

### meosc

FAIR-IMPACT Expanding FAIR solutions across EOSC

### Synchronisation Workshop 2023

Metadata, semantics and interoperability November 28th 2023

Esteban Gonzalez, UPM and Anne-Sofie Fink, DeIC



# Let start with an small exercise

https://www.mentimeter.com/

Code: 1819 5806



Metadata, semantics and interoperability



### Glossary

**Metadata**: "Metadata represents data about data. Metadata enriches the data with information that makes it easier to find, use and manage." (source: ontotex)

Semantic artefacts: "Semantic artefacts is a variety of digital representation of formats, e.g. RDF, Turtle, OWL-RDF, XML, JSON-LD. Artefacts may have a broad range of formalisation: Loose set of therms, taxonomies, thesauri, higher order of logistics including concepts, terms and classes. " (source: FAIRSFAIR)

Interoperability: "Ability of organisations to interact towards mutually beneficial goals, involving the sharing of information and knowledge between these organisations, through the business processes they support, by means of the exchange of data between their ICT systems." (source:EIF European interoperability framework )



### Recommendations from previous edition

- More cross-disciplinary work is needed to align semantic artefacts with the same terms or concepts.
- Maintenance, sustainability, and governance of semantic artefacts deserve attention and agreement across disciplinary communities.
- The FAIR-at-large community should intensify the work on crosswalks and mapping

### CODE: 2489 3480



### Objective of this session

The aim of the workshop will be to explore the semantic and technical interoperability adopted by different research communities to explore their compliance with the EOSC IF recommendations...

... and try to *interoperate* all together.





### Agenda

- 1. Introduction to EOSC IF and recommendations.
- 2. Inspiring talks
  - a. Mappings and crosswalks by Yann Le Franc
  - b. FAIRsharing: mapping the interoperability landscape of standards, databases and policies by Allyson Lister
- 3. Interactive session.
- 4. Discussion session.



### Some details for this session.

- The session will be recorded but only for internal use for the rapporte report writing.
- Shared <u>spreadsheet</u> and <u>note taking document</u> for use in the session afterwards).
  - In the spreadsheet please keep information factual, short and include links w possible.
  - In the note taking document you may add more detail and background.
  - Please be careful when editing spreadsheet cells.
  - Please refrain from editing other people's information even typos.



### Some more details for this session.

- You may wish to add more information later, spreadsheet & note taking document will be available until **December 10th**
- Survey responses have been added to the spreadsheet
- 4 questions for this session but you may not have answers for them all.



### Questions

- What does your project/institution/community do to implement semantic artefacts (ontologies, schemas, standard, vocabularies, etc) into your (meta)data?
- 2. In your project/institution/community do you publish your (meta)data following FAIR principles?
- 3. Are you reuse data from other project/institution/community?
- 4. What are the worst data practices adopted in your project/institution/community?

We have received 10 responses. Thank you!



### Questions

- 70% use any kind of semantic artefact (only 2 ontologies).
- 70% have some FAIR guideline to publish their data
- 40% reuse data from other domains
- Worst data practices
  - metadata using free text
  - no agreement on how to provide metadata inside communities
  - editors don't use PID provided by authors to cite data
  - carelessness in using the concepts of open and private

### Do you know more bad data practices?



# EOSC IF

(Oscar Corcho et al.)





### The EOSC Interoperability Framework is a set of policies and guidelines that enable **interoperability of resources and services**, and will facilitate service composability.







The European Interoperability Framework four levels of interoperability





### Expanding FAIR solutions across EOSC



#### Problems

Lack of (or overabundance of)

P1: explicit definitions P2: common semantics (general ontologies) P3: reference repository P4: common metadata scheme across communities P5: metadata models

#### Needs

N1: principle approaches/tools for ontology and metadata schemes N2: harmonisation across disciplines N3: harmonisation of data of the same type N4: federated access to existing research data repositories

#### Recommendations

R1: definitions of concepts, metadata and data schemes R2: creating semantic artefacts with open licenses R3: associated documentation for semantic artifacts R4: repositories of semantic artefacts R5: minimum metadata model and cross walks discovery R6: extensible options for disciplinary metadata R7: apply a broad definition of data (datasets, workflows, lab protocols, software, methods, hardware design, etc.) R8: clear protocols and building blocks for catalogues





