Moving towards FAIR mappings and crosswalks

Jana Martínková[1,2], Nick Juty[3], Alejandra Gonzalez Beltran[4], Carole Goble[3], and Yann Le Franc[1,*]

[1] *e-Science Data Factory S.A.S.U., Montpellier, France*
[2] *Czech Technical University in Prague, Faculty of Information Technology, Prague, 160 00, Czech Republic*
[3] *The University of Manchester, Department of Computer Science, Oxford Road, Manchester,M13 9PL UK*
[4] *Scientific Computing, Science and Technology Facilities Council, Harwell Campus, OX11 0QX, UK*

Abstract

Mappings and crosswalks are key elements to ensure semantic interoperability as well as metadata and data integration between different information systems. Designing FAIR compliant systems requires making sure all the elements that constitute the systems are themselves FAIR to support machine-actionability and automation. This paper describes the ongoing European and international effort to build a framework for FAIR mappings and crosswalks. This framework aims to be generic enough to capture the diverse set of use cases and methodologies across domains and communities. It should be composed of a set of technical recommendations to aid compliance with FAIR principles, a set of models for machine-actionable mappings and crosswalks as well as a practical framework with aligned good practices to support the creation of mappings by scientific communities. Developed in the context of FAIR IMPACT, a Horizon Europe project, this work will be pursued within a more international context as a Research Data Alliance Working Group.

Keywords

Semantic artefacts, Mapping, Alignment, FAIR principles, Crosswalks, mapping exchange model, mapping practices

1. Introduction

A mapping defines connections or relationships between different information elements by identifying similarities, correspondences, and alignments. Mappings include different types of connections, depending on the level of the elements that are being mapped. Semantic alignments, for instance, involve relationships between ontologies. Metadata mappings, on the other hand, link different metadata properties from a source to a target schema. Possible applications of these mappings are to enhance the quality of search results or to transition metadata from one schema to another. Thirdly, data mappings enable aligning two datasets, and may include tasks such as unit conversions and scaling when dealing with multi-scale data.

These different mapping types require capturing different pieces of information about the relationship, for example, complex mappings involving unit conversion require expressing mathematical equations for the conversion itself, whereas semantic alignment requires a predicate

Proceedings of the Joint Ontology Workshops (JOWO) - Episode X: The Tukker Zomer of Ontology, and satellite events co-located with the 14th International Conference on Formal Ontology in Information Systems (FOIS 2024), July 15-19, 2024, Enschede, The Netherlands.

* Corresponding Author
jmartinkova@esciencefactory.com (J. Martínková); nick.juty@manchester.ac.uk (N. Juty); alejandra.gonzalez.beltran@gmail.com (A. Gonzalez Beltran); carole.goble@manchester.ac.uk (C. Goble); ylefranc@esciencefactory.com (Y. Le Franc)
0000-0002-2036-8350 (N. Juty); 0000-0003-3499-8262 (A. Gonzalez Beltran), 0000-0003-1219-2137 (C. Goble), 0000-0003-4631-418X (Y. Le Franc)

indicating equivalence or broader/narrower relationship. Despite their differences, the main components of a mapping remain consistent across their different types.

Mappings have been identified as key elements for interoperability between information systems in various documents, such as the SEMAF framework [1], the FAIR Semantics recommendations [2] proposed by the FAIRsFAIR project[2], the EOSC interoperability framework [3] developed by the EOSC Executive Board Working Groups FAIR and Architecture, and the EOSC Semantic Interoperability Task Force final report [4]. Despite their key role, these particular digital entities are hard to find and reuse, and are not represented with a common exchange format.

Mappings and crosswalks, defined as sets of mappings designed for specific purposes or directly linked to particular use cases, serve as vital connectors between different elements, thereby facilitating interoperability among diverse information systems. Consequently, they play a key role in realising the FAIR principles [5], particularly as integral components for FAIR semantic artefacts. However, mappings are often represented in tabular form indicating an unspecified correspondence, but with no clear semantics of the relationship between the entities, or those mappings may even only be embedded in tools for data preparation or analysis. Given their importance and the points above, we argue that mappings and crosswalks, being digital entities themselves, must also adhere to FAIR principles. This involves making mappings accessible in relevant repositories where they can be curated, integrated, found and made available for reuse.

