

## Collaborative notes for Session 2 - “Metadata, semantics and interoperability”

---

**Chair:** Esteban Gonzales, UPM

**Rapporteur:** Anne-Sofie Fink, DeIC

**Synch Force Team:** Sara Plttonet (Trust-IT) Marjan Grootveld (DANS)

**Workshop date:** Tuesday, November 28, 10.30-12.00 CET

### Session overview

Interoperability is one of the FAIR Principles that recently has taken up more and more attention. Interoperability will often be the more challenging one of the four principles - indeed when we want to address interoperability across scientific domains. In this session we will explore the subject of semantic and technical interoperability within and across scientific domains. The aim of the workshop will be to explore the semantic and technical interoperability adopted by representatives from different research communities and institutions to explore their compliance with the EOSC Interoperability Framework's recommendations. We expect participation from different scientific domains in order to get a picture of their semantic and technical practices. Our point of departure for addressing semantic and technical interoperability is the [EOSC interoperability framework](#).

Questions explored during the session

- What does your project/institution/community do to implement semantic artefacts (ontologies, schemas, standards, vocabularies, etc) into your (meta)data? Please provide any relevant links (data policy, schema, repository, catalogues, etc.) and challenges to implement them.
- In your project/institution/community, do you publish your (meta)data following FAIR principles? Please provide any relevant links (data policy, guidelines, training courses, etc.) and challenges to implement them.
- Do you reuse data from other projects/institutions/communities? Please provide any relevant links of examples and the challenges met to reuse them (For example, provide sufficient information to reuse/ trustworthy). Specify if you plan to reuse data in the future.
- What are the worst data practices adopted in your project/institution/community? (For example, not using PID, clear data policy, etc.)

Useful references:

- [EOSC Interoperability Framework](#) for semantic and technical recommendations within EOSC environment
- [Turning FAIR into reality - Rec4](#) for developing interoperability frameworks for FAIR sharing within disciplines and for interdisciplinary research
- [FAIRsFAIR white paper - Rec1](#) for developing domain and cross-domain interoperability frameworks at the level of vocabularies, ontologies, and metadata schema

## Expected outcomes

The cross-workshop expected outcome is to assess the status of the implementation of the EOSC IF recommendations in the components and services present in cross-domain environments. These results will be used to create guidelines for the usage of components for semantic and technical interoperability in cross-domain use cases.

## Preliminary questions and responses received

Project or initiative	1. What does your project/institution/community do to implement semantic artefacts (ontologies, schemas, standard, vocabularies, etc) into your (meta)data?
<b>OpenAIRE</b>	<a href="https://guidelines.openaire.eu/en/latest/">https://guidelines.openaire.eu/en/latest/</a>
<b>RDA France</b>	
<b>GO FAIR Foundation</b>	We use a sheet2RDF pipeline and provide trainings for newcomers about how to create vocabularies and how to upload them in vocabulary repositories. We have developed the Semantic Interoperability Profile (SIP) to focus on FAIR Supporting Resources addressing semantic interoperability issues: <a href="https://sip-wizard.ds-wizard.org/">https://sip-wizard.ds-wizard.org/</a> We also use the I-ADOPT Framework for describing variables in train people on the use of it
<b>FAIR-IMPACT</b>	We develop ontology repositories (and semantic artefact catalogues) for multiple disciplines and communities. We standardize them, and eventually will federate them.
<b>FAIR-IMPACT</b>	We support scientists in developing semantic artefacts with a dedicated INRAE service (support action) called VO ouverts. We promote and share our semantic artefacts thru AgroPortal mainly.
<b>FAIR-EASE</b>	FAIR-EASE uses data from many different environmental disciplines. We are currently analysing these data to identify what kind of semantic artefacts they are using. Some semantic repos used are, vocabulary service ( <a href="https://vocab.nerc.ac.uk/">https://vocab.nerc.ac.uk/</a> ) for the marine community, gcmd keywords, OBO Foundry ontologies e.g. ENVO NCBI taxon, Chebi, EU eionet, INSPIRE, EFO, NCEI vocabs, ozcar-theia, aeris vocabs and a lot more.
<b>Blue-Cloud2026</b>	Blue-Cloud makes use of federated data from many different Blue Data Infrastructures. The Data Discovery and Access service drives on metadata, standards and semantics, and provides harmonised search capabilities to find and download datasets from the BDI's. The marine domain makes widely use of the NERC Vocabulary Service. In Blue-Cloud I the first set of BDI's have been made accessible (e.g. Seadatanet, Eurobis, EMODnet, Ecotaxa, Wekeo) improving FAIRness on the way, and in BC2026 other BDI's will be connected, and where needed will also improve their metadata models and semantics where needed.

