



Metadata, semantics and interoperability Preliminary information gathered by workshop participants

Thursday 24th November, 14.00-15.30 CET

Theme/Workshop Chair: **Clément Jonquet**, INRAE

Rapporteur: Oscar Corcho, UPM

Event link: <https://fair-impact.eu/events/synchronisation-force-events/synchronisation-force-1st-workshop-november-2022>

Context: FAIR-IMPACT's WP4 focuses on several aspects of the lifecycle of FAIR semantics artefacts – a broader term to include ontologies, terminologies, taxonomies, thesauri, vocabularies, metadata schemas and standards – such as governance, creation, sharing, reuse, FAIRness assessment. FAIR-IMPACT's T6.1 focuses on semantic interoperability for EOSC and how semantic artefacts help to support it. In the context of the 1st FAIR-IMPACT Synchronisation Force workshop we launch this session to gather different communities to make an inventory of the state of use of semantics artefacts in their scientific domain. We will survey multiple scientific communities directly involved within FAIR-IMPACT and beyond: agri-food, biodiversity/ecology, biomedicine, photons&neutrons, astronomy, materials, industry, earth sciences, social sciences & humanities. Which metadata standards and ontologies are used? Is there a reference semantic artefact catalogue in this area? Is there a governance for semantic artefacts? etc. The session will feature short summary presentations of experts in each area and discussion for path of improvement, cross-community fertilisation.

Recommendations assessed during the session:

- **Recommendation 1:** Develop domain and cross-domain interoperability frameworks at the level of vocabularies, ontologies, and metadata schema
- **Recommendation 2:** Further develop and implement semantic technologies, particularly in domains where their use is less advanced

Relevant EOSC-A Task Forces

- [EOSC Semantic Interoperability Task Force](#). To a secondary degree, the [EOSC FAIR Metrics and Data Quality Task Force](#).

Useful references

- [Recommendation for a FAIR EOSC: White paper, Rec. n1 and n 2](#)
- [Turning FAIR into Reality report Rec. n 4](#)

Preliminary questions

1. Which are the semantic artefacts that you develop or use in your scientific domain and how are they governed (developed and maintained)? Please provide any relevant links. 2
2. What limitations do you see in your scientific domain with respect to the use of semantic artefacts to describe and/or search/find research datasets? 7
3. Is there any type of semantic artefact catalogue in your scientific domain? 10
4. Is there a strategy to deal with crosswalk and mappings between semantic artefacts in your scientific domain? 12
5. People who contributed to the preliminary information gathered 15

1. Which are the semantic artefacts that you develop or use in your scientific domain and how are they governed (developed and maintained)? Please provide any relevant links.

Project or initiative	Input
INRAE	I coordinate the INRAE Thesaurus (https://consultation.voculaires-ouverts.inrae.fr/thesaurus-inrae/fr/) which is the institutional resource that gathers and offers concepts of interest for use. It covers agriculture, food and environment thematics as well as concepts to describe research works. It is maintained by an editorial committee of 15 persons in collaboration with research teams and individual scientists who bring their thematic expertise. Users can ask for modifications and enrichment. Though for internal use in INRAE's information systems, the thesaurus is publically exposed and can be used by anyone through its discovery portal and its API.
Life in Kyrgyzstan Study	We do not develop but use what already exists. For example, the sectors of economy, classification of occupations, etc.
CESSDA	CESSDA vocabularies and ELSST thesaurus, developed and maintained by CESSDA ERIC https://www.cessda.eu/ DDI Alliance standards and vocabularies https://ddialliance.org/products/overview-of-current-products . DDI Alliance is a member-based organisation, development is done mostly by in-kind contributions.
BlueCloud	NERC Vocabulary Server (NVS) (https://vocab.nerc.ac.uk/) Hierarchical Linked Data vocabularies related to the marine and wider domain.
Open Biological and Biomedical Ontologies Foundry	https://obofoundry.org/ Personally, I am involved in many anatomy ontologies (Uberon, ZFA, XAO, etc), phenotype ontologies (HPO), disease (MONDO).
EOSC-A Long Term Data Preservation task force	NA
AgroPortal	Within AgroPortal we have semantic artefacts form different sources, e.g. : - Ontologies developed in with the OBO guidelines e.g., PO, TO, AGRO, FOODON - Ontologies / trait dictionaries developed in group project such as Crop Ontology project - Reference thesaurus developed by organizations: AGROVOC, INRAE Thesaurus, ANAEE Thesaurus - Application ontologies e.g., OFPE, AFEO, PO2