There have not been many proposals for standards for mapping definitions and exchange. When considering semantic mappings, Expressive and Declarative Ontology Alignment Language (EDOAL)[3] [6] has been developed by the Ontology Alignment Evaluation community [7] to represent ontology mappings. More recently, a new model, the Simple Standard for Sharing Ontology Mappings or SSSOM [8], emerged from the biomedical community. This model proposes a tabular representation of the mapping with an extensive set of metadata necessary to understand the rationale behind the mappings, the justifications and its provenance. This additional information is of high relevance for the further reuse of the mappings.

In this paper, we describe our approach to address the challenges of making mappings FAIR by focusing on both the technical aspects of a FAIR representation, and the practical aspects of the mapping process. In the first part of this paper, we present the methodology used to develop a set of technical recommendations for FAIR mappings and our effort to leverage and extend the SSSOM model to cover a more extensive set of mapping types, while conserving an appropriate level of metadata. Subsequently, we describe our effort to build a generic and coherent mapping practice framework to support the mapping process for scientific communities. We then reflect on the future work to be done toward a commonly agreed upon framework supporting the creation and the sharing of FAIR mappings.

2. Towards a framework for FAIR mappings and crosswalks

To facilitate the creation of FAIR-compliant mappings and crosswalks, we are currently developing a framework tailored to this purpose. Our ongoing work aims to establish a standardised approach to machine-actionable mappings and crosswalks while offering practical guidelines. The objective of this framework is twofold: firstly, to ensure that mappings and crosswalks adhere to the FAIR principles, maximising their usability and impact, and secondly, to engage with diverse communities to capture a wide range of perspectives and use cases, ensuring its relevance across diverse domains.

At its current stage of development, this framework for FAIR mappings and crosswalks is structured into three interconnected parts. Firstly, the Recommendations for FAIR mappings provide technical guidance and requirements for ensuring FAIR compliance. Secondly, a proposal for a

[2] https://www.fairsfair.eu/
[3] https://moex.gitlabpages.inria.fr/alignapi/edoal.html

machine-actionable common exchange model should enable sharing of the different types of mappings. Lastly, a practical framework offers guidelines and best practices for creating and maintaining mappings and crosswalks. Each component of this ongoing work is detailed in the following sections.

2.1. Recommendations on how to make mapping FAIR

While it is possible to propose technical implementations at the level of each individual FAIR principle, we have chosen to develop our recommendations based on the grouping of those principles within the following four key concept categories: i) Metadata, ii) Persistent Identifiers, iii) Service and API, and iv) Format and Model. In this section, we present our preliminary analysis for each concept category.

The primary concern that was addressed was to select a **model** to exchange mappings and to investigate **formats** supporting machine-actionability. The Format and Model topic covers mainly the principles related to interoperability (I1, I2 and I3) and reusability (R1, R1.1, R1.2, R1.3).

Our first recommendation to implement the FAIR Principles is to use the SSSOM model which provides an extensive documentation of the mappings and crosswalks (principle I1). Although this model has been developed for a specific type of mapping (entity mappings), it could accommodate simple mappings between metadata fields for instance. However, we are aware of the limitations of SSSOM for “complex mappings” and for different types of mappings. In section 2.2, we describe our approach to leverage and extend SSSOM.

The main motivation for supporting this model lies in its simplicity and its format which is close to the format usually used to create and share mappings i.e. tabular format. Compared to the classical approach of correspondence tables where the relation between the mapped entities is not specified, SSSOM emphasises the use of defined relationships such as *skos:broader, skos: exactMatch, owl:sameAs* which enable more semantically described mappings. One of the key features of the SSSOM model is that it can be serialised as Tabular Separated Values (TSV) but can also be converted to RDF. Since mappings are often used in web-based environments, we recommend the use of RDF serialisations to share and publish the mappings, whenever possible (principle I3). These standards offer clear specifications, ensuring precision and clarity in mapping representation. They also enable the use of standard and FAIR vocabularies (principle I2).

FAIR principles emphasise the importance of rich **metadata**. Our second focus has been on investigating the needs for metadata. This topic covers the principles related to findability (F2, F3) and reusability (R1, R1.1, R1.2, R1.3).