Project or initiative	1. What does your project/institution/community do to implement semantic artefacts (ontologies, schemas , standard, vocabularies, etc) into your (meta)data?
<b>Semantics in Astronomy, Planetary Sciences and Heliophysics</b>	<p>The IVOA (ivoa.net) astronomy alliance implements vocabulaires and they are accessible (<a href="http://ivoa.net/rdf">http://ivoa.net/rdf</a>). The IVOA interfaces are using XML, with XML schemas, but they are not using RDF/XML. The IPDA (<a href="https://planetarydata.org">https://planetarydata.org</a>) is formalising its information model using OAIS and can be exported to OWL, although it is not used in this form. <a href="https://pds.nasa.gov/datastandards/about/">https://pds.nasa.gov/datastandards/about/</a></p> <p>The IHDEA (<a href="https://ihdea.net">https://ihdea.net</a>) is using an XML based registry (not RDF/XML): working repository <a href="https://github.com/hpde/">https://github.com/hpde/</a> and data model: <a href="https://spase-group.org/data/index.html">https://spase-group.org/data/index.html</a></p>
<b>Reactome database of biomolecular pathways</b>	<p>All our data uses external reference resources/ontologies. Most used examples are UniProt for proteins, ChEBI for chemical entities, Pubmed for literature. Main challenge is incompleteness of the external reference system. This is only relevant for ChEBI, which adds relevant chemicals on demand. UniProt, Pubmed, and most external ontologies are complete for our purpose. External ontologies may become unmaintained, as happend to us for an external ontology representing protein modifications. This requires costly adaptations.</p>
<b>A CRIS system for Open Science and FAIR publications</b>	
<b>University of Bologna (FAIR Champion + Data Stewards)</b>	<p>The institution, via the data stewardship service, aims to increase researchers' awareness on semantic artefacts via advocacy and informative moments. Each DS, expert in a specific knowledge domain, can advise researchers on the most appropriate semantic artefacts for their data.</p>
<b>PerSciDo: The French Labex Persyval-Lab</b>	<p>We have developed an ontology that integrates various specialized vocabularies into a single RDF database organized into namespaces identified by prefixes associated with URLs providing access to these namespaces. To achieve this, we followed three main principles: (1) adhere to existing or emerging standards, (2) anticipate usage and encourage best practices, and (3) evolve and enrich the platform with user input. The PerSciDo platform is built on a database based on semantic web tools in the form of an RDF Fuseki triple store. Therefore, the ontology and knowledge graph are intricately connected.</p>

Project or initiative	1. What does your project/institution/community do to implement semantic artefacts (ontologies, schemas , standard, vocabularies, etc) into your (meta)data?
<b>Integrated Carbon Observation System (ICOS)</b>	ICOS has developed our own ontology (see <a href="https://meta.icos-cp.eu/ontologies/cpmeta/">https://meta.icos-cp.eu/ontologies/cpmeta/</a> ) that covers variables, units, instrumentation, measurement stations, people/roles and organisational aspects. In those cases that other semantically consistent definitions exist, we attempt to provide qualified references to these. In the context of the ENVRI-FAIR project (and its soon-to-start follow-up ENVRI-Hub NEXT) we have worked on also implementing semantic artefacts describing our Carbon Portal services, as well as typing and definitions of more variables and concepts (following the RDA I-ADOPT recommendations). We also note that our landing pages are encoded with relevant schema.org attributes to support machine interpretability and actionability.
<b>Blue-Cloud 2026</b>	
<b>IVOA, ~ESCAPE</b>	(see above description by Baptiste Cecconi, it fits also my case)
<b>ELIXIR</b>	We utilise and integrate over 50 community ontologies into our own applicaiton ontologies for use within FAIRsharing ( <a href="https://github.com/FAIRsharing/subject-ontology">https://github.com/FAIRsharing/subject-ontology</a> , <a href="https://github.com/FAIRsharing/domain-ontology">https://github.com/FAIRsharing/domain-ontology</a> ). We create new ontologies (e.g. terms4FAIRskills, <a href="https://github.com/terms4fairskills/FAIRterminology/">https://github.com/terms4fairskills/FAIRterminology/</a> ) to fulfil community needs. We describe nearly 800 terminologies and the relationships they have to the databases, other standards and data policies within research data.
<b>FAIRCORE4EOSC</b>	Our project builds the new services or extends existing ones that also relate to the above topics. An overview of these services at: <a href="https://faircore4eosc.eu/eosc-core-components">https://faircore4eosc.eu/eosc-core-components</a> The project does not directly work with data, except with some specific case studies for testing and validating the components under development.
<b>BY-COVID</b>	We provide recommendations and cross-domain mappings. We integrate metadata into the COVID-19 Data Portal. See also <a href="https://by-covid.org/outputs">https://by-covid.org/outputs</a>
<b>Virtual Atomic and Molecular Data Centre</b>	Our project organize the knowlege representaiton in form of controlled vocabularies <a href="http://dictionary.vamdc.eu">http://dictionary.vamdc.eu</a>