Project or initiative	Input
	They are developed by different groups of scientists with specific guidelines and/or governance under the umbrella of a specific action or project often driven by concrete application tasks sometimes without any coordination or no global governance. Hence the goal of WP4.
LifeWatch Italy, the National Hub for biodiversity and ecosystem data and research products	<ul style="list-style-type: none"> • Thesauri on functional traits of aquatic organisms and on alien species developed for LifeWatch Italy using the editing tool VocBench and published through the repository of semantic resources for the ecological domain EcoPortal. Lifewatch Italy Thesauri are used to label data and in metadata. • EnvThes, BODC vocabularies, DwC controlled vocabularies and others available through Ecoportal or external semantic services used to label data and in metadata. • CIDOC CRM, CRM digital, CRM science, Parthenos and LUPO ontologies used in the semantic model that describes different resources and their metadata in the LifeWatch Italy Semantic Platform. Semantic artefacts developed in-house (e.g. LW Italy thesauri) follow the EcoPortal governance. All others follow guidelines and/or governance of a specific organization or initiative.
TRIPLE	The project TRIPLE developed its own multilingual thesaurus for SSH contents (data and publications), based on LCSH thesaurus (multidisciplinary library thesaurus). Set up by the consortium with manual review. in the SSH, semantic artefacts are often related to disciplines or domains (eg, linguistics, social sciences, archeology). Related to disciplines which traditionally use more and bigger datasets. Other SA are not scientific specific: coming from libraries, publishers, or about generic information (eg languages).
The project of French national catalogue of individual health data collections (FRéSH, for France Recherche en Santé Humaine) is currently in preparation phase	The candidate data sources for the catalogue cover different domains (Epidemiology, Clinical medicine and SSH) and use their proper terminologies: MeSH (Medical Subject Headings) for ClinicalTrials.gov, CESSDA (Consortium of European Social Science Data Archives) for PROGEDO, ICD-10 (International Classification of Diseases, 10th revision) and ad hoc vocabulary for the Portail Epidémiologie France. The challenge is to identify one or more bilingual (French-English) terminologies that cover the entire thematic scope of the portal in a satisfactory and balanced way. We consider using several existing bilingual controlled vocabularies for the topic fields. The bilingual MeSH version is produced by the Inserm's team that translates it into French. The bilingual ICD-10 is produced and translated by WHO.



Project or initiative	Input
	Multilingual CESSDA vocabularies have been created and maintained by the DDI Alliance.
LifeWatch ERIC	On the ERIC level, only an upper level ontology has been created so far (domain-specific semantic artefacts are created at the national node level). The upper level ontology (LUPO - http://ecoportal.lifewatch.eu/ontologies/LUPO) was created through extended meetings that included all LifeWatch stakeholders, in which classes and relationships were discussed and their inclusion was decided with general consensus.
ENVRI-FAIR	<p>In ENVRI-FAIR in the Solid Earth subdomain the catalog uses CERIF. It has a syntactic (structure) layer and a semantic layer. Both have base entities and link entities. Base entities in the semantic layer are terms; link entities are relationships between terms (i.e. ontological structure). The subdomains within earth science maintain vocabularies and relationships through various tools and various governance structures.</p> <p>In the marine subdomain the vocabularies incorporated in the metadata are WoRMS taxonomy, MarineRegions ontology, NERC Vocabulary Server (NVS), Aquatic Sciences and Fisheries Thesaurus (ASFA). SeaDataNet and Argo vocabularies are operated, managed and governed by BODC, and a lot of vocabularies are community led and requested, discussed and handled through github: https://github.com/nvs-vocabs</p> <p>Marine Regions and Marine Species is managed and operated by VLIZ</p> <p>In the Ecosystem subdomain, the DCAT-AP is being used to describe assets listed in catalogues, including datasets and research sites. In the soil water content use case we are further creating a semantic model out of the dataset metadata schema and the data series structure, which will eventually reuse as much from existing semantic artefacts as possible, including for example DCAT, the Countries ontology by the Object Management Group, the AnaEE thesaurus, I-ADOPT, etc.</p> <p>For the ENVRI Catalogue of Services (in Task Force 1), we proposed the use of EuroSciVoc and the NASA GCMD vocabularies to specify the domain of interest and domain-specific keywords for the described services. Every subdomain may also use keywords from their own, community-approved controlled vocabularies. We are</p>