In particular, principle F2 requires provision of rich metadata without any explicit specification of what exactly constitutes ‘richness’ of metadata. Therefore, one of our goals was to compile a list of essential metadata elements (fields or properties) that are sufficient for describing mappings and crosswalks in a comprehensive manner. These elements should be rich enough to primarily enhance their discoverability, but should also capture the context of the mapping, and provide sufficient information so as to facilitate reuse of existing mappings, and address the redundancies that currently exist (and reusability).

To identify these key elements, we compiled a list of queries that would be used to retrieve these resources automatically or manually. Our aim was to identify the information necessary to retrieve mappings and therefore to derive from these queries a set of metadata elements which should provide this information. Queries were written in pseudo-code as shown in the following example: “*Give me the* ***mappings for the resource type*** *(class, concept, instance, property) with the* ***resource <identifier>*** *or with the* ***resource <label>***”.

Based on this list of queries, we established an extensive list of metadata elements (see Appendix - Table 2) for describing these resources, where this list ensures a comprehensive support for findability and reusability objectives. This list has been then extended with community inputs gathered during various workshops. We then compared the list of metadata elements with the SSSOM specification and aligned them based on their semantic meaning and intended content. Out of the 19 metadata elements we defined, only two were not fully addressed by the SSSOM metadata ("*Context*" and "*Source and target Semantic Artefact Name*"), and one highlighted some ambiguities in the naming of SSSOM properties ("*Mapping method*", see Appendix-Table 1: query 17). The latter can be mapped to the SSSOM "justification" property, which provides the type of approach used to create the mapping such as lexical matching, manual curation, semantic similarity threshold-based matching (see https://mapping-commons.github.io/sssom/mapping-justifications/), rather than the reason for mapping, as we interpreted. The "*Source and target Semantic Artefact Name"* (see Appendix-Table 1: queries 2,3,5,6 and 10) has been identified as necessary for Findability and is not considered in the SSSOM metadata model. However, this information can be accessed by connecting to ontology repositories, such as OntoPortal. [9]

In our analysis, the main missing element "*Mapping Context*", identified from community feedback gathered in various workshops, has been deemed essential for reusability. This missing element would be a perfect placeholder for linking the mappings with the practical information collected through the use of the Mapping Practice Framework presented in section 2.3.

This analysis demonstrated that SSSOM provides an extensive set of metadata descriptors which support both findability and reusability. Although we found some ambiguities and a missing element, we recommend using SSSOM metadata descriptors even if one chose not to use the model. It is important to note that SSSOM is still developing, and a number of items are under active discussion. Furthermore, it is worth noting that while many of the SSSOM metadata elements correspond to our identified metadata elements aimed at enhancing the findability and reusability of mappings, they are often labelled as optional or recommended within the SSSOM framework. To ensure full compatibility with our findings, it may be necessary to treat the corresponding SSSOM metadata elements as mandatory for supporting findability and/or reusability objectives. Alternatively, the directly overlapping set of terms could be considered "core" with additional fields being more use case defined; there is discussion on custom fields in the SSSOM open issues.

Another major topic related to the FAIR principles is the need of **Globally Unique Persistent and Resolvable Identifiers (GUPRIs)** to unambiguously identify and access the various information entities. This topic is covered by principles F1 and F3 which specify that GUPRIs should be attached to both data and metadata records and that metadata records should include a reference to the data. In the context of mappings, this means that every component within mappings and crosswalks should have a GUPRI assigned to it. Therefore, it is recommended to assign a GUPRI not only to the crosswalk itself but also to each individual mapping within it, along with their metadata. By assigning GUPRIs to individual mappings, it becomes possible to create collections or crosswalks that use mappings from other existing collections. Therefore we recommend that GUPRI should be provided for both individual mappings and crosswalks (collections of mappings). This recommendation would require modifying the SSSOM model which does not provide GUPRI for individual mappings. This is currently being discussed within the SSSOM community[4].

The reason behind associating GUPRIs with metadata for individual mappings and collections of mappings is that systems can first retrieve the metadata, perhaps through content negotiation, before obtaining mappings that are deemed relevant. Furthermore, separating the metadata record from the mapping itself may ensure the longevity of the metadata even if the mapping ceases to exist.

Mappings and crosswalks, whether they're published, stored, or used on the web, are considered as web-based resources that facilitate connections between two distinct online digital entities. We therefore recommend using web-based GUPRIs such as PURL or w3id.org that support content negotiation, giving clients the flexibility to choose their preferred representations of the metadata.