Project or initiative	2. In your project/institution/community do you publish your (meta)data following FAIR principles?
<b>OpenAIRE</b>	<a href="https://guidelines.openaire.eu/en/latest/">https://guidelines.openaire.eu/en/latest/</a>
<b>RDA France</b>	
<b>GO FAIR Foundation</b>	We have a whole training programme about the FAIRification of metadata: <a href="https://osf.io/bthf8">https://osf.io/bthf8</a> . We are following FAIR Impact challenges to improve our method and align with the suggested standards for metadata for vocabularies.
<b>FAIR-IMPACT</b>	As much as we can. We also assess the level of FAIRness of the resources we host with the O'FAIR tool.
<b>FAIR-IMPACT</b>	As much as we can. We support the DataINRAE platform, part of the French Research Data Gouv data repository. We support the scientists in adopting the FAIR principles at all level.
<b>FAIR-EASE</b>	In FAIR-EASE project we don't publish data per se but we use published datasets and we publish them internally via the asset register for our own purposes. This internal publication follows the FAIR principles.
<b>Blue-Cloud2026</b>	Yes, the DDAS metadata/data is following the FAIR principles. And also data products created within the project are published following the FAIR principles as much as possible.
<b>Semantics in Astronomy, Planetary Sciences and Heliophysics</b>	There is no community wide guidelines, but local guidelines or examples of FAIR implementation for specific projects. In IVOA, this is often presented in the DCP (Data Curation and Preservation) Interest Group: <a href="https://wiki.ivoa.net/wiki/bin/view/IVOA/IvoaDCP">https://wiki.ivoa.net/wiki/bin/view/IVOA/IvoaDCP</a>
<b>Reactome database of biomolecular pathways</b>	Yes. All our data and software is well indexed (for example through schema.org support), accessible under permissive licences (CC0 or similar), using external reference systems, and exported in community standards.
<b>A CRIS system for Open Science and FAIR publications</b>	
<b>University of Bologna (FAIR Champion + Data Stewards)</b>	An institutional policy on FAIR (meta)data management is being approved by the end of the year. Up to now, FAIR (meta)data management is mandatory for researchers with EU fundings and strongly recommended to other researchers, e.g., through a training course for PhD students and a video series on Open Science principles for the university research community. The university has an institutional repository for data deposit, AMS acta ( <a href="https://amsacta.unibo.it/">https://amsacta.unibo.it/</a> ), using Datacite and Dublin core as metadata standards for deposited records.