Project or initiative	Input
	also developing our own, ENVRI-specific Service Type vocabulary, because we could not find one that satisfied our needs for the Catalogue.
ELIXIR-NO, EuroScienceGateway, FAIRtracks	<p>In the FAIRtracks project, we are developing a minimal metadata exchange standard for genomic track files , which are condensed data files which are routinely generated as part of genomic datasets. The standard itself is a semantic artefacts, registered mainly in FAIRsharing (https://fairsharing.org/).</p> <p>We only want to support a single ontology term for each concept, to which end we have tried to select non-overlapping high-quality branches from different ontologies for different metadata fields, and have stringent validation for conformance. See https://fairtracks.net/fair/#fair-04-ontologies . Selecting the best ontologies are still difficult, but easier when you only need to consider a particular sub-branch at a time and can mix-and match the strength of ontologies</p>
BlueCloud	<p>We maintain and support the evolution of the ISO 21127:2014 CIDOC CRM, and its family of extensions. Amongst them the MarineTLO ontology, that is a top level ontology for the marine domain. Moreover, we have developed a schema mappings definition language and the corresponding services and tools to support data transformation from one schema to another.</p>
EERAdata project	<p>We try to identify semantic artefact for the energy domain with a specific focus on the low-carbon transition. If no standards are found, we create missing artefacts. Typical standards are set in the domain by statistical offices, the European Commission, the International Energy agency and other. There is no overall platform which serves as a portal to all these standards, see European Environmental Agency Vocabularies here</p>
ELIXIR	<p>FAIRsharing.org lists our semantic artifacts (https://fairsharing.org/) - the ontologies, reporting guidelines, formats. OLS (https://www.ebi.ac.uk/ols/index) and the various portals of the Ontoportals Alliance (https://ontportal.org/) are used for the ontologies. Bioschemas (https://bioschemas.org) is our convention of using schema.org for a range of resources. OBO foundry is part of this community.</p>



Project or initiative	Input
IOC-UNESCO Ocean InfoHub / the Environment Ontology / Helmholtz Ontology Base	For knowledge graph construction, we leverage schema.org semantics and augment them where necessary by linking to ontologies such as ENVO (and other OBO resources), thesauri such as the NERC Vocabularies, and others. There are varying governance schemes for these, but most support community input, which is then subject to editorial control before merging
ExPaNDS	Photon and Neutron PaNET ontology, DCAT vocabulary
FAIRsharing	<p>* Subject Ontology: an open ontology built from 7 community ontologies, describes subject areas / academic disciplines across all research domains with PIDs, definitions and synonyms for all terms. Available for anyone to use. Used within FAIRsharing records by curators and the user community (see also the SRAO FAIRsharing record for details of the ontologies it incorporates and other information). Developed within FAIRsharing but happy to have external collaborators.</p> <p>* Domain Ontology: this is an open FAIRsharing application ontology that integrates >50 community ontologies, many of them OBO Foundry ontologies. While not as directly relevant to the wider community as our Subject Ontology is, it showcases a practical implementation of actually using other ontologies rather than developing your own.</p> <p>* terms4FAIRskills: terminology for the skills necessary to make data FAIR and to keep it FAIR. Now approved and incorporated within the OBO Foundry, and used by a variety of projects such as the FAIR Cookbook, FAIRsharing and those listed at its homepage. Github, OBO Foundry entry. Can be used to annotate training materials, job advertisements, CVs, and more. Use cases include: To assist with the creation and assessment of stewardship curricula; To facilitate the annotation, discovery and evaluation of FAIR-enabling materials (e.g. training) and resources; To enable the formalisation of job descriptions and CVs with recognised, structured competencies. Developed by a collaboration with a number of resources as outlined in the homepage.</p> <p>* Software Ontology: Also an OBO Foundry ontology, describing software tools, their types, tasks, versions, licensing, provenance and associated data. (Github)</p>
Bioschemas	extensions to schema.org targeting life sciences primarily, though many analogous efforts cross domain (see RDA grp Research Schemas)
ISIDORe / By-COVID	sensitive data vocabularies, dealing with countries / regions languages and laws, rules



