

FAIR Ontologies, FAIR Ontology Alignments^{*}

Cassia Trojahn^{1,†}

¹*Institut de Recherche en Informatique de Toulouse, Université de Toulouse 2, Toulouse, France*

Abstract

A number of recommendations has been proposed so far for making FAIR data, including more recent ones on how to publish FAIR ontologies on the web. However, less attention has been given to producing FAIR ontology alignments. This paper reviews existing FAIR data initiatives and discusses the required efforts for generating and publishing FAIR alignments on the Web. It aligns the four principles (F, A, I and R) to the actions and requirements towards the generation and sharing of FAIR alignments. It ends with a discussion on further developments.

Keywords

Ontology alignment, FAIR principles, FAIR alignment

1. Introduction

Since their proposal in 2016, the FAIR (Findable, Accessible, Interoperable, and Reusable) principles [1] have become increasingly important in (scientific) data management. They have been the subject of numerous research projects and initiatives, such the ones pushed by the European Open Science Cloud (EOSC) or the Research Data Alliance (RDA), and have been adopted by both private and public institutions. These principles correspond to a set of recommendations that aims to facilitate data reuse by humans and machines. They are domain-independent and may be implemented principally by: (F), assigning unique and persistent identifiers to data and metadata, and describing them with rich metadata that enable their indexing and discovery; (A), using open and standard protocols for data access; (I), using formal languages, and (FAIR) vocabularies to represent (meta)data; and (R), documenting (meta)data with rich metadata about usage license, provenance and data quality, using domain-relevant standards. While such recommendations are close to what has been so far advocated by the Linked Open Data (LOD) initiative, as stated in [2], unlike Linked Data, the FAIR principles make an explicit and strong focus on metadata management in order to enable resource findability and reusability.

Resulting from those projects and initiatives, a number of recommendations has been proposed for making data FAIR, as for example the FAIRsFAIR recommendations [3] that provide of 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 practices are related to one or more FAIR principles and links to existing recommendations and related stakeholders

EKA-C 2022: Companion Proceedings of the 23rd International Conference on Knowledge Engineering and Knowledge Management

✉ cassia.trojahn@irit.fr (C. Trojahn)

ORCID 0000-0003-2840-005X (C. Trojahn)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

(e.g: practitioners, repositories or the Semantic Web community). Several frameworks have also been proposed to assess the degree of FAIRness of resources¹, as the well-known “FAIR Data Maturity Model” [4] which mainly consists of a list of evaluation items in a spreadsheet format. Other works have proposed more automated approaches for FAIRness evaluation [5, 6] based on web applications. More recently, proposals have addressed the evaluation of vocabularies and ontologies as well as best practices for implementing FAIR vocabularies and ontologies on the Web [7, 8], with tools available to support this evaluation process, such as FOOPS! [9] and O’FAIRe (for Ontology FAIRness Evaluator) for the evaluation of semantic resources in general (e.g., vocabularies, terminologies, thesaurus) [10].

Despite this wave of efforts on making data and ontologies FAIR, few attention has been given to producing FAIR ontology alignments, in particular to the generation of (rich) alignment metadata. While alignment representation languages have become the *standard de facto* in the ontology matching field, such as the RDF Alignment API format and the EDOAL (Expressive and Declarative Ontology Alignment Language) [11] languages, they were not designed to provide rich metadata on the alignments. Metadata in terms of interpretation, explanation, quality, provenance, usage license, version history, etc., both at the level of the alignment and at the level of the correspondences, are so far missing. This lack of documentation definitely makes hard the task of exploiting, combining and reproducing alignments.

