Project Title	Expanding FAIR solutions across EOSC
Project Acronym	FAIR-IMPACT
Grant Agreement No.	101057344
Start Date of Project	2022-06-01
Duration of Project	36 months
Project Website	fair-impact.eu

1 FAIR-Impact's workshop: Why semantic mappings matter and how to make them FAIR?

Work Package	**WP 4, Metadata and ontologies**
Lead Author (Org)	**Jana Martínková (eSDF), Yann Le Franc (eSDF)**
Contributing Author(s) (Org)	**Alejandra Gonzalez Beltran (UKRI-STFC), Pascal Flohr (DANS), Sophie Aubin (INRAE), Maria Poveda Viallalon (UPM), Gabriela Mejias (DataCite), Nick Juty (UNIMAN), Carole Goble (UNIMAN)**
Date	**2023-07-13**
Version	**V1.0**

Dissemination Level

X	PU: Public
	PP: Restricted to other programme participants (including the Commission)
	RE: Restricted to a group specified by the consortium (including the Commission)
	CO: Confidential, only for members of the consortium (including the Commission)

Funded by the European Union

2 Versioning and contribution history

Version	Date	Author	Notes
0.1	1.06.2023	Jana Martínková, Alejandra Gonzalez-Beltran	Initial write-up summarising the workshop
0.2	1.06.2023	Pascal Flohr, Gabriela Mejias, Sophie Aubin, Carole Goble, Nick Juty, Maria Poveda Villalon	Analysis of inputs from workshop
0.3	21.06.2023	Yann Le Franc	Executive summary, conclusion

3 Table of Contents

Funded by the European Union

1.1 List of Tables

Funded by the European Union

TERMINOLOGY

Terminology/Acronym	Description
API	Application Programming Interface
BioDT	Biodiversity Digital Twin
CERIF	Common European Research Information Format
ChatGPT	Chat Generative Pre-trained Transformer
CSV	Comma-separated Values
DCAT	Data Catalog Vocabulary
DISSCo	Distributed System of Scientific Collections
DTR	Data Type Registry
EBI	European Bioinformatics Institute
EMBL	European Molecular Biology Laboratory
ENVRI	Environmental Research Infrastructures
EOSC	European Open Science Cloud
EPOS	European Plate Observing System
FAIR	Findable, Accessible, Interoperable, Reusable
JSON	JavaScript Object Notation
LCSH	Library of Congress Subject Headings
MSCR	Metadata Schema and Crosswalk registry
NSC	Natural science collection
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OLS	Ontology Lookup Service
OWL	Web Ontology Language
OXO	Ontology Xref Service
PID	Persistent Identifier
SIOC	Semantically-Interlinked Online Communities
SKOS	Simple Knowledge Organization System
SSH	Secure Shell
SSSOM	Simple Standard for Sharing Ontological Mappings
TSV	Tab-Separated Values
TTL	Turtle

Funded by the European Union

Executive Summary

This document is a short report on the first workshop organised by the FAIR Impact project on the topic of mappings. We are summarising the various presentations which were given during the workshop. This short summary is then complemented by a short summary of the inputs collected from the participants on four main topics: FAIR Mappings, methodologies for mappings, mappings tools and identifying the various types of mappings. In conclusion, we are providing a short overview of the projected outcomes of this workshop.

Funded by
the European Union

1 Introduction

FAIR-IMPACT is a European project aimed at expanding FAIR (findable, accessible, interoperable and reusable) solutions across Europe and supporting the establishment of the European Open Science Cloud (EOSC). Within this project, Task 4.4 focuses on mappings and cross-walks, specifically on how to make them FAIR. This task is part of Work Package 4, which covers the lifecycle of semantic artefacts.

The main objectives of our Task 4.4 are twofold. Firstly, we aim to understand the requirements for FAIR mappings and establish guidelines on how to make mappings FAIR. We recognise the importance of making the sharing of FAIR mapping easier, and therefore, we will propose a common exchange model that can be easily understood and utilised by machines, based on existing models such as SSSOM. This model should cover a wide range of mappings (from simple to complex) and will facilitate the smooth exchange of mapping information between different systems.