Project or initiative	Input
EOSC4Cancer	Not yet defined. But we will align with standard ontologies in the field. DUO codes for data access, etc.

2. What limitations do you see in your scientific domain with respect to the use of semantic artefacts to describe and/or search/find research datasets?

Project or initiative	Input
INRAE	<p>The domains covered by INRAE are numerous, from fundamental sciences like physics to applied domains like agriculture. Some domains are well covered while others are not covered at all with publically available, trustworthy and technically fitted semantic resources.</p> <p>Lack of methodological guidelines and user-friendly tools to reuse, edit, document and publish semantic artefacts</p>
Life in Kyrgyzstan Study	Local classifications can be difficult to generalize to international standards, maybe. Educational systems, for example, can be different.
CESSDA	<p>Lack of strategies e.g. for managing vocabularies and vocabulary services, and lack of best practices for versioning of vocabularies</p> <p>Lack of tooling that supports using vocabularies in an optimal way in documentation.</p> <p>Challenges related to multilinguality</p> <p>Lack of SKOS/RDF expertise to produce interoperable and FAIR voc exports</p>
BlueCloud	A standardised quality framework is missing to be able to map between different quality schemes
Open Biological and Biomedical Ontologies Foundry	
EOSC-A Long Term Data Preservation task force	NA, but just to note that also semantic artifacts such as the terminologies, ontologies etc. should be preserved to guarantee FAIRness over Time of the digital objects.

Project or initiative	Input
AgroPortal	<ul style="list-style-type: none"> - Always changing contexts - Long term availability, strategy and governance needed - Problem of selecting Mr right semantic Artefact and Ms right ontologie - Then select the appropriate term/concept - Need for a mapping repository <p>(beyond what's available for now in AgroPortal)</p> <ul style="list-style-type: none"> - Long term support of infrastructure to access and use SA - Interoperation with data repositories is still hard
LifeWatch Italy, the National Hub for biodiversity and ecosystem data and research products	<p>Difficulty to select the right term/concept due to a large collection of independent semantic artefacts in the same domain.</p> <ul style="list-style-type: none"> • Too many semantic artefacts with the same terms/concepts without mapping among them. • Lack of semantic annotation tools/services and their integration on the data management systems. • Agreed strategy and governance needed.
TRIPLE	<p>In the SSH, the disciplinary boundaries make it difficult to have a unified environment: not only the perimeter but the granularity and the terminology changes. Having all existing SA connected is challenging.</p> <p>Also, a complete list of available resources in the SSH still has to be made.</p> <p>Another challenge comes with multilingualism, which is in the SSH adds another level of complexity of the mappings (technical terms of IT are more "universal").</p> <p>Generally, SSH is an abstraction not clearer than STM.</p>
The project of French national catalogue of individual health data collections (FReSH, for France Recherche en Santé Humaine) is currently in preparation phase	<p>Some fields have more terminology resources than others. In particular, the clinical disciplines are very well supplied with semantic artefacts, as they need them for the description of medical acts in hospital IS, whereas I have not found a satisfactory level of granularity for epidemiology/public health field.</p>