Recently, the EOSC has addressed the problem of “semantic mapping sharing”, reporting on the requirements for creating, documenting, and publishing alignments and cross-walks within a particular scientific community, as well as across scientific domains [12]. This effort resulted from 26 interviews that have been carried out and on observations about the work in the ESFRI (European Strategy Forum on Research Infrastructures) initiatives, in the realm of EOSC discussions, and the work in RDA. A complementary effort is the *Simple Standard for Sharing Ontological Mappings* (SSSOM) [13] that proposes a machine-readable and extensible vocabulary to describe metadata that makes imprecision, inaccuracy and incompleteness in correspondences explicit. Tools and software libraries working with the standard are made publicly available. However, there is still no clear ‘alignments’ between FAIR principles and requirements for making FAIR alignments.

This paper reviews the required efforts for generating and publishing FAIR alignments on the Web, what brings to light many still unsolved issues in the field such as the lack of rich metadata alignment models, lack of ontology alignment repositories for alignment publishing and sharing (as LOV for ontologies and vocabularies), common good practices, etc. It aligns the four principles (F, A, I and R) to the actions and requirements towards the generation and sharing of FAIR alignments on the Web. The paper ends with a discussion on further developments.

2. FAIR alignment requirements

The list of FAIR guiding principles defined in [1] is as in the following. These principles are presenting together with the efforts for making FAIR alignments.

¹Most of which are listed on <https://fairassist.org/> (accessed on 10th August 2022).

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: *Alignments should be findable in order to facilitate their reuse. Data here are alignments themselves. They have to be exposed, stored, in dedicated repositories (e.g., github), and ideally indexed in alignment (searchable) catalogs. They have to be described with enough information allowing their full exploitation (see R below): rich describing of the alignment content (data) and rich description of the alignment (metadata). The question of having globally unique and persistent identifiers arises and has to be discussed in the ontology matching community.*

Accessible 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. *Alignments should be accessible in order to facilitate their reuse. They should be made available along with (open) mechanisms and protocols for content negotiation, allowing for both automated and human exploration (with at least one RDF serialization and HTML, as recommended by FOOPS! for ontologies).*

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. *The main question that arises here is: what should be interoperable alignments? A debatable view but consistent with the 'I' principle is that their metadata should be described with descriptions that are formally represented and accessible, using FAIR ontologies and vocabularies. Again, in the sense of FOOPS! evaluation checks, such a vocabulary has to include references to existing vocabularies in their metadata annotations, classes and properties. They have to be documented both in a human-readable manner and with formal languages that are expressive enough to be able to capture the semantics of each correspondence (meaning of the relation between the ontology entities being aligned, how the confidence should be considered, explanations on how the correspondence has been found, etc.). Alignments have to be clearly described to be consumed for other communities than the ones producing them. Alignments have to be understood.*

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. *Alignments should definitely be produced to be reusable. Besides type of documentation referred in 'I', alignments have to be exposed with clear information on the usage license and their provenance (who has created the alignment, which tool and tool version, on which ontologies version, etc.).*

3. Discussion

While the previous section has introduced the ideal requirements for making FAIR alignments, this section discusses the further developments and the need for joint efforts in those directions.

With respect to making alignments *findable*, this is still an open issue in the field. In fact, alignments generated in research papers are rarely available and OAEI alignments are rather available as zip files on dedicated web pages without any indexation. While some (domain-specific) portals, such as BioPortal² and AgroPortal³ catalog curated alignments, there is still a lack of catalog services to index general purpose alignments, as the LOV⁴ does for ontologies. A good practice could be exposing alignments on repositories as github, with an indexation by alignments catalogs that can alternatively also offer the storing service. The field urgently needs a LOA (Linked Open Alignment) service. Versioning has also to be taken into account at the metadata level (version annotations, as `owl:priorVersion` and `owl:versionInfo` for ontologies).

Alignments (data) have to be *accessible* (besides their metadata being findable). As stated above, they should be made available along with mechanisms for content negotiation, allowing To docu While there is specialised tools for generating ontology documentation in a human-readable format, including content-negotiation configuration, as WIDOCO [14], alignment documentation tools should be able to deal with both alignment metadata and alignment content visualisation.