Secondly, we strongly believe in the importance of collaborating with the wider community. By actively involving researchers, data providers, and domain experts, we aim at gathering valuable insights and feedback from the different communities to reach a consensus on working with mappings. This collective effort will be instrumental in developing the exchange model, identifying effective methodologies, and establishing mapping best practices.

This report represents the outcomes of our first workshop in a series designed to interact with the different communities. It includes detailed descriptions of all the presentations and discussions that took place during the workshop. Additionally, the inputs collected from the participants are documented and thoroughly analysed.

Funded by the European Union

2 FAIR mappings - General Information

The first workshop of Task 4.4 of the FAIR-IMPACT project, titled '*Why Mappings Matter and how to make them FAIR?',* was successfully conducted online on April 13th 2023. This workshop --the first one of a workshop series around semantic mappings– introduced participants to the motivation behind doing mappings between semantic artefacts and how those mappings could benefit them even more by being FAIR.
The workshop featured an engaging series of 'Show and Tell' presentations, describing use cases and practices from a wide range of domains. These presentations served as examples of used mappings and possible methods.
The main objectives of this workshop were: to start interacting with different communities, to provide a place for initial discussion, and to define collaboratively the initial requirements for developing a useful framework centred around FAIR mappings, including the requirements to adhere to the FAIR principles.
The workshop boasted an impressive attendance, with 120 registered participants, and a peak of 95 participants (organisers and speakers included), many of whom were actively engaging. The event drew participants from 24 countries, representing 24 named organisations and communities, and approximately 32 domains of knowledge. Notably, only 30% of participants identified themselves as ontologist-metadata experts. The remaining attendees consisted mainly of software developers (15%), data stewards (19%) and project managers (18%), and surprisingly, almost 20% of participants do not have a direct involvement with mappings at all. This diverse representation highlights the broad impact and relevance of semantic mappings across different roles and areas of expertise.
Nearly half of the participants expressed their interest in contributing to the ongoing task of developing a FAIR mapping framework, and several attendees were willing to share their use cases. This response demonstrates the commitment within the community to participate actively and contribute to the advancement of FAIR mappings.

3 Summary of the presentations

The workshop included multiple presentations - from an introduction to the topic, to a series of 'Show and Tell' presentations where use cases and practices from a range of communities were presented – spanning domains, research infrastructures, projects, task forces, and working groups. Then a series of presentations from other projects and outputs were made.
During the introduction, **Yann Le Franc** presented:

- the objectives and schedule of the workshop
- the FAIR-impact project (see the video at time 13:30)
- a brief state-of-the-art: "Why do we need mappings/crosswalks and why FAIR mappings/crosswalks would be a good idea?" (see video at time 25:05)

During the workshop, collaborative notes were taken and a Miro board was used during the breakout sessions, which are described later in this report. The slides are available on Zenodo (see https://doi.org/10.5281/zenodo.7828268).

Funded by the European Union

3.1 Biomedical perspective by Henriette Harmse

Henriette Harmse from EMBL-EBI presented a Biomedical perspective on why mappings matter and how to make them FAIR (see video at time 42:39). Henriette motivated the use of ontologies at EBI due to two data related problems when trying to integrate data: different words may refer to the same concept or the same word may refer to different concepts. EBI created multiple tools for the use of ontologies: Ontology Lookup Service (OLS) for ontology search, ZOOMA for ontology annotation, OXO for ontology mapping and ONTOString that works on top of ZOOMA for string to term mapping. Henriette also explained why support for SSSOM is needed.