Project or initiative	Input
LifeWatch ERIC	Coordination has proven difficult, because an ERIC is large and discussions/decisions should involve all interested members. Agreement on the semantic artefact to use (or create/extend) is very time-consuming and not easy to achieve, as it requires knowledge of each member's individual needs and dataset peculiarities, as well as very good coordination, problem-solving and feedback, just like in an EU-level project.
ENVRI-FAIR	<ul style="list-style-type: none"> · Only the ability of the subdomains to maintain the terms current with scientific practical use of the terms. This work is continuous and ongoing. · Finding ways to encourage those creating metadata records to use these (or other) vocabularies when they provide us with their metadata and within their datasets · In TF1, we could not find a suitably descriptive vocabulary for service types. · Finding terms/concepts/classes/etc. inside vocabularies is not always straightforward. · Using existing vocabularies and their terms often requires compromises that subdomains are not ready to make. · Communication gap between the people who pick the terms (data providers) and people who implement IT solutions using these terms.
ELIXIR-NO, EuroScienceGateway, FAIRtracks	Different repositories use different semantic artefacts, at different levels of semantics, conformity, etc. Mapping/harmonization across repositories, data models, remains a huge challenge. Particularly the management of ontology versioning have been difficult in the context of metadata schemas, specifically how to map metadata schema versions with ontology versions (see https://fairtracks.net/fair/#fair-04-ontologies for a discussion).
BlueCloud	The mapping definition process, can be a quite challenging and time consuming process that involves persons with different knowledge and expertise
EERAdata project	Low technical skills, semantic techniques are rarely used, no institutions coordinating the efforts
ELIXIR	a large number of established ontologies from different communities (bioscience is a big space) so we do a lot of cross-walking. this is expected, Sustainability of the ontology services and registries. A great deal of work is spent on curation pipelines using the semantic artifacts. Tools like Zooma https://www.ebi.ac.uk/spot/zooma/

Project or initiative	Input
IOC-UNESCO Ocean InfoHub / the Environment Ontology / Helmholtz Ontology Base	Poor awareness of how to evaluate the quality and fitness-for-purpose of a semantic resource. Poor understanding of what is an ontology vs a vocabulary vs a thesaurus etc. Poor and ad hoc mappings, often changing the semantics without qualification. Poor technical workflows for operational maintenance and release, poor quality control (e.g. with reasoners). Weak definitions that are "pragmatic" but generally damaging.
FAIRsharing	People aren't reusing ontologies or ensuring that they are orthogonal, and are much more likely to create their own ontologies (often to the detriment of its quality)
ISIDORe / By-COVID	DURC data (confidentials and/or protected) and difficulty to train experts on these aspects, other limitations due to national priorities on these topics

3. Is there any type of semantic artefact catalogue in your scientific domain?

Project or initiative	Input
INRAE	Yes, Agroportal, Bioportal, OBO Foundry -- we are quite lucky ;) Many semantic artefacts produced at INRAE are also recorded in Research Data Gouv (data catalog/repository for French research data)
Life in Kyrgyzstan Study	No.
CESSDA	CESSDA CVS https://vocabularies.cessda.eu/ DDI Alliance standards and vocabularies https://ddialliance.org/products/overview-of-current-products SSH Vocabulary Commons https://vocabs.sshopencloud.eu/vocabularies/sshocterm/en/
BlueCloud	<u>NVS is a semantic artefact catalogue:</u> https://vocab.nerc.ac.uk/ <u>It currently hosts 290 vocaularies related to the marine and wider domain.</u>
Open Biological and Biomedical Ontologies Foundry	https://obofoundry.org/
AgroPortal	With resources in the agri-food domain, we can cite: <ul style="list-style-type: none"> - SA libraries: OBO Foundry, FAIRsharing (all disciplines) - SA repositories: EBI OLS, NCBO BioPortal, etc.. - specific web applications such as the Map of agri-food data standards, or croponontology.org - AgroPortal AgroPortal develops a focus on agri-food with and open

