



A short personal introduction

- CEO and Scientific Director of e-Science Data Factory (France)
- Head of the EUDAT CDI Secretariat
- Trained as Neuroscientist 10+ years academic experience
- Specialised on Semantic Web and Linked Data, Ontology design -15 years experience



A short personal introduction

- Contributing to H2020 and HEurope Research projects
 - EUDAT 2020(2015-2018) Task lead
 - EOSC-Hub (2018-2020) Participant
 - FAIRsFAIR (2019-2021) Task lead
 - OntoCommons (2020-2023) Task lead
 - EOSC-Pillar (2019-2022) Technical Coordinator as CINES employee
 - FAIR IMPACT (2022-2025) Task lead
- Leading and contributing to international expert groups on Semantics and FAIR
 - Co-chair of the RDA Interest Group on Vocabulary and Semantic Services since 2017
 - Co-chair of the RDA FAIR Mapping WG
 - Member of the EOSC Semantic interoperability Task Force
 - Member of the GOFAIR Implementation Network on Cross disciplinary Semantic interoperability
 - Co chair of the FAIR Digital Object Semantic Working Group



Expanding FAIR Solutions across Europe

Call HORIZON-INFRA-2021-EOSC-01-05

Enabling discovery and interoperability of federated research objects across scientific communities

Expanding FAIR solutions in Europe

Partly following up on FAIRsFAIR

EU funded project

Coordination and Support Action

10 million euro

36 months, starting 1 June 2022 28 partners and affiliate entities

From 10 EU member states: NL, FI, FR, DK, IT, DE, ES, NO, BE, RO

and the UK



Semantic artefacts* are a key elements to achieving FAIR and these artefacts and their catalogues have to be FAIR too

*ontologies, terminologies, taxonomies, thesauri, vocabularies, metadata schemas and standards...

WP4 will develop and foster the uptake of a semantic framework for the governance, creation, mapping, sharing, reuse, FAIRness assessment and interoperability of semantic artefacts for EOSC.





Greater and more harmonised use of semantic artefacts throughout the EOSC ecosystem, leading to semantic interoperability within and between disciplines.



WP4's tasks



T4.1 (governance)

Nicola Fiore



T4.2 (lifecycle and catalogues)

Daniel Garijo +
Clement Jonquet
+ Alejandra
Gonzalez-Beltran



T4.3 (for research software)

Morane Gruenpeter



T4.4 (crosswalks and mappings)

Yann Le Franc



T4.5 (in-use in data repos)

Sophie Aubin

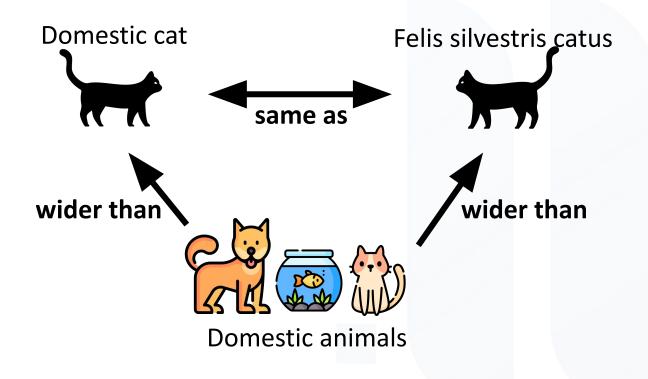






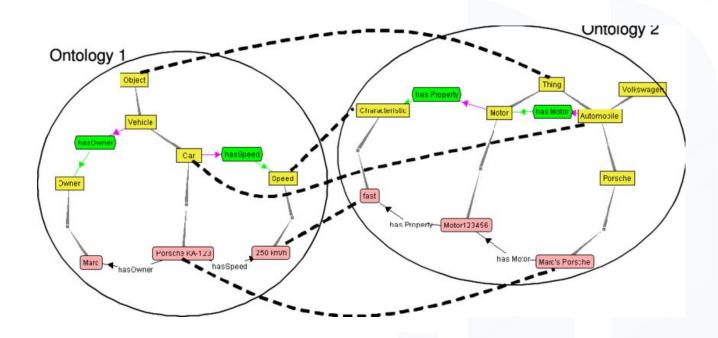


What are mappings and crosswalks?