3.2 Geoscience perspective by Keith Jeffery

Keith Jeffery introduced the environmental and geoscience perspective (see video at time 59:43) with a presentation on "Semantic Mapping in ENVRI (Environmental Research Infrastructures) and EPOS (European Plate Observing System)". He presented on multiple aspects of the work, including vocabularies, harmonisation, and a subset of common terms and catalogue. As regards to vocabularies, ENVRI is finding it difficult to find vocabularies that are ready to use, and is therefore currently going through a process which defines the different concepts and how to publish them. The Research Infrastructure also works together with 'thematicians', and produces an incubator register that aims to get a standard set of terms and definitions, but it remains a challenge to identify relationships. Related to harmonisation, the main challenge for EPOS is to harmonise concepts across domains - such as seismology and volcanology - as it is a slow and human-resource-heavy work. There is an important question to answer: is it worth doing this work and mapping terms across domains?
Related to common subsets, the community is working on agreeing on a very small number of common terms to all subdomains (atmosphere, hydrosphere, geosphere, biosphere). Finally, Keith described the catalogue that is based on the CERIF model, and relies on metadata conversions – i.e. manual mappings – between the metadata records that describe digital assets to EPOS-DCAT-AP, and finally to the CERIF model. EPOS-DCAT-AP is an extension of DCAT-AP, and all of this work in progress would benefit from the work of mappings discussed in the workshop.

3.3 Social Science perspective by Luca De Santis

The Social Science and Humanities (SSH) perspective was presented by **Luca De Santis** (see video at time 1:15:49), as he described the GoTriple case. GoTriple is a multilingual discovery platform for the SSH domain, so it can be used as a search engine to discover publications, projects, authors, and researcher profiles. It provides a configurable harvesting and processing pipeline relying on the OAI-PMH protocol. They are traversing the FAIR journey, improving the metadata they have on the platform. Their data model relies on schema.org and SIOC (Semantically Interlinked Online Communities), and they use multiple data enrichment techniques to populate the metadata. The GoTriple system proposes a classification for SSH based on 27 disciplines. The TRIPLE vocabulary is a multilingual SKOS-based taxonomy of concepts related to SSH. It currently contains 3375

Funded by the European Union

manually-curated concepts, and uses labels in 11+ languages, and every concept is linked to the Library of Congress Subject Headings (LCSH), as well as other mappings to taxonomies such as Wikidata and CESSDA ELSST. They are formalising all the vocabularies and mappings into a TRIPLE ontology, which documents and links the controlled vocabularies used during enrichments. The data on the website is available via RESTful APIs, and is provided in JSON or JSON-LD.

3.4 BioDT and DISSCo perspective by Claus Weiland

The final presentation of the 'Show and Tell' session was on the BioDT (Biodiversity Digital Twin) and DISSCo perspective by **Claus Weiland** (see video at time 1:29:06). DISSCo is a Distributed System of Scientific Collections that focuses on natural science collections (NSCs), including physical specimen collections, subject databases and aggregators, as well as data from research platforms. Its aim is to unify digitally all the current fragmented landscape of European NSCs into a single virtual data collection following common curation practices and access policies. The core data model is embedded in the concept of FAIR Digital Objects, i.e. a unified conceptual data model that enables interoperability by PID/model layer.
The main topics presented were:

- Concerns about existing mappings and what to do with these.
- Where to put mappings (repositories, but also where to put them temporarily/ad hoc while a more refined registry is being built).
- The need for learning resources about mappings (from the beginner level onwards).

3.5 Semantic Interoperability Task Force

In the presentation, **Milan Ojstersek** discussed the work of the EOSC Task Force on Semantic Interoperability (see the video at time 1:46). The task force is divided into three subparts: Semantic Interoperability landscape and metadata, Semantic artefacts catalogues, and Case studies/use-case studies. Milan explained the concept of metadata and its types, including descriptive, structural, administrative, reference, statistical, and legal metadata as defined by the Task Force. He also discussed metadata schemes, the need for mappings between them, and common approaches to metadata mapping such as field-to-field, value-based, business rule-based, and ontology-based mapping. The presentation highlighted the importance of metadata and the efforts to enhance interoperability through mapping.

3.6 SSSOM: a machine actionable model for simple mappings

During the presentation, **Nicolas Matentzoglu** introduced the Simple Standard for Sharing Ontological Mappings (SSSOM) (see the video at time 2:08). The presentation provided insights into understanding mappings and highlighted the distinction between two categories of mappings: entity mapping and schema mapping. The importance of semantic precision in mappings was emphasised, as well as the significance of making mappings shareable.