[4] https://github.com/mapping-commons/sssom/issues/359

Finally, the fourth topic, **Service and API**, largely covers the principles related to accessibility (A1, A1.1, A1.2, A2) and some aspects of findability (F4). The key service in the realm of mappings and crosswalks is undoubtedly the need for a mapping repository such as the Metadata Schema and Crosswalk Registry (MSCR), currently being developed by the FAIRCORE4EOSC project[5,6]. From what we discussed earlier, it is apparent that we have specific expectations for mapping repositories and their Application Programming Interfaces (APIs) to meet FAIR standards.

Mapping repositories need to ensure that every mapping, crosswalk, or related metadata has a GUPRI to fulfil the PIDs recommendations. Simply relying on associated metadata for the discoverability of mappings or crosswalks is not enough. Even with detailed descriptions, these resources might slip under the radar. To tackle this issue, it is advisable that such repositories incorporate an appropriate indexing mechanism, as well as exposing a search function for users.

When it comes to accessibility, mappings, crosswalks, or their related metadata records do not necessarily need to be openly available to everyone. Instead, they may have specific access conditions and criteria. Encouraging users to register for a repository account is a sensible practice. This allows for automated provenance tracking of the information about the owner or contributors of the mappings, crosswalks, or metadata, with the flexibility to establish user-specific rights, for example to update or modify, as needed. In line with the reusability principle, and as seen in the list of attributes, mappings should have a visible license that will indicate how they can be reused.

Mappings and crosswalks often become obsolete over time due to changes that may occur within the sources that are being mapped i.e. version change of an ontology, a metadata schema, concept obsolescence, semantic drift, etc. To avoid the frustration of searching for mappings and crosswalks that no longer exist or that are deprecated, it is highly recommended that the associated metadata remain accessible, particularly version metadata or associated timestamping of the resources that are mapped and to describe the status of the mapping e.g. tombstoned or deprecated, Leveraging these additional metadata fields to automate the tracking information of mapping and resource status would be a cost-effective way to enable a curator to update their mapping accordingly.

For enhanced accessibility, and thus improved findability and reusability, it is recommended to describe APIs using open standards and machine-readable solutions such OpenAPI[7] or smartAPI[8]. This empowers API users to understand how to integrate the API into their applications and even generate them automatically.

2.2. Toward a common exchange model for multiple types of mappings

Mappings and crosswalks can be used to bridge information elements at various levels (e.g. semantic artefact, metadata, data, API,...). The existing mapping models for exchange are focused mainly on one type of mapping i.e. entity mappings between semantic artefacts. Despite the difference of the type of information that is being mapped, the core model is similar as you are linking a source element with one or more target elements. This holds true whether you work at the semantic artefact level or at other levels (metadata, data, API). The main difference lies in the relation linking these elements. Therefore, in order to provide a common standard exchange model, we should consider a core model of mapping and provide the possibility to represent the wide variety of mappings. For instance, we should consider the cases of “complex” mappings where one source element is mapped to more than one target element. A simple example would be to map the metadata element “full name” in one metadata schema or ontology to the elements “first name” and “last name” in the target schema. To represent this relation we need to be able to identify the two target elements and also specify the order in which they should be used to be transformed into “full name”. As one of our core

[5] https://faircore4eosc.eu/
[6] https://faircore4eosc.eu/eosc-core-components/metadata-schema-and-crosswalk-registry-mscr
[7] https://swagger.io/specification/
[8] https://smart-api.info/

recommendations for FAIR mapping is to use as much as possible the SSSOM model, we should thus consider extending SSSOM. Its main advantage is the large amount of metadata to describe the mapping itself. As we saw previously, the metadata provided by SSSOM is rather extensive and applies to any type of mapping. Therefore, we propose extending the SSSOM model by extracting the core elements representing the mapping by itself i.e. *source_id*, *target_id* and *predicate_id* from the rest of the model (see Figure 1) and focusing on proposing variants of this core model specific to the other types of mappings. In practice, the core metadata remains unchanged, and thus it provides a common metadata model for all mapping types. Only a subset of the columns in the TSV model should be changed per mapping type. In order to build these variations, we will need to create an exhaustive typology to identify their peculiarity, and propose a specific extension for each mapping type.