### Take-home messages

- Semantic interoperability is a clear need to achieve more inside and across scientific communities, and to realise the EOSC vision
  - Also applicable to other contexts: data spaces, open data, etc.
- Although well-studied and described, we are still "fighting" with some key problems...
  - Lack of common explicit definitions of terms used inside and across communities
  - Lack of common semantic artefacts across communities
- ... and with the design and implementation of the key enabling components / building blocks
  - Catalogues of semantic artefacts and their metadata
  - Crosswalks across metadata models (mapping repository)
- And finally, a proper governance of all these artefacts inside communities is needed (but that's for another talk and discussion)



### The EOSC Interoperability TF is conducting a landscape overview of semantic interoperability

### They are looking for use cases

Visit the Semantic Interoperability Profile wizard! https://sip-wizard.ds-wizard.org/



FAIR IMPACT Synchronisation Force 2nd Workshop - 28 November 2023 Topic Session 2: Metadata, semantics and interoperability



# Mapping the interoperability landscape of standards, databases and data policies

Allyson Lister, FAIRsharing Content & Community Lead Prof. Susanna-Assunta Sansone











### An informative and educational resource, and a service



 FAIRsharing provides curated descriptions and relationship graphs of

 standards, databases and policies

 FAIRsharing.org
 C BY-SA 4.0 International

### Our mission and how we deliver it

Promote the *value* and *use* of standards, databases and policies in all disciplines, by engaging stakeholders across all sectors, through all stages of the research life cycle



Guides consumers to discover, select and use these resources with confidence.

Helps producers to make their resources more visible, more widely adopted and cited.

Provides humans and tools with access to trustworthy content to enable data management

tasks.

FAIRsharing.org Educational

### FAIRsharing in numbers: content, contributions, coverage



### Working with and adopted by all stakeholders, all disciplines

Users, adopters and collaborators include:



https://fairsharing.org/communities



### The landscape of (meta)data standards

#### Example: life and biomedical sciences





### **Educational Content**



Series, focusing on:

- Content
- Users
- Functionalities
- Behind the scenes
- Tips



https://fairsharing.org/educational





- helps *developers* to make their standards more visible, more widely adopted and cited

#### Standards ...

set of requirements,

Are a **collectively agreed-upon** 

characteristics that can be used

for the **description**, **structure**,

sharing, and/or preservation of all kinds of data and metadata

interoperability, and use of data

with little/no human intervention;

observations, a list of measurements, descriptions of certain objects,

be of many types, including descriptive, administrative, and legal

enable humans to understand and reuse data at scale

\* Where data can simply be a piece of information, e.g.,

FAIRsharing

CC BY-SA 4.0 International

specifications, guidelines or

harmonisation, citation,

computational accessibility,

Help **machines** with

#### FAIRsharing categorises standards with four types:

#### Reporting guidelines

Outline in narrative form the necessary and sufficient information that should be reported about data, such as in itemised, prescriptive checklists; or the features and behaviours that should be followed, such as in general guiding principles

#### Models and formats **Z**

Define the representation of information for use by machines; these range from conceptual models to transmission formats, facilitating data retrieval and exchange between systems

#### Terminology artefacts

Add an interpretive, semantic layer for use by machines and humans; these range from controlled vocabularies (lists of terms, often with definitions) to ontologies (complex hierarchical groupings), providing unambiguous identification of concepts and aiding data querying

#### Identifier schemata

Are formal systems to identify information in an unique, machine-readable way: these persistent identifiers (PIDs), minted by recognised registries, build reliable metadata specifies the relevant information about the data, and can and long-lasting links between data, people, organisations and infrastructures

> Cite this: 10.5281/zenodo.8186982 🍋 Find more at Malin Sandström 0000-0002-8464-2494 💿 fairsharing.org/educational Allyson Lister 0000-0002-7702-4495 🧿 Susanna-Assunta Sansone 0000-0001-5306-5690 😳

#### https://fairsharing.org/educational

#### Cite this: 10.5281/zenodo.8186982 🚭

Malin Sandström 0000-0002-8464-2494 🗈 Allyson Lister 0000-0002-7702-4495 💿 Susanna-Assunta Sansone 0000-0001-5306-5690 💿















### Collections: tailored views for education and promotion

**Collections** are branded pages that *group* selected **standards** and/or **repositories** Initiatives and projects have created them for several purposes, e.g. to list resources:

#### **Developed by the community** International Virtual Observatory Alliance (IVOA) Maintainers carviset **Subjects** Collection URL: fairsharing.org/IVOA **Developed by a SDO** cdisc Maintainers awhite $(\bullet \bullet)$ **Subjects Biomedical Science Preclinical Studies**