Funded by the European Union

The presentation also included a brief overview of the SSSOM metadata model, which encompasses mapping justification. This aspect of SSSOM involves providing reasons to justify the mapping decisions made by the author.

3.7 FAIRCORE4EOSC: toward a mapping registry

In the presentation, **Tommi Suominnen** introduced the FAIRCORE4EOSC project (see the video at time 2:29), which aims to develop and implement essential components for the European Open Science Cloud (EOSC). The presentation provided insights into the project's focus areas, including persistent identifiers (PIDs), metadata and ontologies, interoperability, and research software.
The FAIRCORE4EOSC project comprises nine components, two of which were highlighted during the presentation: the Data Type Registry (DTR) and the Metadata Schema and Crosswalk Registry (MSCR). The Data Type Registry (DTR) provides a hierarchical model of basic data types descriptions. Through the DTR toolkit, users can register their data types and conduct searches to prevent duplication. The Metadata Schema and Crosswalk registry (MSCR) component is used for creation, registration, and versioning of schemas and crosswalks with PIDs.

3.8 Mapping with ChatGPT

This section of the workshop highlighted the transformative potential of Large Language Models (LLMs), particularly ChatGPT, in academic metadata schema mapping. **Kristian Garza**, from DataCite, showed the dynamic capabilities of ChatGPT through three examples (see the video at time 3:33).
The first one shows the transformation of metadata from one schema to another, specifically from Crossref schema to Datacite schema. The output generated consisted of a valid Datacite schema, showcasing the successful conversion.
The second scenario delved into the transformation from a prospective schema to an existing one, specifically targeting the DCAT schema. The output was a valid schema, showcasing the remarkable adaptability of ChatGPT to handle yet-to-be-released schemas, and its ability to perform accurate translations.
The final example showed the generation of crosswalks between Datacite schemas and Schema.org. This produced fairly accurate mapping when compared to mappings done by experts.

4 Summary of the discussion

During the workshop, a structured and organised discussion was held, aiming to gather valuable inputs from participants. To facilitate the discussion a Miro board was utilised, providing a collaborative virtual space for attendees to contribute their ideas and perspectives. The discussion was around four main topics, which were selected to address key aspects of the FAIR mapping:

- **Requirements for FAIR mappings** - Participants engaged to define the essential requirements for FAIR mappings.

Funded by the European Union

- **Different types of mappings** - Participants shared their knowledge and experiences regarding the different types of mappings they were familiar with or commonly employed in their work.
- **Mapping tools** - List of tools that participants use with the desired needs for features or tools in general, and what are the requirements for these tools.
- **Mapping methodologies** - Participants had the opportunity to share the mapping methodologies and guidelines used by their communities or organisations.

5 Summary of the collected inputs

The following section provides a comprehensive summary of the inputs gathered during the workshop, highlighting the key points and insights shared by the participants. These inputs revolve around the four main topics that were discussed during the workshop: requirements for FAIR principles, types of mappings, existing tools and requirements for them and mapping methodologies.

5.1 Requirements for FAIR principles

The input of the participants regarding the requirements for FAIR mappings can be divided into three main topics:

1. Mapping metadata
2. Mapping repositories
3. Mapping governance

The input will be further used to create a first version of specifications or recommendations of the FAIR principles applied to mappings and crosswalks.

5.1.1 Mapping metadata

The need for rich metadata for mappings, including PIDs, is clear from the input of the workshop participants, and not surprising given the centrality of metadata in the FAIR principles. Key points raised by the participants are:

- The need to give a **persistent ID (PID)** to mappings and mapping attributes. There appears to be discussion on, and therefore the need to reach consensus on, exactly what needs PIDs (what granularity is needed). Should ‘everything’ get a PID, or certain aspects only (i.e. alignment, individual mappings, versions)?
- The need for (machine readable) **metadata (/standards)** to identify / include how and why mappings were created and if they can be reused:
    - The mapping justification
    - The context in which the mapping was created
    - A clear definition of the scope of the mapping
    - The mapping creator’s intentions
    - The licence
    - Definitions of source and target data structures
    - The meanings of mappings