Alignments have to be *interoperable*. At least, their metadata has to be formally represented and accessible, using FAIR vocabularies. A number of vocabularies has been proposed to represent metadata in general (Dublin Core, VoID, Schema.org, DCAT, DCAT-AP), with extensions for accommodating specific kinds of data, such as geo-spatial data (GeoDCAT-AP) or statistical data (StatDCAT-AP). The question that arises is “what are FAIR vocabulaires”? In [15], eleven features for FAIR vocabulary are proposed, covering requirements for identifiers, access protocols, knowledge representation, etc. First, FAIR vocabularies for alignment metadata have to be proposed and then reused. SSSOM brings a format that improves on the metadata and provenance information associated with the correspondences, enhancing their understanding and potential future reuse. It seems to be a good candidate towards a FAIR model for alignment metadata.

Alignments have be generated to be *reusable* as much as possible. It is still difficult to reuse alignments mainly because they are hardly findable, accessible, and interoperable. As stated above, alignment representation languages still lack descriptive metadata, explanation and justification components. It is difficult to interpret a correspondence involving complex constructors, the truth relation expressed between the involved ontologies entities within a correspondence, etc. The languages fail as well in offering a support for representation “how” a given correspondence has been found, in both a human-readable format and formally. This goes beyond provenance, which can be documented using existing vocabulaires, as PROV-O⁵ to represent and interchange provenance information (agent, activity, entity, etc).

²<https://bioportal.bio>

³<https://agroportal.lirmm.fr>

⁴<https://lov.linkeddata.es/dataset/lov/>

⁵<https://www.w3.org/TR/prov-o/> (accessed on 10th August 2022)

Last but not at least, the Ontology Alignment Evaluation Initiative (OAEI), as an international coordinated forum for matching systems developers, has also led the initiative towards the generation of FAIR alignments: agreement on metadata model for adoption (SSSOM is a good candidate), coordination with the EOSC group [12] on the developments on the semantic mapping framework, for citing few concrete actions.

Summing up, there is a path to producing and exposing alignments fully compliant to the FAIR principles, in particular: i) what is the vocabulary to use to explicitly describe them both at alignment level (descriptive metadata, provenance, licence, versioning) and at correspondence level (relation interpretation, confidence interpretation, explanation, justification), ii) how to index them (LOA); iii) how to assist alignment producers on sharing alignments; iv) how to evaluate the degree of FAIRness of alignments.

4. Future Work

Concretely, in a short term, the next steps to integrate the FAIR principles into ontology alignment have to be: adopting SSSOM in the well-known OAEI campaigns; working on SSSOM extensions to provide alignment metadata at both correspondence and alignment levels; providing clear guidelines to make FAIR alignments (with an identification of the requirements to do so), what involves ‘aligning’ the FAIR guidelines and recommendations ([3, 8, 15, 12], for citing a few); developing a LOA.

