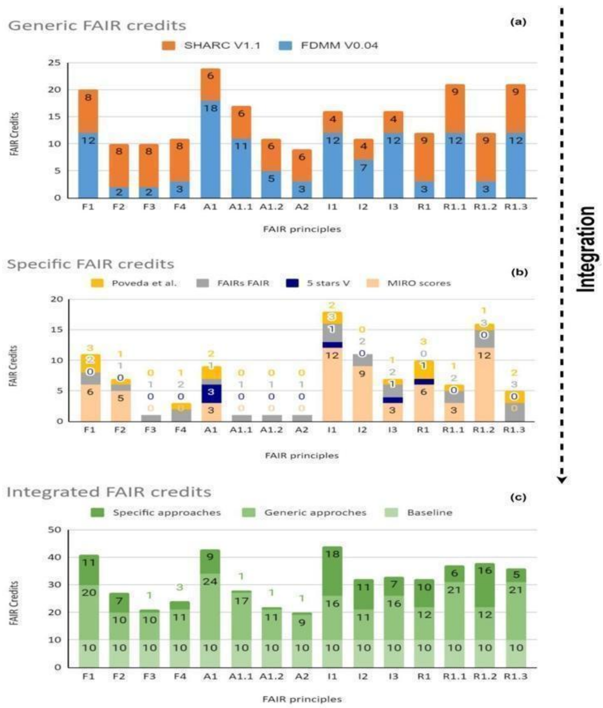

Fig. 1. Credits are assigned to each FAIR principle by generic approaches (a), specific approaches (b), and sums with a common baseline in our integrated grid (c).

A quick analysis of Table 3 and Fig. 1 reveals interesting points:

- The most important principles for generic and specific approaches are not the same. Generic approaches tend to emphasize principles F1 (identifier), A1 (access protocol), R1.1 (license), and R1.3 (community standards), while specific approaches emphasize principles I1 (knowledge representation), R1.2 (provenance), and I2 (use of vocabularies). This confirms our hypothesis that being FAIR is strongly dependent on the type of digital object considered and therefore FAIRness assessment methods must be customized for each type.
- In the integrated grid, F1, A1, and I1 are the three sub-principles with the higher number of credits. These aspects being "generally" well addressed by ontologies, it will contribute to an overall good level of FAIRness.
- Four sub-principles, important for FAIR, were completely ignored/avoided by specific approaches, except the FAIRsFAIR recommendations: F3 (link data-metadata), A1.1 (protocol openness), A1.2 (protocol security), and A2 (long term metadata). Consequently, three of these four keep the minimum number of credits in the integrated grid.
- None of the specific approaches covered all of the FAIR sub-principles. This is not surprising for MIRO and 5-stars V, which preexist the FAIR movement, but it is more surprising for FAIRsFAIR and Poveda et al. whose recommendations were done specifically for ontologies or semantic resources to be FAIR. Only A1, I1, and I3 were found in the four approaches studied. This point backups our methodology, which mixes both generic and specific approaches.
- Despite differences in credits assigned to the sub-principles, the sums by principles are relatively close, with a mean of 119,5. Only the R group is significantly above the mean. The group I is slightly under, mainly because it is made of only three sub-principles instead of four.
- R being the most important principle may reveal the concern that ontologies and semantic resources, often developed by means of semantic Web technologies (RDFS, OWL, SKOS) are naturally equipped with good findability, accessibility, and interoperability features (e.g., URIs for identifiers, HTTP for accessibility, W3C standards for knowledge representation, claim to use vocabularies, etc.) whereas they lack reusability.

4 Candidate metadata properties for FAIR ontologies

In the second phase of our work, we elicited candidate metadata properties that can be used to encode information relevant for each FAIR sub-principle. Indeed, we found out most sub-principles (about 93%) might be partially or totally implemented and assessed with a series of metadata properties. In this section, we review candidate metadata properties that could be used by anyone developing (i) an ontology or semantic resource or (ii) a FAIRness assessment tool to obtain associated credits as listed in the previous section.