Project or initiative	Input
	and use case driven approach Also include a Recommender, an Annotator, a FAIRness assessment tool, a (simple) mapping repository Standardized approach with the OntoPortal Alliance effort. Towards an EOSC ready component?
LifeWatch Italy, the National Hub for biodiversity and ecosystem data and research products	EcoPortal, BioPortal, AgroPortal, OBO Foundry, Gfbio Terminology Service, Research Vocabularies Australia, NERC Vocabulary Server, FAIRsharing.
TRIPLE	For the SSH, the organizers know them better than me. Like said above: a complete list of available resources in the SSH still has to be made. Another question would be, once identified these SA in SSH: who is using them and to which extent?
The project of French national catalogue of individual health data collections (FReSH, for France Recherche en Santé Humaine) is currently in preparation phase	Yes, Bioportal (except for CESSDA vocabularies) and SIFR BioPortal for the bilingual versions.
LifeWatch ERIC	We maintain EcoPortal, a repository for semantic artefacts pertaining to Ecology, Biodiversity and related disciplines. LUPO is an upper ontology for our organisational resources in general, but we still published it in EcoPortal because it will eventually link everything LifeWatch, including domain-relevant semantic artefacts, forming an important component in the internal interoperability framework.
ENVRI-FAIR	See (1). CERIF provides the necessary capabilities. See list in question 1 for the marine subdomain, also used are ENVO · For the ecological domain the EcoPortal (http://ecoportal.lifewatch.eu/) · AgroPortal (http://agroportal.lirmm.fr/) for the AnaEE thesaurus
ELIXIR-NO, EuroScienceGateway, FAIRtracks	We also use external ontologies from EMBL-EBI Ontology Lookup Service (https://www.ebi.ac.uk/ols/index) and NCBO Bioportal (https://bioportal.bioontology.org/) extensively, an also plan to extend existing ontologies with terms that we need. The FAIRtracks project in itself is an approach at harmonising metadata that in many cases are following



Project or initiative	Input
	existing standards. As such we are developing generic tools to simplify and automate metadata transformations
BlueCloud	We maintain several instations of the ISO 21127:2014 CIDOC CRM ontology and its extensions, as well as the entire knowledge base of the Global Record of Stock and Fisheries, that was built on top of MarineTLO
EERAdata project	<u>European Environmental Agency, Open Energy Ontology</u>
ELIXIR	FAIRsharing.org lists our semantic artifacts (https://fairsharing.org/) - the ontologies, reporting guidelines, formats. OLS (https://www.ebi.ac.uk/ols/index) and the various portals of the Ontoportals Alliance (https://ontportal.org/) are used for the ontologies. Also see OBO Foundry.
FAIRsharing	<u>FAIRsharing stores descriptions of >800 terminologies (all disciplines) as just one of the standards types listed within our registry; please note these store descriptions of terminologies, but not the terminologies themselves. For the life science only, the OBO Foundry is vital to good ontology development and management within the research space (https://obofoundry.org/). We also regularly use the EBI's Ontology Lookup Service (OLS).</u>
ISIDORe / By-COVID	ongoing work within the RDA

4. Is there a strategy to deal with crosswalk and mappings between semantic artefacts in your scientific domain?

Project or initiative	Input
INRAE	Yes, absolutely. Crosswalks and mappings are (almost) first class citizens in the Ontoportals based SA repositories. We are very interested in SSSOM which allows to provide rich metadata for mappings, facilitating their reuse. (mappings included in SA as it is often the case actually are hardly reusable as they lack documentation)
CESSDA	Not that I know of. SSH Vocabulary Commons is a 'collaboration of the willing' that brings together experts and managers from SSH research infrastructures CESSDA, CLARIN, DARIAH and E-RIHS that have agreed to share their expertise and work towards common recommendations for firstly, using and managing vocabularies used in SSH and secondly operating, sharing and managing of vocabulary services useful for the broad SSH community. https://docs.google.com/document/d/18v0J1Xbo25YuGLyC37aGLc7O6MIZaxH9iCV1mrwM9JA/edit?usp=sharing