References

- [1] M. Wilkinson, M. Dumontier, *et al.*, The FAIR Guiding Principles for scientific data management and stewardship, *Scientific data* 3 (2016) 1–9.
- [2] M. Poveda-Villalón, P. Espinoza-Arias, D. Garijo, Ó. Corcho, Coming to terms with FAIR ontologies, in: C. M. Keet, M. Dumontier (Eds.), *Knowledge Engineering and Knowledge Management - 22nd International Conference, EKAW, 2020*, pp. 255–270. URL: https://doi.org/10.1007/978-3-030-61244-3_18. doi:10.1007/978-3-030-61244-3_18.
- [3] Y. Le Franc, J. Parland-von Essen, L. Bonino, H. Lehväslaiho, G. Coen, C. Staiger, D2.2 fair semantics: First recommendations, 2020. URL: <https://doi.org/10.5281/zenodo.3707985>. doi:10.5281/zenodo.3707985.
- [4] FAIR Data Maturity Model Working Group RDA, FAIR Data Maturity Model. Specification and Guidelines, 2020. URL: <https://doi.org/10.15497/rda00050>. doi:10.15497/rda00050, <https://doi.org/10.15497/rda00050> Accessed 6 May 2022.
- [5] M. Wilkinson, M. Dumontier, *et al.*, Evaluating FAIR maturity through a scalable, automated, community-governed framework, *Sc. Data* 6 (2019) 1–12.
- [6] A. Devaraju, R. Huber, M. Mokrane, P. Herterich, L. Cepinskas, J. de Vries, H. L’Hours, J. Davidson, A. White, FAIRsFAIR Data Object Assessment Metrics 0.5, Technical Report, Research Data Alliance (RDA), 2020. URL: <https://zenodo.org/record/6461229>. doi:10.5281/zenodo.6461229, <https://zenodo.org/record/6461229> Accessed 3 May 2022.
- [7] D. Garijo, M. Poveda-Villalón, Best practices for implementing FAIR vocabularies and

- ontologies on the web, CoRR abs/2003.13084 (2020). URL: <https://arxiv.org/abs/2003.13084>. arXiv:2003.13084.
- [8] S. J. D. Cox, A. N. Gonzalez-Beltran, B. Magagna, M.-C. Marinescu, Ten simple rules for making a vocabulary fair, *PLOS Computational Biology* 17 (2021) 1–15. URL: <https://doi.org/10.1371/journal.pcbi.1009041>. doi:10.1371/journal.pcbi.1009041.
- [9] D. Garijo, Ó. Corcho, M. Poveda-Villalón, FOOPS!: An Ontology Pitfall Scanner for the FAIR principles, in: *Proc. of the ISWC 2021 Posters, Demos and Industry Tracks*, 2021. URL: <http://ceur-ws.org/Vol-2980/paper321.pdf>.
- [10] E. Amdouni, S. Bouazzouni, C. Jonquet, O’FAIRe: Ontology FAIRness Evaluator in the AgroPortal semantic resource repository, in: *ESWC 2022 Poster and demos*, Greece, 2022. URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-03630543>.
- [11] J. David, J. Euzenat, F. Scharffe, C. Trojahn, The alignment API 4.0, *Semantic Web 2 (2011)* 3–10. URL: <https://doi.org/10.3233/SW-2011-0028>. doi:10.3233/SW-2011-0028.
- [12] D. Broeder, P. Budroni, E. Degl’Innocenti, Y. Le Franc, W. Hugo, K. Jeffery, C. Weiland, P. Wittenburg, C. M. Zwolf, SEMAF: A Proposal for a Flexible Semantic Mapping Framework, 2021. URL: <https://doi.org/10.5281/zenodo.4651421>. doi:10.5281/zenodo.4651421.
- [13] N. Matentzoglou, J. P. Balhoff, S. M. e. a. Bello, A Simple Standard for Sharing Ontological Mappings (SSSOM), *Database* 2022 (2022). URL: <https://doi.org/10.1093/database/baac035>. doi:10.1093/database/baac035. arXiv:<https://academic.oup.com/database/article-pdf/doi/10.1093/database/baac035/43832024/baac035.pdf>, baac035.
- [14] D. Garijo, Widoco: a wizard for documenting ontologies, in: *International Semantic Web Conference*, Springer, Cham, 2017, pp. 94–102. URL: <http://dgarijo.com/papers/widoco-iswc2017.pdf>. doi:10.1007/978-3-319-68204-4_9.
- [15] F. Xu, N. S. Juty, C. A. Goble, S. Jupp, H. E. Parkinson, M. Courtot, Features of a FAIR vocabulary, in: *13th International Conference on Semantic Web Applications and Tools for Health Care and Life Sciences, SWAT4HCLS*, 2022, pp. 118–148. URL: <http://ceur-ws.org/Vol-3127/paper-15.pdf>.