Funded by the European Union

- There was positive feedback to the SSSOM metadata (specifically mapping date, mapping justification; request to add “context”), but also the wish to extend it to complex mappings.
- The need for (controlled) vocabularies (terms) for mapping metadata

5.1.2 Mapping repositories

A considerable amount of the participants’ feedback regarding requirements of FAIR mappings concerned **the need to be able to find, access, and register mappings (metadata) through repositories / registries**. The use of lookup services and APIs helps not only findability and accessibility but also interoperability, since the communities will be able to find and reuse relevant mappings rather than come up with new ones. The idea of initiating the process of setting up a federation of registries including different disciplines (life sciences, SSH, ESS) was also mentioned. Questions were raised regarding funding and sustainability.

5.1.3 Mapping governance

There is a need for clarity on the **governance** of a mapping / crosswalk. What was specifically mentioned was the question of who is going to operate and sustain a mapping registry, but one could also take this further to **who is responsible** for a mapping / crosswalk and decides, for example, on changes to be made etc. This should be clarified in the metadata.

5.2 Types of mappings

During the discussion, the issue of accurate understanding of the mapping was often encountered. The biggest source of discussion was the difference between entity mapping and schema (or data models/structures) mapping. These are very different, but entity mapping is usually part of schema mapping.
The contributions related to the types of mappings can be qualified according to:

1. Nature of the mapped elements
2. Purpose of the mapping

5.2.1 Nature of the mapped elements

The contributions from the participants can be conceptually divided according to the origin of the mapping as well as the origin of the mapped elements. By analysing the contributions to this topic, four natures of mapping were determined, specified in the table below:

Table 1 – Natures of mapping

From	To
Data model	Data model
Semantic artefact	Semantic artefact
Data/free text	Semantic artefact
Data	Rules

Funded by the European Union

5.2.2 Purpose of the mapping

1. **Data standardisation:** the mapping is used to transform the data from a proprietary or local representation to a standardised, shared, documented data model. This contribution was made in the context of archiving, which requires transforming data formats to more sustainable formats.
2. **Data enrichment:** the mapping is used to document/annotate in a formal way a model or the data itself with entity types, characteristics and relations.
3. **Indexation:** the mapping is used to enhance discovery and search results by describing data with concepts that provide synonyms, terms in other languages, hierarchies to extend search, facets, etc.
4. **Integration:** the mapping between term A and term A' is used to infer the semantic equivalence of the data they respectively define or document.

Some inputs relate to value/quality of the mappings (reproducibility capacity)

1. **Low value** : fully automatic : no added value, reproducible
2. **Relative value** : automatically produced and curated : mappings are selected by an expert (wrong mappings are discarded; missing mappings may or may not be added)
3. **High value** : produced by an expert / several experts

Value augments if guidelines were used to define criteria for respectively valid and invalid mappings.

5.3 Existing tools and requirements for tools

The participants undertook two activities:

1. Identify required practical mapping tool features
2. Collect tools that were commonly used

Given time constraints a follow-on study is needed to systematically analyse and map the tools against the features.

An overall message is that spreadsheets and simple TSV/CSV formats are commonly used, as the tool; within tools, exchange between tools; or as output of tools. Regardless of the tool, spreadsheets/CSV must be supported, and features for conversion into more formal or "better" formats are needed.

5.3.1 Identify practical mapping tool features

Features broadly adhere to:

- FAIR mappings - that mappings should be Findable, Accessible, Interoperable and Reusable as first class entities in their own right.
- Support for common formats (CSV, TSV) and standardised representation (SSSOM).
- Support for the life cycle of mappings (see https://rdmkit.elixir-europe.org/data_life_cycle) - acquisition, registration, validation, management, legacy management.
- Access to mappings through PID resolution, registries, APIs, user interfaces.