Project or initiative	Input
BlueCloud	<p>NVS hosts mappings between NVS concepts but also between NVS and externally hosted concepts. It provides a unique URI to each mapping and also maintains and publishes the provenance of the mapping.</p> <p>A possible strategy to crosswalk on observable properties is the usage of the I-ADOPT framework across domains. That would seriously help on the harmonisation of the observable properties across domains.</p>
Open Biological and Biomedical Ontologies Foundry	<p>We have a community standard for representing mappings https://w3id.org/sssom, with a community for debating new elements (https://github.com/mapping-commons/sssom/issues).</p> <p>And a bunch of tools around it. But the situation is far from solved. The word "crosswalk" alone causes many raised eyebrows, when prefer "mapping chaining".</p>
AgroPortal	<ul style="list-style-type: none"> - Lack of coordination & governance - Need for long term strategies and support. If SAs are really important research outputs, let's get organized for their support on the long term - Still a variety of metadata approach and models (not restricted to agri-food) - Taking into consideration the whole life cycle of SA (from designing to sharing)
LifeWatch Italy, the National Hub for biodiversity and ecosystem data and research products	<p>There is not a strategy for crosswalk and mappings in the scientific domain. We need to analysis what exists and adopt a common solution that could be integrated in our catalogues and tools.</p> <p>At the moment, EcoPortal offers a basic mapping tool, but it should be improved in terms of algorithms, validation and other technical issues . Vocbench tool also allows the management of alignment from two different source: 1. File: loads a document expressed according to model of INRIA's Alignment API; 2. Remote Alignment System: allows to run an alignment task exploiting an external alignment system. This task, when completed, returns a list of alignments found and the project manager can validate or reject the proposed alignments one by one, or apply some quick actions (for more specifications see http://vocbench.uniroma2.it/doc/user/alignment_validation.jsf).</p>
TRIPLE	<p>That would probably be a project in itself.</p> <p>In TRIPLE, we took a generic SA to cover all the SSH fields (LCSH thesaurus), but lack of granularity and domain specificity. so, issues with terms too broad or common.</p> <p>In other projects, the objective was to try to integrate many voc together, but issues with granularity and specific terms.</p>



Project or initiative	Input
	My guess would be to start in the middle: identify most used and general voc, find mappings and/or terms adaptations that could work for as many fields as possible.
LifeWatch ERIC	No. We expect to uptake guidelines and suggestions from FAIR IMPACT.
ENVRI-FAIR	<p>For the Solid Earth subdomain: CERIF provides (through linking entities in the semantic layer) crosswalks between terms (and between vocabularies) including the capability of multilinguality. In the Marine subdomain it is done on a case-by-case and often manual basis at present, we are however starting to develop such a semantic service in-house.</p> <p>NVS maintains a manual mapping service where each mapping has a URL that lists its provenance since mappings are submitted by different submitters.</p> <p>In the soil water content use case we tend to either use the relevant classes/concepts directly or to create sub-classes/super-classes, on a case-by-case basis. Eventually, a mapping using skos properties will also be created where required, to link vocabulary terms in thesauri of different RIs. The links will be decided by the RI representatives and will be inserted directly in the file of the semantic model. I-ADOPT: Magagna, Barbara, Moncoiffé, Gwenaëlle, Devaraju, Anusuriya, Stoica, Maria, Schindler, Sirko, Pamment, Alison, & RDA I-ADOPT WG. (2022). Interoperable Descriptions of Observable Property Terminologies (I-ADOPT) WG Outputs and Recommendations (1.1.0). https://doi.org/10.15497/RDA00071. The I-ADOPT Interoperability Framework is a set of guidelines based on a simple non-domain specific ontology that supports the decomposition of complex observable properties into their essential atomic parts represented through the concepts in FAIR terminologies. The I-ADOPT Interoperability Framework is a substantial contribution to enabling the interoperability of terminologies, within and across domains without sacrificing any prior efforts. It serves as a common layer of abstraction through which concepts from different terminologies can be systematically aligned and extended, as needed. The annotation of observational data with these interoperable concepts supports the findability and reusability of datasets across repositories facilitating the federation of data systems. It represents the variables' context and provides the level of detail needed to identify suitable datasets and to connect different datasets with related observations. The I-ADOPT Interoperability Framework will facilitate the development of tools to enable researchers to describe and find their data more systematically.</p>
ELIXIR-NO, EuroScience	We are developing a generic Python library for structured and scalable transformation of (meta)data schemas