4.1 Candidate metadata properties to support FAIRness

Here, we reuse the MOD ontology metadata model (v1.4) [24] as a reference to pick up metadata properties. MOD1.4 reviewed 346 metadata properties from 23 standard metadata vocabularies (such as Dublin Core, DCAT, VoID, ADMS, VOAF, Schema.org, etc.) to provide a list of 127 "aligned or crosswalked" properties that can be used to describe an ontology or a semantic resource. MOD allows us to unambiguously identify which property may be used; however, our grid could be implemented with any other metadata standard or combination of standards that cover all the sub-principles.

The outcome of this process is a list of **58 candidate metadata properties** that may be used to support FAIRness assessment and assign some credits from our grid. These metadata properties might help to assign **276 credits** over a total of 478 (57%). We have separated the metadata properties for any principles from the ones for F2, which has to be treated apart. Indeed, F2 ("Data are described with rich metadata") was assigned all the properties that MOD1.4 has reviewed as relevant for ontologies that have not been assigned to another sub-principle. We refer to the first group as *core metadata properties* (Table 4) and to the second group as *extra metadata properties* (Table 5). The idea is that any ontologies using some of the **69 extra metadata properties** in addition to the core 58 ones, will be "FAIRer".

Table 4. List of core metadata properties from MOD1.4 to help make an ontology FAIR.

Principle		Credits	Metadata properties
F	F1	29	owl:ontologyIRI, owl:versionIRI, dct:identifier
	F4	24	schema:includedInDataCatalog
A	A1	36	owl:ontologyIRI, dct:identifier, sd:endpoint
	A2	4	omv:status, owl:deprecated
I	I1	44	omv:hasOntologyLanguage, omv:hasFormalityLevel, omv:hasOntologySyntax, dct:hasFormat, dct:isFormatOf
	I2	22	owl:imports, voaf:hasEquivalenceWith, owl:priorVersion, voaf:similar, voaf:metadataVoc, dct:relation, dct:isPartOf, voaf:specializes, schema:translationOfWork, voaf:generalizes
R	R1	8	mod:prefLabelProperty, mod:synonymProperty, mod:definitionProperty, mod:authorProperty, bpm:obsoleteProperty, mod:hierarchyProperty, mod:obsoleteParent, mod:maxDepth
	R1.1	37	dct:license, dct:rightsHolder, dct:accessRights, cc:morePermissions, cc:useGuidelines
	R1.2	36	dct:creator, dct:ontributor, pav:curatedBy, schema:translator, dct:source, prov:wasGeneratedBy, prov:wasInvalidatedBy, dct:accrualMethod, dct:accrualPeriodicity, dct:accrualPolicy, omv:versionInfo, vann:changes, dct:hasVersion, omv:usedOntologyEngineeringTool, omv:usedOntologyEngineeringMethodology, omv:conformsToKnowledgeRepresentationParadigm, omv:designedForOntologyTask, mod:competencyQuestion, foaf:fundedBy
	R1.3	36	mod:ontologyInUse, omv:endorsedBy, mod:group, dct:accessRights
Total		**276**	**58 metadata properties**

We identified that 46% of the FAIR principles (i.e., F2, I1, I2, R1, R1.1, R1.2, and R1.3) are totally evaluable with metadata properties, 33% are partially evaluable (i.e., F1, F4, A1, A1.2, and A2). Two principles for which we have not found any metadata property are A1.1 ("*The protocol is open, free, and universally implementable.*") and A1.2 ("The protocol allows for an authentication and authorization where necessary.") because they are completely related to the evaluation of the communication protocol, not the ontology persay. A sub-principle is not totally evaluable with metadata properties when it is about an aspect independent of the ontology itself but related to the library/repository hosting the ontology. For instance: F4 (“*(Meta)data are registered or indexed in a searchable resource.*”) concerns also ontology repositories.