uthor_id	mapping_date	**subject_id**	**subject_label**	**predicate_id**	**predicate_label**	**predicate_modifier**	**object_id**	**object_label**	confidence	mapping_justif
rcid:000)-0002-7 356-1779	2022-05-02	KF_FOOD :F001	apple	skos:exactMa tch	has exact match	*Can be "Not"*	FOODON: 00002473	apple (whole)	0.95	semapv:Huma n
rcid:000)-0002-7 356-1779	2022-05-02	KF_FOOD :F002	gala	skos:exactMa tch	has exact match	*Can be "Not"*	FOODON: 00003348	Gala apple (whole)	1	semapv:Huma n
rcid:000)-0002-7 356-1779	2022-05-02	KF_FOOD :F003	pink	skos:exactMa tch	has exact match	*Can be "Not"*	FOODON: 00004186	Pink apple (whole)	0.9	semapv:Huma n
rcid:000)-0002-7 356-1779	2022-05-02	KF_FOOD :F004	braeburn	skos:exactMa tch	has exact match	*Can be "Not"*	sssom:No Mapping		1	semapv:Huma n
rcid:000)-0002-7 356-1779	2022-05-02	KF_FOOD :F004	braeburn	skos:broadMa tch	has exact match	*Can be "Not"*	FOODON: 00002473	apple (whole)	1	semapv:Huma n

Figure 1: Core mapping model within SSSOM from [8]

2.3. Mapping practice framework

Besides the technical considerations, we consider it essential to establish a proper method for creating and curating mappings from a general perspective, as the main process of mapping creation remains the same regardless of mappings purpose or type. Thus, we collaborated with various communities across different domains through workshops[9,10], events, and an online survey[11] to gather their practices and processes. This collaboration aims to develop guidelines that encompass all necessary steps and considerations in the mapping creation process. It allowed us to collect information on existing mapping practices, tools & resources, governance, hosting and mapping activities and methodologies, subsequent sharing of outputs and of their availability.

From these workshops, we identified key issues from the community relating to mapping practices, what worked well, and what was difficult and determined the common steps in all such mapping activities, the objective being to derive from this a common and generic framework which would work for all future mapping activities. Simultaneously, we tried to identify the key properties that such a framework would have. An analysis indicated the following desired traits:

- generic framework; applicable without modification cross-domain
- modular; clearly divided into identifiable 'phases'
- agile; flexible enough to accommodate a diversity of use cases
- context, provenance; a means to capture metadata on context and provenance

[9] https://fair-impact.eu/events/fair-impact-events/documenting-mapping-community-practices
[10] https://fair-impact.eu/events/fair-impact-events/developing-mapping-process-framework
[11] https://fair-impact.eu/news/collecting-ways-doing-mappings-take-survey

- iterative/refinement; living mapping that can persist and be refined (versioned)
- findability; reduce duplication of effort
- maintenance; a commitment to keep up to date
- confidence was also mentioned, and discussion on how confidence in the mappings should be reflected

Given the identification of the different ‘phases’ that constitute a mapping, as well as the traits that such a framework should possess, we built a framework, which consists of 5 phases: ‘pre-mapping’, ‘mapping’, ‘review’, ‘hosting’, and ‘maintenance’ (Figure 2). The purpose of the framework is to guide researchers who wish to undertake a mapping to systematically address practical considerations, and hence collect the appropriate context and use case underlying it, and thereby the key metadata that needs to be shared with the mapping itself, to make it useful and usable by other researchers.