Collection URL: fairsharing.org/CDISC

EXAMPLES

#### **Recommended by a community**



Collection URL: fairsharing.org/RDACovid19WG

#### Mapped to each other



Collection URL:

fairsharing.org/CrosswalkOfMostUsedMetadataSchemesAndGuidelines

FAIRsharing.org Educational

## FAIRsharing makes your resource discoverable to an entire ecosystem of resources



### 'Live' Knowledge graphs to complement lists

To discover and search the 230 related **standards**, part of the specification developed by the ISO Technical Committee on Biotechnology Processes





URL: https://committee.iso.org/standard/68848.html



### 'Live' Knowledge graphs to complement lists



URL: https://fairsharing.org/ISO20691

### Powering 3rd party tools

A growing number of tools and services access

FAIRsharing API, and use it for look-up, selection and

content retrieval for standards and repositories in:

- creation of data management plans
- enrichment of guidance and training material
- assessment of FAIRness



"Apples to Apples" Comparisons



b Wilkinson, Mark D; b Sansone, Susanna-Assunta; b Grootveld Marjan; b Nordling, Josefine; b Dennis, Richard; b Hecker, David





Recommendation 4: Define a clear PID policy and a common security and privacy framework (EOSC Interoperability Framework - Technical recommendation p.6)

### **Defining PID-compliant Identifier Schemas**

Via this EOSC workshop, FAIRsharing will be implementing the current EOSC PID definition (<u>https://doi.org/10.2777/926037</u>) and curate all identifier schemas within FAIRsharing to provide a common point of truth for which identifier schemas are PIDs for the purposes of FAIR evaluation and assessment

meosc

FAIR Metrics and Data Quality Task Force

#### FAIR Assessment Tools: Towards an "Apples to Apples" Comparisons

Wilkinson, Mark D;
 Sansone, Susanna-Assunta;
 Grootveld Marjan;
 Nordling, Josefine;
 Dennis, Richard;
 Hecker, David

doi.org/10.5281/zenodo.7463421





- FAIRsharing provides content to support the **Plan** and **Assess** themes
- Within the **Track** theme, FAIRsharing is both a knowledge graph (for standards, databases, data policies and organisations) and provides content to other KGs
- FAIRsharing also provides a service for improving how people search and find information on these resources



### What can FAIRsharing do for you?

You require	FAIRsharing.org helps you via
Visibility for the <b>resources</b> you produce and endorse	<b>Collections</b> of resources to provide <b>live</b> <b>representations</b> and <b>showcase</b> <b>interoperability</b> of these resources
Visibility for the <b>organisations</b> responsible for them	Organisation-specific FAIRsharing pages that list all relationships
Integration of <b>resources</b> with an ecosystem of tools	<b>API</b> access used by many <b>tools</b> including those for FAIR assessment and data management planning
Ways to discover <b>resources</b> of use to your stakeholders	Search and graph functionality to discover resources for collaboration and use
	FAIRsharing.org 🔊 Educational

CC BY-SA 4.0 International

### FAIR-IMPACT recommendations and FAIRsharing

To enhance	You should
Recommendation 1: Develop domain and cross-domain interoperability frameworks at the level of vocabularies, ontologies, and metadata schema (FAIRSFAIR white paper - Rec1).	Discover <b>relationships</b> and use FAIRsharing record links in your guidance to members, <b>utilising stable PIDs</b> for resources you want to recommend.
Recommendation 2: Define maintenance of a repository of semantic artefacts and a governance framework for such a repository ( <u>EOSC Interoperability</u> <u>Framework</u> - Semantic recommendation p.6)	<b>Register</b> your <b>repository</b> , the <b>terminologies</b> within it (if not already present), and crosswalks in FAIRsharing to improve <b>visibility</b> to <b>humans</b> and <b>machines</b>
Recommendation 3: Develop interoperability frameworks for FAIR sharing within disciplines and for interdisciplinary research ( <u>Turning FAIR into reality</u> - <u>Rec4</u> )	Suggest that members use FAIRsharing to <b>discover</b> <b>standards</b> and <b>databases</b> for this work
Recommendation 4: Define a clear PID policy and a common security and privacy framework (EOSC Interoperability Framework - Technical recommendation p.6)	Implement and join EOSC FAIR Metrics and Data Quality TF common approaches to FAIR evaluation
<b>Representation</b> and <b>visibility</b> of your resources in FAIRsharing	Nominate representatives to <b>join</b> the <b>FAIRsharing</b> <b>Community Champions</b> to aid delivery of all of these FAIRsharing.org <b>For Educational</b>

CC BY-SA 4.0 International

### FAIRsharing Community Champion Programme

#### **Community Champions**



Our Maintainers and Community Champions are continue to update across

all disciplines, according to

their needs and interests

#### **Record Maintainers**





### FAIRsharing Community Champion Programme



fairsharing.org/community\_champions/our\_champions

A thriving community of domain and discipline experts who:

1. **act as advocates** to promote the value of standards, databases and policies for digital objects (incl. data, software).