F3 (“*Metadata clearly and explicitly include the identifier of the data they describe.*”) is excluded from Table 4 as MOD1.4 do not yet offer a property to establish the link between an ontology and its metadata (necessary when metadata are not explicitly included in the same file than the ontology itself). Such a property is currently being discussed in the FAIR Digital Object working group of GO FAIR that shall soon release a new metadata vocabulary including `fdo:hasMetadata` and `fdo:metadataOf` properties. Even if I3 is totally evaluable with metadata, the currently proposed candidate metadata are not covering its evaluation. Here again, we need some extension to MOD to enable encoding all information required by this principle (especially alignment qualification). MOD is currently being extended as a new model compliant with DCAT2 within the RDA VSSIG and H2020 FAIRsFAIR.

Table 5. List of extra metadata properties from MOD1.4 to make an ontology FAIRer.

Principle		Credits	Metadata properties
F	F2	27	omv:acronym, dct:title, dct:alternative, skos:hiddenLabel, dct:description, foaf:page, omv:resourceLocator, omv:keywords, dct:coverage, foaf:homepage, vann:example, vann:preferredNamespaceUri, void:uriRegexPattern, idot:exampleIdentifier, dct:publisher, dct:subject, owl:backwardCompatibleWith, door:comesFromTheSameDomain, mod:sampleQueries, omv:knownUsage, dct:audience, doap:repository, doap:bugDatabase, doap:mailing-list, mod:hasEvaluation, mod:metrics, omv:numberOfClasses, omv:numberOfIndividuals, omv:numberOfProperties, mod:numberOfDataProperties, mod:numberOfObjectProperties, omv:numberOfAxioms, mod:numberOfLabels, mod:byteSize, vann:preferredNamespacePrefix, dct:language, dct:abstract, mod:analytics, dct:bibliographicCitation, rdfs:comment, foaf:depiction, foaf:logo,voaf:toDoList, schema:award, schema:associatedMedia, owl:isIncompatibleWith, dct:hasPart, schema:workTranslation, door:hasDisparateModelling, voaf:usedBy, voaf:hasDisjunctionsWith, omv:keyClasses, void:rootResource, mod:browsingUI, mod:sampleQueries, void:propertyPartition, void;classPartition, void:dataDump, void:openSearchDescription, void:uriLookupEndpoint, schema:comments, dct:created, dct:modified, dct:valid, dct:dateSubmitted, pav:curatedOn, omv:IsOfType
Total		**27**	**69 metadata properties**

4.2 FAIR or FAIRer: How FAIR is a semantic resource?

Qualifying the degree of FAIRness of a semantic resource or even comparing it with other semantic resources necessarily implies the use of a metric delimiting the range of values for each qualification (e.g., not FAIR, FAIR, or FAIRer). In that context, our proposed integrated quantitative grid allows for defining *thresholds*. For instance, the median value of the resulting total credits can be considered a minimum threshold to be FAIR. A semantic resource with a degree/score under this threshold will not be considered FAIR. Similarly, a semantic resource might be considered as “FAIRer” if it is described with extra metadata properties. In other words, answering the question: *“how much is a semantic resource FAIR?”* becomes possible with such a metric. In our grid, the total credits are 478, so the ***first threshold could be at 240*** (478/2+1) and the ***second threshold at 451*** (478-27), as illustrated in Fig. 2.

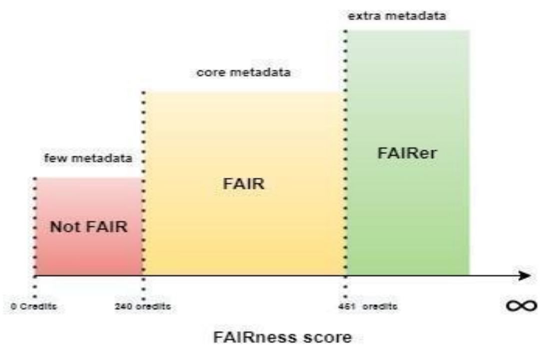

Fig. 2. Not FAIR, FAIR, or FAIRer: using the metric of the integrated quantitative grid.

