What Should be the Minimum Requirements for FAIR Alignments?

Cassia Trojahn[1][0000−0003−2840−005X] and Nicolas Matentzoglu[2][0000−0002−7356−1779]

[1]Institut de Recherche en Informatique de Toulouse, France
cassia.trojahn@irit.fr
Semanticly, United Kingdom
nicolas.matentzoglu@gmail.com

1 Introduction

The FAIR (Findable, Accessible, Interoperable, Reusable) principles [10] have become increasingly important in data management. A number of recommendations has been proposed for making FAIR data[1], such as the "FAIR Data Maturity Model" [5]. Best practices for implementing FAIR vocabularies and ontologies on the Web [7, 3] have been also proposed and the evaluation of their degree of FAIRness addressed [6, 1].

Despite this wave of efforts, few attention has been given to producing FAIR ontology 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 [2]. A complementary effort is the Simple Standard for Sharing Ontological Mappings (SSSOM) [9] that proposes a machine-readable and extensible vocabulary to describe metadata that makes imprecision, inaccuracy and incompleteness in correspondences explicit.

This paper goes towards a set of minimum requirements for generating and publishing FAIR alignments, 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, common good practices for alignment engineering (as for ontology engineering), and so on. This is an early proposal that is inspired from what has been done so far in different FAIR recommendations. However, they have to be further discussed in the ontology matching community.

2 Minimum requirements?

Four minimum requirements are proposed, as described in the following:
Requirement 1: alignments have to be described with rich metadata While alignment representation languages have become the *standard de facto*, in practice, in the field as the RDF Alignment API format and EDOAL [4], they lack on providing rich metadata on the alignments. Alignments have at least to be described with metadata on their provenance (who, when, tool, tool version, etc.), usage license, version (as ontologies evolve),and so on.

[1] Most of which are listed here: `https://fairassist.org/` (accessed on 08/10/2022).

Requirement 2: correspondences have to be described with rich metadata A fine-grained metadata is required at correspondence level, in terms of relation interpretation, confidence interpretation, explanation, justification (how the correspondence has been found). In fact, it is hard to interpret the truth relation expressed between the involved ontologies entities within a correspondence, without clear statements.
Requirement 3: alignments have to be published For being reusable, alignments have to the accessible. They have to be at least exposed and stored, in dedicated repositories (e.g., github), and ideally indexed in alignment (searchable) catalogs. They have to be published with standard formats. It is evident that the field lacks searchable services (Linked Open Alignment service, LOA), as LOV for ontologies.
Requirement 4: alignments have to be exposed with content negotiation Alignments have to be made available along with (open) mechanisms and protocols for content negotiation, with at least one RDF serialization and HTML, as recommended by FOOPS! for ontologies.
This set of minimum requirements has not been defined from questionnaires and deeper discussion in the community, but rather from observations on the FAIR recommendations in general.
As a future work, an alignment of recommendations ([8, 3], for citing a very few) is required to provide further concrete guidelines; together with an extensive involvement of the community in order to refine the requirements presented here and/or define necessary and sufficient conditions for making FAIR alignments.

References

1. E. Amdouni, S. Bouazzouni, and C. Jonquet. O'FAIRe: Ontology FAIRness Evaluator in the AgroPortal semantic resource repository. In *ESWC 2022 Poster and demos*, Greece, 2022.
2. D. Broeder, P. Budroni, E. Degl'Innocenti, Y. Le Franc, W. Hugo, K. Jeffery, C. Weiland, P. Wittenburg, and C. M. Zwolf. SEMAF: A Proposal for a Flexible Semantic Mapping Framework, Mar. 2021.
3. S. J. D. Cox, A. N. Gonzalez-Beltran, B. Magagna, and M.-C. Marinescu. Ten simple rules for making a vocabulary fair. *PLOS Computational Biology*, 17(6):1–15, 06 2021.
4. J. David, J. Euzenat, F. Scharffe, and C. Trojahn. The alignment API 4.0. *Semantic Web*, 2(1):3–10, 2011.
5. FAIR Data Maturity Model Working Group RDA. FAIR Data Maturity Model. Specification and Guidelines, June 2020. https://doi.org/10.15497/rda00050 Accessed 6 May 2022.
6. D. Garijo, Ó. Corcho, and M. Poveda-Villalón. FOOPS!: An Ontology Pitfall Scanner for the FAIR principles. In *Proc. of the ISWC 2021 Posters, Demos and Industry Tracks*, 2021.
7. D. Garijo and M. Poveda-Villalón. Best practices for implementing FAIR vocabularies and ontologies on the web. *CoRR*, abs/2003.13084, 2020.
8. Y. Le Franc, J. Parland-von Essen, L. Bonino, H. Lehväslaiho, G. Coen, and C. Staiger. D2.2 fair semantics: First recommendations, Mar. 2020.
9. N. Matentzoglu, J. P. Balhoff, and S. M. e. a. Bello. A Simple Standard for Sharing Ontological Mappings (SSSOM). *Database*, 2022, 05 2022. baac035.
10. M. Wilkinson, M. Dumontier, and *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3(1):1–9, 2016.