2 . **create educational material** describing these resources helping researchers and other stakeholders to find, use and adopt them.

3 . **enrich the content** of FAIRsharing, adding and enhancing the description and discoverability of these resources.

변 University of Oxford – FAIRsharing: Oxford, GB			
2023-01-01 to 2023-12-31   FAIRsharing Community ( Service JRI: <u>FAIRsharing user page for Inanderscience, 2023</u> JRI: <u>FAIRsharing Community Champions Programme 2</u>		University of Oxford – FAIRsharing) Show more detail	
Source: 🔗 FAIRsharing.org	The following	community curators have contributed to this reco	
FAIRsharing.org: Oxford, GB		kcopas@gbif.org (p), jmiller@gbif.org (	
2022-01-01 to 2022-12-31   FAIRsharing Community ( Service JRI: FAIRsharing <u>user page for Inanderscience</u> JRI: FAIRsharing Community Curation Programme	Curator (Omics) 2022 (FAI	IRsharing.org) Show more detail	

#### **Stakeholder Advisors**

- Amye Kenall, Research Square
- Adam Leary, Oxford University Press
- Catriona MacCallum, Wiley
- Dagmar Meyer, European Research Council, Executive Agency
- David Carr, Global Biodata Coalition
- Dominic Fripp, JISC
- Emma Ganley, Protocols.io
- Geraldine Clement-Stoneham. Medical Research Council
- Graham Smith, Springer Nature
- Helena Cousijn, DataCite
- Imma Subirats, FAO of the United Nations
- Kiera McNiece, Cambridge University Press
- Lauren Cadwallader, PLoS
- Lorenzo Feri, Elsevier
- Luiz Olavo Bonino, GO-FAIR
- Mark Leggott, Digital Research Alliance of Canada
- Marta Teperek, Open Science NL
- Michael Ball, Medical Research Council
- Mike Huerta, NIH National Library of Medicine
- Nick Everitt, and Matthew Cannon, Taylor and Francis
- Peter McQuilton, (FAIRsharing Founding Member), GSK
- Rebecca Grant, F1000
- Richard Brown, BBSRC UKRI
- Robert Hanisch, National Institute of Standards and Technology
- Sarah Callaghan, Research Strategy & Policy Unit, University of Oxford
- Sarah Stewart, St. George's, University of London; Oxford Internet Institute
- Scott Edmunds, GigaScience, Oxford University Press
- Simon Hodson, CODATA
- Theo Bloom, British Medical Journal
- Thomas Lemberger, EMBO Press
- Varsha Khodiyar, Independent Expert
- Wei-Mun Chan, eLife

# **Thank you!**

#### **Community Champions**

fairsharing.org/community champions/our champions



FAIRsharing Registry: Connecting data policies, standards and databases RDA WG WG

www.rd-alliance.org/group/fairsharing-registry-connecting-data-policies-standards-databases.html

#### **RDA FAIRsharing WG Chairs**

- Graham Smith, Springer Nature .
- Holly Murray, Health Data Research UK .
- Peter McQuilton, GSK
- Rebecca Grant, F1000 .
- Simon Hodson, CODATA .
- Allyson Lister, University of Oxford
- Susanna-Assunta Sansone, University of Oxford

#### **Operational Team**







**Allyson Lister** 

Susanna-Assunta

**Philippe Rocca-Serra** 

Content and Community Coordinator

Principal Investigator and Founder

Co-Founder







**Delphine Dauga** 

Curation Manager



**Ramon Granell** 

Data Manager and Research

Software Engineer

**Dominique Batista** 

Front End Manager

Web Developer

DEPARTMENT OF ENGINEERING OXFORD SCIENCE



#### fairsharing.org/communities#governance

**Milo Thurston** 

Sansone

Technical Coordinator







@fairimpact\_eu /company/fair-impact-eu-project