Project or initiative	Input
Gateway, FAIRtracks	(https://fairtracks.net/fair/#fair-07-transformation). We are interested in building on and integrating mapping solutions developed elsewhere. We have also used Ontology Xref Service (https://www.ebi.ac.uk/spot/oxo/) and Zooma (https://www.ebi.ac.uk/spot/zooma/) for mapping, both at EMBL-EBI.
BlueCloud	This is amongst our core activities. We have developed the X3ML schema mappings definition language that supports the collaborative definition of schema mappings in a declarative manner, as well as the corresponding tools (i.e. X3ML engine, schema analyzers, visualizers, etc.) to support the data transformation process.
EERAdata project	EERAdata aims to advance this, there seems to be something upcoming in the context of energy data-spaces
ELIXIR	we do this on a project basis - BY-COVID, EOSC4Cancer etc are each doing cross-walks driven by concrete examples.
FAIRsharing	The OBO Foundry has a defined procedure for importing other foundry ontologies that works very neatly, for all disciplines. This is better than having a separate mapping file. FAIRsharing can provide "live" versions of crosswalks via our collections, such as this one.
Bioschemas	cross walks in RDA work, and some (that need improving) in bioschemas
ISIDORE / By-COVID	through RDA groups and recommendations, integrating international groups around Life Science

5. People who contributed to the preliminary information gathered

Project or initiative	Name	Surname	Affiliation	Country
INRAE	Sophie	Aubin	INRAE	France
Life in Kyrgyzstan Study	Damir	Esenaliev	Leibniz Institute of Vegetable and Ornamental Crops (IGZ)	Germany
CESSDA	Mari	Kleemola	Tampere University	Finland
BlueCloud	Alexandra	Kokkinaki	NOC-BODC, Blue Cloud	UK
Open Biological and Biomedical Ontologies Foundry	Nicolas	Matentzoglou	Semanticly	Greece

Project or initiative	Name	Surname	Affiliation	Country
EOSC-A Long Term Data Preservation task force	Roxanne	Wyns	KU Leuven	Belgium
AgroPortal	Clement	Jonquet	INRAE	France
LifeWatch Italy, the National Hub for biodiversity and ecosystem data and research products	Ilaria	Rosati	National research Council	Italy
TRIPLE	Arnaud	Gingold	OPERAS-Aix Marseille University	France
The project of French national catalogue of individual health data collections (FReSH, for France Recherche en Santé Humaine) is currently in preparation phase	Baudoin	Lesya	Inserm	France
OpenAIRE	Leonidas	Pispiringas	OpenAIRE	Greece
LifeWatch ERIC	Xeni	Kechagioglu	LifeWatch ERIC	EU
ENVRI-FAIR	Katrin	Seemeyer	Forschungszentrum Juelich	Germany
ELIXIR-NO, EuroScienceGateway, FAIRtracks	Sveinung	Gundersen	University of Oslo	Norway
BlueCloud	Yannis	Marketakis	Foundation for Research and Technology - Hellas	Greece
EERAdata project	August	Wierling	Western Norway University of Applied Sciences	Norway
ELIXIR	Carole	Goble	The University of Manchester	UK
IOC-UNESCO Ocean InfoHub / the Environment Ontology / Helmholtz Ontology Base	Pier Luigi	Buttigieg	Helmholtz Metadata Collaboration / Alfred Wegener Institute	Germany
ExPaNDS	Alejandra	Gonzalez Beltran	STFC - UK Research and Innovation	UK



Project or initiative	Name	Surname	Affiliation	Country
FAIRsharing	Allyson	Lister	FAIRsharing (RDA FAIRsharing WG)	UK
Bioschemas	Nick	Juty	The University of Manchester	UK
	Jean-Christophe	Desconnets	french research institute for development	FR
ISIDORE / By-COVID	David	Romain	ERINHA	BE (FR)
EOSC4Cancer	Sergi	Aguiló	BSC	ES