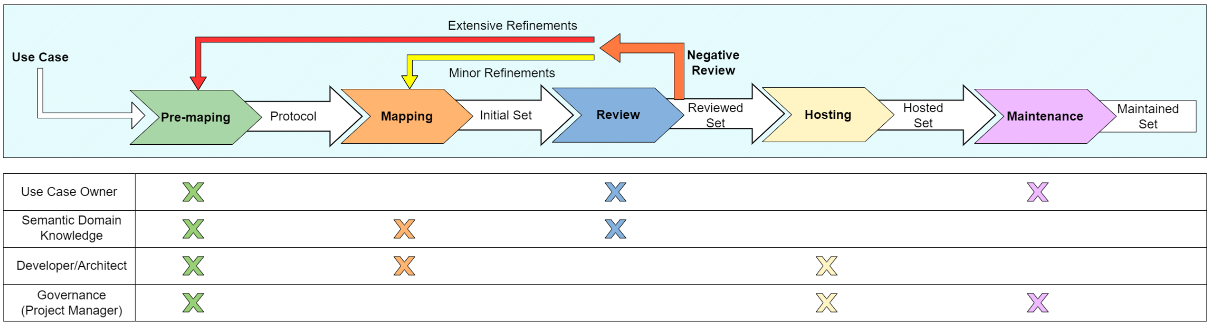

Figure 2: Diagram of Practical Framework

The **pre-mapping phase** is the foundation of the process, where key decisions must be made that influence or define downstream tasks. For instance, this phase involves discussion with various stakeholders to define concretely the source and targets for mapping, the mechanisms or methodologies to be used, and for instance which licence will be applied to the resulting work. The requirements to enter this phase are simply a defined ‘use case’, with the output being a ‘protocol’, which in turn feeds the next phase: mapping. The protocol consists of all the decisions made in this pre-mapping phase, including the chosen mechanisms, methodologies, and other relevant factors.

The **mapping phase** is the practical implementation of the mapping itself, whether by manual work, or through automated or semi automated mechanisms, and as defined in the protocol. The output of this phase is an initial mapping file, in the format described by the protocol, that enters the ‘review’ phase.

The **review phase** assesses the applicability of the mapping exercise in the context of the use case. While ‘confidence’ was a desired property expressed in community events and workshops, for the moment at least we have incorporated this as a reviewer-led judgement. Depending on the collective decision of the reviewers, the mapping file can be accepted and move to the hosting phase, or else the process will have to fall back to the protocol (for minor refinements), or else require further discussion, back in the pre-mapping phase, to get input from the wider stakeholders.

In the **hosting phase**, the ‘accepted’ mapping file is uploaded to the agreed hosting resource, along with metadata as appropriate for that resource. Resources that surface sufficient of the indexed metadata should have been selected in the pre-mapping phase, as well as a suitable license. Once hosted, the mapping process moves into the ‘maintenance’ phase.

The **maintenance phase** targets the long term availability, accuracy and ‘openness to feedback’ of the mapping itself, something that is defined in discussions in the pre-mapping phase. Fundamentally, this phase is initiated once the maintenance mechanisms, *if any*, agreed in the pre-mapping phase are activated, and the file remains in this phase in perpetuity.

As is obvious from this process, it entails numerous discussion and decisions that must be documented. To facilitate this process, we created a narrative guide and a spreadsheet. The narrative guide defines each phase of the framework (pre-mapping, mapping, review, hosting, maintenance). For each phase, there is a description of the activities constituting the phase, a list of anticipated stakeholders that should be involved, an indication of the types of questions that need to be answered, as well as the formalisation of inputs and outputs for each phase, and hence what is needed to move between phases. The 'spreadsheet' provides a convenient means by which discussion and decisions can be recorded. It is not intended to be used as a format in of itself, but we anticipate it may provide a lightweight mechanism for simpler mappings. The spreadsheet is organised into sheets per phase, with a top sheet to record higher level metadata (date of mapping, version, participants per phase, etc). The materials (narrative guide & spreadsheet) are available on zenodo[12]. This mapping practice framework is currently being tested by different communities.

3. Future work

The full first version of these FAIR mapping recommendations will be published in the near future, and will be followed by a period of time for gathering feedback for improvements which will lead to a consolidated version of the recommendations. In addition, we will be working with the communities to identify use cases involving a wide range of mapping types in order to develop the specific mappings models that will be used to create the common exchange model. As this common model relies on SSSOM, we will work extensively with the SSSOM community to establish the integration. Finally, the mapping practice framework will be refined together with the communities during at least another dedicated workshop.

To strengthen and extend our collaborations and the outreach of our work, a Working Group (WG) within Research Data Alliance (RDA) focusing on FAIR mappings is being created. The aim of this group is mainly to engage with the wider research data community and share all the outcomes of our project focused on mappings and crosswalks. This WG will then provide input to the recommendations and will contribute to collect more use cases for defining the common model. In addition to these activities, we will work on creating a terminology describing the different types of mappings based on the collection of use cases and in collaboration with the international communities. This terminology will play a major role in disambiguating communication about mappings and will support the possible categorization of mappings to enhance the search experience.