Project or initiative	2. In your project/institution/community do you publish your (meta)data following FAIR principles?
<b>PerSciDo: The French Labex Persyval-Lab</b>	The metadata is displayed following FAIR principles: •Findable: Metadata can be searched using a simplified web search, advanced querying using the SPARQL language, or through an application using the PerSciDo API. •Accessible: Datasets are stored permanently. Access to the data requires only authentication on the PerSciDo platform. Some datasets have restricted access with a request to the data owner. The authorization request and retrieval process are automated. •Interoperable: Metadata is organized in an ontology that utilizes standard or emerging vocabularies as much as possible. Links with other similar platforms in the national network have been defined. An interface with the Dataverse system is currently under consideration. •Reusable: In the absence of a precise semantic description, each dataset must be accompanied by a document precisely describing the data and how to use it. This document is validated by a restricted committee before the dataset is published.
<b>Integrated Carbon Observation System (ICOS)</b>	In ICOS, all digital objects that we curate and manage are assigned a Handle-based persistent identifier. For data, these are based on the object SHA-256 checksum, and resolve to dynamically created landing pages that are readable by both humans and machines. The rich landing page metadata provided is extracted by SPARQL queries to our RDF-based metadata store. In this sense, the data and metadata share the same PID. The result is highly FAIR digital objects.
<b>Blue-Cloud 2026</b>	
<b>IVOA, ~ESCAPE</b>	(idem)
<b>ELIXIR</b>	Yes
<b>FAIRCORE4E OSC</b>	partially n.a. as some parts of our project does not produce metadata, but for OpenAIRE and DataCite I cannot say
<b>BY-COVID</b>	The BY-COVID project contributes to the FAIR principles by carrying out inter-domain metadata mappings, and by improving data discovery, integration and citation, using tools based on the Covid-19 Data Portal and FAIRsharing.
<b>Virtual Atomic and Molecular Data Centre</b>	Yes, we publish metadata following FAIR principles



Project or initiative	3. Do you reuse data from other project/institution/community?
<b>OpenAIRE</b>	OpenAIRE graph and explore service. <a href="https://graph.openaire.eu/">https://graph.openaire.eu/</a>
<b>RDA France</b>	
<b>GO FAIR Foundation</b>	not directly - the communities in our workshop do
<b>FAIR-IMPACT</b>	Our data are ontologies and semantic artefacts. We do use and encourage other to use ontologies/SA from other especially when they are representing a reference or standard.
<b>FAIR-IMPACT</b>	Possibly
<b>FAIR-EASE</b>	Yes, we are using data from a wide range of data infrastructures from the Earth-Science domains (about 20).
<b>Blue-Cloud2026</b>	Blue-Cloud makes use of federated data from many different Blue Data Infrastructures.
<b>Semantics in Astronomy, Planetary Sciences and Heliophysics</b>	The astronomy/planetary sciences and heliophysics communities have been reusing data for a long time. The citation on data is though not completely implemented, but all data providers are considering it. However, the data have been readily reusable for long. The recent/current developments on reusability is on the citation, the link to datasets from publication, the dataset provenance...
<b>Reactome database of biomolecular pathways</b>	We are importing high confidence molecular interactions from the IntAct database. No problems, as IntAct is FAIR. We are importing gene-disease associations from the DisGeNET database. Their CC BY-NC-SA 4.0 licence restricts access for our commercial users. We are importing KEGG data for a specific functionality. KEGG's restrictions on redistribution mean that we can't redistribute an integrated dataset we provide on our website.
<b>A CRIS system for Open Science and FAIR publications</b>	
<b>University of Bologna (FAIR Champion + Data Stewards)</b>	Many researchers reuse already produced data, particularly in some domains. The main challenge associated to the reuse is the discovery and understanding of the permitted uses as stated by the licenses associated to the data. The second issue is the harmonization of different datasets to combine them for reuse, which would benefit from increased standardization.
<b>PerSciDo: The French Labex Persyval-Lab</b>	Standards: Dublin Core, Friend Of a Friend, Creative Commons DOI: Datacite. Emerging standards when defining the ontology for describing research data: Fabio, RADAR

Project or initiative	3. Do you reuse data from other project/institution/community?
<b>Integrated Carbon Observation System (ICOS)</b>	ICOS itself is mainly concerned with collecting, curating and sharing observational data from our own organisation. However, we also create and supply various data products based on modelling, and these require inputs from external parties such as greenhouse gas emissions inventories, weather reanalysis data and global vegetation models. Challenges include ensuring the alignment of temporal and spatial coordinate systems, the treatment of gaps in time series and the interpretation of various quality flagging conventions.
<b>Blue-Cloud 2026</b>	
<b>IVOA, ~ESCAPE</b>	In 2 projects, dealing one with galactic astrophysics and space weather we are re-using/re-annotating existing data from our own and other institutions
<b>ELIXIR</b>	Yes
<b>FAIRCORE4E OSC</b>	
<b>BY-COVID</b>	The COVID-19 Data Portal brings together and continuously updates the relevant COVID-19 datasets and tools. <a href="https://www.covid19dataportal.org/">https://www.covid19dataportal.org/</a>
<b>Virtual Atomic and Molecular Data Centre</b>	We use standards from other communities to represent our data/metadata