Funded by the European Union

Table 2 – Practical mapping features

Formats	Support simple (CSV, TSV) and legacy formats, including SKOS ETL Libraries to extract data from common formats to defined data structures to format neutral mapping for further use
Models	Support SSSOM, expressed through JSON; and JSON in general
Registry	Registries/storage for mappings, including legacy mappings Namespace to host SSSOM mappings for PIDs Solution should be based on communities of practice, rather than a monolithic centralised system; communities of practice should have their own registries, which can be interlinked or communicate with each other. https://mapping-commons.github.io/sssom/MappingRegistry/
Crosswalks and mappings management	Facilitate crosswalks between simple formats Features and processes to manage the lifecycle of mappings: edit/create/deprecate/analyse/version Reversible mappings, (e.g convert between SSSOM TSV and OWL/TTL or reverse merge a TSV into an ontology)
Acquisition	(semi)-automated workflows that can create/extract mappings, Continually update mappings with versioning and legacy support "FAIR matching" report the reasons for establishing a match User friendly tools for manual mapping
Validation	Support for validating mappings
API	APIs for making, accessing and managing mappings

5.3.2 Tools that are commonly used

The session collected input from participants on which mapping tools and platforms they commonly used. The participants also highlight the Ontology Alignment Evaluation Initiative as a source of tools.

Funded by the European Union

Table 3 – Tools that are commonly used

Cocoda https://coli-conc.gbv.de/cocoda/	A web application to manage and create mappings between knowledge organisation systems, such as classifications, authority files, and thesauri.
EMBL-EBI OxO https://www.ebi.ac.uk/spot/oxo/	A service for finding mappings (or cross-references) between terms from ontologies, vocabularies and coding standards. OxO imports mappings from a variety of sources including the Ontology Lookup Service and a subset of mappings provided by the UMLS.
Alignment Maker Light (AML) https://github.com/AgreementMakerLight	An automated ontology matching system based on the use of element-level matching techniques supported by background knowledge. "FAIR matching" export the reasons for establishing a match.
Mapping commons - https://mapping-commons.github.io/	Gathers mappings from various sources with the goal of offering them as simple to read, well provenanced TSV files in the SSSOM format. Focus on biomedical ontologies and controlled vocabularies, e.g. Mouse-Human Ontology Mapping Effort.
Nanopublications - https://nanopub.net/	A nanopublication is the smallest unit of publishable information expressed in a knowledge graph format that is formal and machine-interpretable, that could be adapted to SSSOM
Ontology Access Kit - http://berkeleybop.github.io/software/OAK/	A Python library for executing common ontology operations over a variety of backends, providing a collection of interfaces for various ontology operations.
Mapping.bio https://mapping.bio/	Mapping Service V. 0.12.3 developed by the Horizon Europe BioDT project, using SSSOM as a representation.
OnaGUI https://github.com/lmazuel/onagui	OnAGUI - Ontology Alignment GUI - Software to help automatic or manual realisation of ontologic alignment. OnAGUI can align SKOS thesauri, OWL ontologies, or arbitrary RDF entities (based on common labelling properties).

Funded by the European Union

Protégé https://protege.stanford.edu/software.php#desktop-protege	Well established feature rich ontology editing environment with full support for the OWL 2 Web Ontology Language, and direct in-memory connections to description logic reasoners like HermiT and Pellet.
OntoPortal https://ontoportal.org/ Instances: BioPortal, AgroPortal, EcoPortal etc https://ontoportal.org/about/	A generic technology to build ontology repositories or semantic artefact catalogues with support to store, retrieve, import and explore mappings between ontology terms, and get lexical overlap of an ontology with all the other ones in a portal.
ontoCAT https://rdrr.io/bioc/ontoCAT/	The ontoCAT R package provides a simple interface to ontologies described in widely used standard formats, stored locally in the filesystem or accessible online.
Spreadsheets: Google Sheets, Excel	Using off-the-shelf spreadsheets to manage mappings
Scibite https://scibite.com/solutions/ontology-management/	Commercial platform for ontology management

5.4 Mapping methodologies

This session, carried out in parallel with the rest of the topic and presented at the end of the common slot, ended up with not enough time to be discussed in depth.
In general the inputs were related to tools used by attendees and references to:

1. The domain-agnostic guidelines by SSSOM.
2. The fact that developing mappings is a time consuming and expensive process.
3. The fact that there is a need for a methodology to handle mapping evolution while mapped resources change.
4. Open discussion about the representation of complex mappings as simple ones.
5. Comments about considering the hierarchical context of concepts in controlled vocabularies and keywords in definitions as a starting point.