4. Conclusion

In conclusion, this paper has outlined our ongoing efforts aimed at fostering FAIR-compliant mappings and crosswalks, which are key components for enhancing semantic interoperability and data integration across diverse information systems. We have set out on a path to establish technical requirements for FAIR mappings and crosswalks, provide a practical framework to guide the creation and maintenance of mappings and crosswalks, and define a machine-actionable common exchange model.

Our forthcoming recommendations for FAIR mapping will soon be published, inviting feedback for further refinement and improvement. Through validation of the practical framework during our last workshop, we have identified key challenges and opportunities, paving the way for refining our proposed approach. Our commitment to testing and refining this framework within a wider community underscores our dedication to ensuring its effectiveness and relevance.

Furthermore, the establishment of a RDA working group focusing on FAIR mappings marks a significant step forward in sustaining and expanding the outcomes of the project. Ongoing

[12] https://zenodo.org/doi/10.5281/zenodo.12521431

discussions on typologies of mappings and crosswalks are expected to drive further advancements in this domain.

As we progress, we welcome collaboration and feedback from the research community to collectively advance the field of mappings and crosswalks, ultimately contributing to enhanced data interoperability and accessibility on a global scale.

Acknowledgements

This work has been funded by the European HORIZON programme under the grant agreement no. 101057344 (FAIR-IMPACT). The authors would like to thank FAIR IMPACT partners namely: Clement Jonquet, Sophie Aubin, Guillaume Alviset, Gabi Mejias, Alain Monteil, Rene van Horik, Pascal Flor, Xeni Kechagioglou, Parham Ramezani, María Poveda-Villalón and Daniel Garijo. The authors also would like to thank in particular Nico Matentzoglu for his support with SSSOM and his active contribution to the discussions. The authors would like to thank all the participants in our workshops for their active contribution in this work.

References

[1] Broeder, D., Budroni, P., Degl'Innocenti, E., Le Franc, Y., Hugo, W., Jeffery, K., Weiland, C., Wittenburg, P., & Zwolf, C. M. (2021). SEMAF: A Proposal for a Flexible Semantic Mapping Framework (1.0). Zenodo. https://doi.org/10.5281/zenodo.4651421

[2] Yann Le Franc, Luiz Bonino, Hanna Koivula, Jessica Parland-von Essen, & Robert Pergl. (2022). D2.8 FAIR Semantics Recommendations Third Iteration (V1.0). Zenodo. https://doi.org/10.5281/zenodo.6675295

[3] European Commission, Directorate-General for Research and Innovation, Corcho, O., Eriksson, M., Kurowski, K. et al., EOSC interoperability framework – Report from the EOSC Executive Board Working Groups FAIR and Architecture, Publications Office, 2021, https://data.europa.eu/doi/10.2777/620649

[4] Nyberg Åkerström, W., Baumann, K., Corcho, O., David, R., Le Franc, Y., Madon, B., Magagna, B., Micsik, A., Molinaro, M., Ojsteršek, M., Peroni, S., Scharnhorst, A., Vogt, L., & Widmann, H. (2024). Developing and implementing the semantic interoperability recommendations of the EOSC Interoperability Framework (27 March 2024). EOSC Association AISBL. https://doi.org/10.5281/zenodo.10843882

[5] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

[6] David, J., Euzenat, J., Scharffe, F., Trojahn dos Santos, C., The Alignment API 4.0 (2011), Semantic Web, vol. 2, no. 1, pp. 3-10, 2011, iOS Press, https://doi.org/10.3233/SW-2011-0028

[7] Euzenat, J., Meilicke, C., Stuckenschmidt, H., Shvaiko, P., Trojahn, C. (2011). Ontology Alignment Evaluation Initiative: Six Years of Experience. In: Spaccapietra, S. (eds) Journal on Data Semantics XV. Lecture Notes in Computer Science, vol 6720. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-22630-4_6