Project or initiative	4. What are the worst data practices adopted in your project/institution/community?
<b>OpenAIRE</b>	
<b>RDA France</b>	
<b>GO FAIR Foundation</b>	Communities need time to align with FAIR Principles. They often come with non-FAIR approaches but willing to learn how FAIR Implementation can look like. We have seen various szenarios from metadata using free text to metadata using lists of terms published on a website or implemented in an application. T he worst case is when communitis have not agreed on a common way on how to provide metadata. In M4Ms they learn how they could agree based on widely-used metadata schemas or can be informed via FIPs on metadata schemas adopted by communities in the same domain.
<b>FAIR-IMPACT</b>	Minting URIs without suporting/providing them as GPRUID (Global, Permament, Resolvable, Unique Identifier). Especially, the resolvable aspect.
<b>FAIR-IMPACT</b>	



Project or initiative	4. What are the worst data practices adopted in your project/institution/community?
<b>FAIR-EASE</b>	<p>As we are currently analysing datasets from different communities the most important issue we find is semantic artefact referencing which is using:</p> <ul style="list-style-type: none"> <li>concatenation of strings and identifiers which makes it difficult to know what semantics have been used thus enable mappings and interoperability.</li> </ul> <p>By the end of this project we will have produced a set of best practices on how to reference semantics from metadata records and data files</p>
<b>Blue-Cloud2026</b>	Some of the metadata elements are not fully interpretable for machines yet, like data quality information or licensing info.
<b>Semantics in Astronomy, Planetary Sciences and Heliophysics</b>	<ul style="list-style-type: none"> <li>- Authors not willing to openly share their data and software.</li> <li>- When authors are willing to share, editors not willingly using author-provided PIDs for citing data in paper : some editor are proposing to cite data, but few are actually including the PID reference in the Crossref metadata, so that the knowledge link can only be inferred parsing (or reading) the paper itself.</li> </ul>
<b>Reactome database of biomolecular pathways</b>	Unclear, restrictive licences (see 3).
<b>A CRIS system for Open Science and FAIR publications</b>	
<b>University of Bologna (FAIR Champion + Data Stewards)</b>	Main worst data practices are related to carelessness in interpreting and using the concepts of open and licenses. Another worst practice is double depositing. Another worst practice is to believe that supplementary materials of publications are properly deposited data.
<b>PerSciDo: The French Labex Persyval-Lab</b>	Datasets are described in heterogeneous ways, meaning that different numbers of metadata are used. This is inherent to data referencing, which requires effort from the user in the role of the depositor. The effort required should be minimal to motivate researchers to reference and deposit their datasets. Therefore, it involves identifying a small subset of required metadata and allowing the user to enter other metadata as needed. As a result, each dataset will be described to varying degrees of richness.
<b>Integrated Carbon Observation System (ICOS)</b>	The ICOS end user communities (we cover three subdomains of environmental & Earth science: atmosphere, ocean and terrestrial ecosystems) are still in the process of adapting to "modern" RDM practices, including how to cite datasets, preparing their research outputs according to Open Science criteria etc. The implementation of a consistent use of semantic artefacts for e.g. describing data contents is also quite slow. But overall, I don't think there are any real bad practices!

Project or initiative	4. What are the worst data practices adopted in your project/institution/community?
<b>Blue-Cloud 2026</b>	
<b>IVOA, ~ESCAPE</b>	Starting a new data and service initiative (project, infrastructure, ...) without FAIR in mind and leaving it all to retrofitting attempts when all is already in place.
<b>ELIXIR</b>	
<b>FAIRCORE4EOSC</b>	
<b>BY-COVID</b>	Metadata standards and/or vocabularies are not always used consistently. Unclear descriptions of data access and use conditions or restrictions.
<b>Virtual Atomic and Molecular Data Centre</b>	When discovered, error fixing may take too much time