Clearly, using a metric and threshold is the first required step in making the FAIRness assessment task machine-actionable and enabling the development of automatic FAIRness assessment tools. We believe it will also be beneficial for researchers to quantify the FAIRness degree of their semantic resources and compare them with other ones.

5 Conclusions and perspective

In this paper, we proposed an integrated quantitative grid for assessing the level of FAIRness of semantic resources and ontologies. Moreover, we provided a list of candidate metadata properties –from the MOD model v1.4– to enable FAIRness assessment and possibly implement systems based on our grid. Our grid was realized by analyzing existing related work (among others, the semantic Web community work before and since the FAIR movement) and summarizing them into one coherent scheme. A distinct feature of our grid is to propose a metric –and thus possible thresholds– for the qualification of any semantic resource. The grid is conceived in a way that can be customized, extended, or improved by other semantic experts in further studies. This work is a starting point for developing machine-actionable FAIRness assessment tools in the semantic Web context.

The motivation of this work was to go beyond the current recommendations to guide semantic stakeholders for making their semantic resources FAIR: We consider these recommendations, harmonize and integrate them to build a grid of 478 credits to assess the 15 FAIR principles.

Currently, we are using the grid to implement a FAIRness assessment tool in AgroPortal (http://agroportal.lirmm.fr/), a vocabulary and ontology repository dedicated to agri-food and based on the generic and open source OntoPortal technology[5]. However, in the future, this work will need to be further tested in other FAIRness assessment approaches and discussed within some international FAIR initiatives, for instance, RDA, GO FAIR, or projects such as FAIRsFAIR.

Acknowledgments

This work has been supported by the *Data to Knowledge in Agronomy and Biodiversity* project (D2KAB – www.d2kab.org – ANR-18-CE23-0017) and the project ANR *French participation in GO FAIR Food Systems Implementation Network* (FooSIN

5 https://github.com/ontoportal-lirmm

– https://foosin.fr – ANR19-DATA-0019). We also thank the VSSIG (Vocabulary and Semantic Services Interest Group) of the Research Data Alliance and the H2020 FAIRsFAIR project T2.2 on "FAIR Semantics" for fruitful discussions and exchanges.