[8] Nicolas Matentzoglu, James P Balhoff, Susan M Bello, Chris Bizon, Matthew Brush, Tiffany J Callahan, Christopher G Chute, William D Duncan, Chris T Evelo, Davera Gabriel, John Graybeal, Alasdair Gray, Benjamin M Gyori, Melissa Haendel, Henriette Harmse, Nomi L Harris, Ian Harrow, Harshad B Hegde, Amelia L Hoyt, Charles T Hoyt, Dazhi Jiao, Ernesto Jiménez-Ruiz, Simon Jupp, Hyeongsik Kim, Sebastian Koehler, Thomas Liener, Qinqin Long, James Malone, James A McLaughlin, Julie A McMurry, Sierra Moxon, Monica C Munoz-Torres, David Osumi-Sutherland, James A Overton, Bjoern Peters, Tim Putman, Núria Queralt-Rosinach, Kent Shefchek, Harold Solbrig, Anne Thessen, Tania Tudorache, Nicole Vasilevsky, Alex H Wagner,

Christopher J Mungall, A Simple Standard for Sharing Ontological Mappings (SSSOM), Database, Volume 2022, 2022, baac035, https://doi.org/10.1093/database/baac035
[9] Clement Jonquet, John Graybeal, Syphax Bouazzouni, Michael Dorf, Nicola Fiore, et al.. Ontology Repositories and Semantic Artefact Catalogues with the OntoPortal Technology. ISWC 2023 - 22nd International Semantic Web Conference, Nov 2023, Athens, Greece. pp.38-58, ⟨10.1007/978-3-031-47243-5_3⟩. ⟨hal-04088537v2⟩

A. Appendix

Table 1
List of queries

No.	Query
1	Give me the mappings for the resource (class, concept, instance, property) with the resource <identifier> or with the resource <label>
2	Give me all the mappings for all the resources (class, concept, instance, property) in a Semantic Artefact (SA) with the SA <identifier> or with the SA <name>
3	Give me the mappings between SA1 and SA2 with SA1 <identifier1> and SA2 <identifier2> or with SA1 <name1> and SA2 <name2> *(NOTE: need to take in account the possibility of directionality: from SA2 to SA1 as well)*
4	Give me the number of mappings for the resource (class, concept, instance, property) with the resource <identifier> *(NOTE: will be computed, no needed extra metadata)*
5	Give me the number of mappings for all the resources in a Semantic Artefact (SA) with the SA <identifier>
6	Give me the number of mappings between SA1 and SA2 with SA1 <identifier1> and SA2 <identifier2> or with SA1 <name1> and SA2 <name2> *(NOTE: need to take into account the possibility of directionality: from SA1 to SA2, from SA2 to SA1)*
7	Give me all the mappings for SA <identifier> or SA <label> where the mapping relations are of type <T> (e.g., skos:exactMatch, owl:sameAs...).
8	Give me the number of mappings for SA <identifier> or SA <label> where the mapping relations are of type T (e.g., skos:exactMatch, owl:sameAs...)
9	Give me all the metadata about one specific mapping given its <identifier>
10	Give me the source and the target SAs of one specific mapping given its <identifier>
11	Give me all the mappings that respect one specific license given the <license>
12	Give me all the mappings for SA <identifier> that were created before/after <DATE>
13	Give me all the mappings produced by one specific author (person/organisation) given the author <identifier> or author <name>
14	Give me all the mappings that were created by specific person/organisation given the author <identifier>
15	Give me all the mappings that were reviewed by specific person/organisation given the reviewer <identifier>
16	Give me all the mappings that were created using one specific software tool

17	Give me all the mappings that were created using the mapping method M (manual, automated, lexical,..)

Table 2
Metadata elements with their SSSOM equivalent and requirement level

Metadata element	SSSOM equivalent	Mandatory by SSSOM?
Author	author_id author_label	No
Creator	creator_id creator_label	No
Reviewer	reviewer_id reviewer_label	No
Accuracy score	confidence	No
Date of creation of this version	mapping_date	No
Version	mapping_set_version	No
License	license	Yes (mapping) No (crosswalks)
Context		
Justification	mapping_justification	Yes
Clarification of the rationale of this mapping	curation_rule	No
Source and target SAs‘ IDs	object_source subject_source	No
Source and target SAs’ NAME		
Version of the SAs	object_source_version subject_source_version	No
Mapping method	mapping_justification	Yes
Software used to create the mapping (if any)	mapping_tool	No
Specific types of SA that are mapped	object_type subject_type	No

Source term and target term label	subject_label object_label	No
Source term and target term ID	subject_id object_id	Yes
Mapping predicate	predicate_id	Yes