For the next workshop we propose to address the same questions again as well as the following new ones:

- How could the methods be automated?
- Do you use any external platforms like LOV, Bioportal, etc?
- What do you use mappings for?
- How do you define requirements?

Funded by the European Union

6. Conclusions and next steps

This workshop drew a lot of attention and revealed the urge for solutions to the problem of mappings from the different communities involved in the workshop and also beyond. It enabled to engage with the interested parties and to kickstart discussion on various aspects supporting the creation of a coherent framework for FAIR mappings. The strong commitment of the participants to share their use-cases and to contribute to the work initiated by the project will allow us to move forward on these matters with large community support. During the workshop, we collected inputs from the participants on four main topics: how to make mappings FAIR? What are the different types of mappings that should be considered? What are the existing tools and requirements for tools? What are the existing generic and community specific methods for doing mappings? Based on these inputs, we will further develop a set of recommendations and technical requirements to define how to make FAIR mappings. The first version of this document will be shared, presented and open for comments for the wide community. In addition, we will work on establishing a list of existing methodologies presented as a FAIR mapping cookbook to support communities which are not yet using mappings in order to establish a common set of guidelines and good practices, therefore avoiding too much divergence of practices. Finally, we will continue interaction with the communities to collect use-cases and identify the various types of mappings. These activities will help us identify the necessary extension of existing models such as SSSOM to cover the diversity of mappings in order to create a common exchange model for mappings.

Funded by
the European Union

7. Appendices

Inputs from Miro board collected during discussion of the workshop.

Topic 1: Requirements for FAIR mapping

General inputs

- Give persistent identifiers to everything
- Agree on the granularity of persistent identifiers – assign to alignment, individual mappings plus versions.
- Agreement on attributes of mappings PID record to decide if mappings (or alignments) are useful before downloading
- Initiate the process to set up a federation of registries carried by communities (life science, SSH, ESS)
- Where can we find mappings or data models?
- Definition of source and target data structure
- Solutions: identifiers for data types, data type identification tools
- Make governance of the mapping clear
- Licence the mappings to make clear how they can be used
- Who is going to operate and sustain the mapping registry to live beyond the life of funded projects
- Challenge: how to register metadata for mappings? Nano publications

Inputs to Findable principle

- Aggregation for mappings themselves?
- For humans: clear user interface
- Mapping registry (comment: Probably multiple registries?)
- And allow versioning and updating of md schemas and crosswalks over time

Inputs to Accessible principle

- Need to be hosted/served in mapping repo
- Open up whats already there (OAEI mappings 2SSSOM)

Inputs to Interoperable principle

- Standard vocabularies for mapping metadata
- Lookup services
- APIs

Inputs to Reusable principle

- Having standards to express how mappings were created (justification, in which context were created)
- Clear definition of the scope of the mapping
- Know the mapping creator´s intentions
- (open) licence in the metadata