References

[1] M. D. Wilkinson *et al.*, "The FAIR Guiding Principles for scientific data management and stewardship," *Sci. Data*, vol. 3, no. 1, Art. no. 1, Mar. 2016.
[2] M. L. Zeng and P. Mayr, "Knowledge Organization Systems (KOS) in the Semantic Web: a multi-dimensional review," *Int. J. Digit. Libr.*, vol. 20, no. 3, pp. 209–230, Sep. 2019.
[3] Y. Le Franc, J. Parland-von Essen, L. Bonino, H. Lehväslaiho, G. Coen, and C. Staiger, "D2.2 FAIR Semantics: First recommendations," Mar. 2020.
[4] C. Caracciolo *et al.*, "39 Hints to Facilitate the Use of Semantics for Data on Agriculture and Nutrition," *Data Sci. J.*, vol. 19, no. 1, Art. no. 1, Dec. 2020.
[5] M. D. Wilkinson, S.-A. Sansone, E. Schultes, P. Doorn, L. O. Bonino da Silva Santos, and M. Dumontier, "A design framework and exemplar metrics for FAIRness," *Sci. Data*, vol. 5, Jun. 2018.
[6] R. David *et al.*, "FAIRness Literacy: The Achilles' Heel of Applying FAIR Principles," *Data Sci. J.*, vol. 19, no. 1, Art. no. 1, Aug. 2020.
[7] C. Bahim *et al.*, "The FAIR Data Maturity Model: An Approach to Harmonise FAIR Assessments," *Data Sci. J.*, vol. 19, no. 1, Art. no. 1, Oct. 2020.
[8] "SurveyMonkey Powered Online Survey." https://www.surveymonkey.com/r/fairdat (accessed Apr. 19, 2021).
[9] "FAIR self-assessment tool." https://satifyd.dans.knaw.nl/ (Jul. 23, 2021).
[10] S. Cox and J. Yu, "OzNome 5-star Tool: A Rating System for making data FAIR and Trustable," Oct. 2017, Accessed: Apr. 12, 2021. [Online]. Available: https://publications.csiro.au/rpr/pub?pid=csiro:EP175062
[11] M. D. Wilkinson *et al.*, "Evaluating FAIR maturity through a scalable, automated, community-governed framework," *Sci. Data*, vol. 6(1) Sep. 2019.
[12] M. Mokrane, L. Cepinskas, V. Åkerman, J. de Vries, and I. von Stein, "FAIR-Aware," 2020, Accessed: Mar. 14, 2021. [Online]. Available: https://pure.knaw.nl/portal/en/publications/fair-aware
[13] C. Bizer, T. Heath, and T. Berners-Lee, "Linked Data: The Story so Far," *Semantic Services, Interoperability and Web Applications: Emerging Concepts*, 2011. www.igi-global.com/chapter/linked-data-story-far/55046.
[14] K. Janowicz, P. Hitzler, B. Adams, D. Kolas, and C. Vardeman II, "Five stars of Linked Data vocabulary use," *Semantic Web*, vol. 5, no. 3, pp. 173–176, 2014.
[15] A. Hasnain and D. Rebholz-Schuhmann, "Assessing FAIR Data Principles Against the 5-Star Open Data Principles," in *The Semantic Web: ESWC 2018 Satellite Events*, Cham, 2018, pp. 469–477.
[16] D. Garijo and M. Poveda-Villalón, "Best Practices for Implementing FAIR Vocabularies and Ontologies on the Web," *ArXiv*, 2020.
[17] N. Matentzoglu, J. Malone, C. Mungall, and R. Stevens, "MIRO: guidelines for minimum information for the reporting of an ontology," *J. Biomed. Semant.*, vol. 9, no. 1, p. 6, Jan. 2018.

[18] B. Dutta, A. Toulet, V. Emonet, and C. Jonquet, “New Generation Metadata Vocabulary for Ontology Description and Publication,” in *Metadata and Semantic Research*, Cham, 2017, pp. 173–185.
[19] W. Hugo, Y. Le Franc, G. Coen, J. Parland-von Essen, and L. Bonino, “D2.5 FAIR Semantics Recommendations Second Iteration,” Dec. 2020.
[20] M. Poveda-Villalón, P. Espinoza-Arias, D. Garijo, and O. Corcho, “Coming to Terms with FAIR Ontologies,” in *Knowledge Engineering and Knowledge Management*, Cham, 2020, pp. 255–270.
[21] M. Lei Zeng and J. Clunis, “FAIR + FIT: Guiding Principles and Functional Metrics for Linked Open Data (LOD) KOS Products,” Mar. 16, 2020. https://sciendo.com/article/10.2478/jdis-2020-0008.
[22] S. J. D. Cox, A. N. Gonzalez-Beltran, B. Magagna, and M.-C. Marinescu, “Ten simple rules for making a vocabulary FAIR,” *PLOS Comput. Biol.*, vol. 17, no. 6, p. e1009041, Jun. 2021.
[23] J. Frey, D. Streitmatter, F. Götz, S. Hellmann, and N. Arndt, “DBpedia Archivo: A Web-Scale Interface for Ontology Archiving Under Consumer-Oriented Aspects,” in *Semantic Systems. In the Era of Knowledge Graphs*, vol. 12378, 2020, pp. 19–35.
[24] C. Jonquet, A. Toulet, B. Dutta, and V. Emonet, “Harnessing the Power of Unified Metadata in an Ontology Repository: The Case of AgroPortal,” *J. Data Semant.*, vol. 7, no. 4, pp. 191–221, Dec. 2018. .