Funded by the European Union

Topic 2: Different types of mapping

- Archiving – complex/proprietary file format/data structure without qualified data values to a well-known data structure with qualified data values
- Indexing – mapping subset of data to a model used for discovery/search
- Computer assisted – Identify and assign a score for potential matches in a target data structure for manual curation
- Inference – Agument/enrich a data structure by mapping implicit relationships defined in ontologies and other knowledge organisation models to explicit one
- Contextual mappings – If property X=Y then do Z else do something else
- Describing & Mapping skills that permit to map concepts
- Have a semantic artefact (e.g thesaurus) as a hub of concepts that links other semantic artefacts, possibly more specific, concurrent,etc. --> rather loose mappings
- Are predicates expressing domain knowledge still considered as mappings ? e.g. A_concept taxon_name_for B_concept
- Mapping against closed vocabularies: E.g. in B2FIND we parse all subjects/keywords and when the value matches a string in the FOS vocabulary, it's mapped on field 'discipline'.
- huge sets of "low" value automatically computed mappings VS small sets of high expertise mappings: how to discriminate from each other?
- Some example transformations sound so complicated that workflow preservation is probably more appropriate than FAIR mappings (?)
- Some complex mappings: string to GEO coordinates, string to ODRL file with more information about licence, string to more information about project and funder, automatic usage of mappings from one vocabulary to another, string to citation in BIBTex, string to restrictions or term of access of usage of digital object, string to temporal coverage represented with vocabulary...

Topic 3: Mapping tools and needs for tools

Needs

- @Carol, CSV; that‘s why we need to support legacy formats- but to create LOD mappings/crosswalks we need to do a format upgrade in order to give those CSV-schemas PIDdeexistencece as well as the capacity to host LOD enabled crosswalks between a CSV and something else
- Better support for SSSOM JSON (added link: https://github.com/mapping-commons/sssom/issues/241)
- Is there an existing namespace we could use for SSSOM mapping commons PIDs?
- How to get an SSSOM TSV out of an OWL/TTL or reverse merge a TSV into an ontology?
- Which applications exist for end-users to create and manage mappings?
- Handling legacy SKOS mappings
- JSKOS mapping data format
- Need for user-friendly mapping validation tool to work with thematic experts (no skills expected)
- Mappings repo inside OntoPortal

Funded by the European Union

- Legacy mappings
- Extract Transform Load libraries to extract data from common domain specific data file formats to more widely-used data structures with qualified data types
- APIs for accessing mappings according to a FAIR data model
- This (points to 13) allows reducing costs in building UI tools: if the APIs return exactly data model conformant instances.
- (semi-)automated workflow to create mappings from spreadsheets (and possibly continuously update them, in a CI/CD-like way)
- Avoid excessive centralisation of infrastructure - communities of practice beat central registries. Registry systems should be easily deployable by different groups and be able to talk to each other: https://mapping-commons.github.io/sssom/MappingRegistry/

Tools

- Cocoda https://coli-conc.gbv.de/cocoda/
- OxO https://www.ebi.ac.uk/spot/oxo/
- Alignment Maker Light (AML) https://github.com/AgreementMakerLight (FAIR matching tools are tools that export the reasons for establishing a match.)
- Mapping commons - https://mapping-commons.github.io/
- Nanopublications - https://nanopub.net/ (nanopub templates based on sssom would be nice in the meantime)
- Ontology Access Kit - http://berkeleybop.github.io/software/OAK/
- Mapping.bio (http://biomapping.net/)
- OnaGUI //github.com/lmazuel/onagui
- Protégé (https://protege.stanford.edu/software.php#desktop-protege)
- Google Sheets or any other spreadsheet software (comment: Hold on ... spreadsheets are good to work... but not to release data.. ;eventually repository for mappings and versioning are necessary)
- OntoPortal mapping repo and mapping tools (LOOM (lexical mappings) and SAME_URI, SAME_CUI (identifiers-based mappings)

Topic 4: Mapping methodologies

- Q: How do you go about making mappings?
  - A: need a methodology to handle mapping evolution while mapped resources change
- Q: Do you have any community specific guidelines?
  - A: some generic (domain agnostic) guidelines by SSSOM : https://mapping-commons.github.io/sssom/tutorial/
- Q: How do you validate your mappings?
  - A: we make thematic experts work on them (but it is difficult - timely)
  - A: sssom has a natural validation system
- PROVOCATIVE STATEMENT: a complex mapping can ALWAYS be represented as a set of simple mappings.=> There are only simple mappings
  - why not, but you may need a way to formalise the relationship between the simple mappings. it is not always A+B

Funded by the European Union

- i agree. Likely any complex mapping can be 'encoded' as a crosswalk comprising n 'simple' mapping steps, but when n>5 it gets maybe not manageable any more.

Funded by the European Union