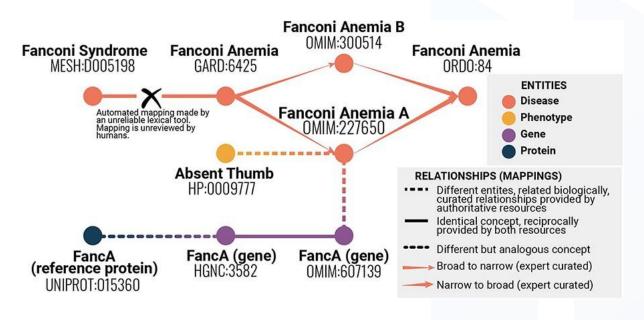
Mapping examples



From Ehrig, Marc & Staab, Steffen. (2004). QOM – Quick ontology mapping. 3298. 356-361. 10.1007/978-3-540-30475-3_47.



Mapping examples



From N. Matentzoglu et al., "A Simple Standard for Sharing Ontological Mappings (SSSOM)", *Database*, Volume 2022, 2022, baac035, https://doi.org/10.1093/database/baac035



When do we need mappings?

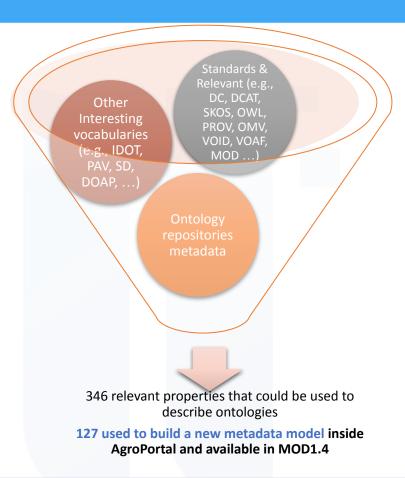
 Needed to integrate different information sources either as one information system or for enabling federated search

Part of the ontology development process: reducing ambiguities in information systems



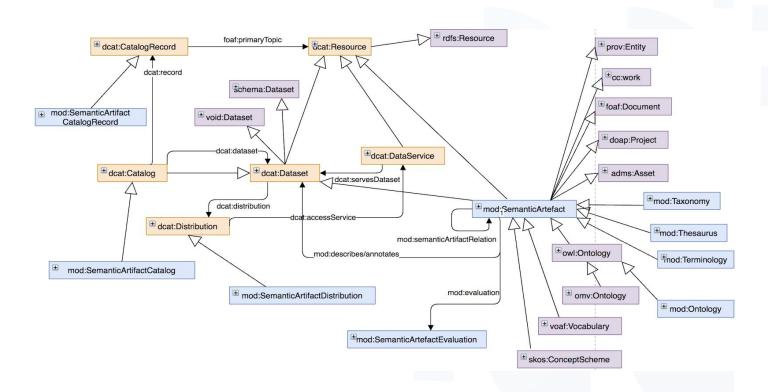
Review of ontology metadata practices to produce MOD

- Analysis of the use of metadata vocabularies in describing ontologies (by ontology developers)
 - 805 ontologies analyzed
- Analysis of the existing metadata vocabularies
 - 23 metadata vocabularies
- Analysis of the uses of metadata vocabularies in various ontology libraries and repositories (e.g., BioPortal, MMI, LOV, OBO Foundry)
 - 13 libraries



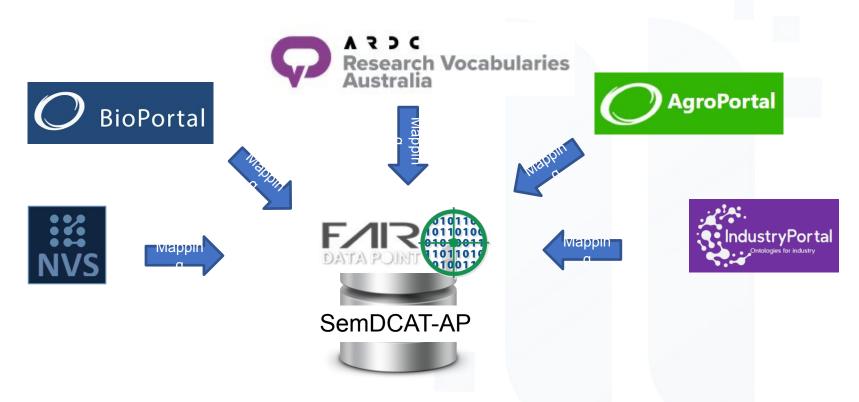


SemDCAT-AP: minimum metadata profile

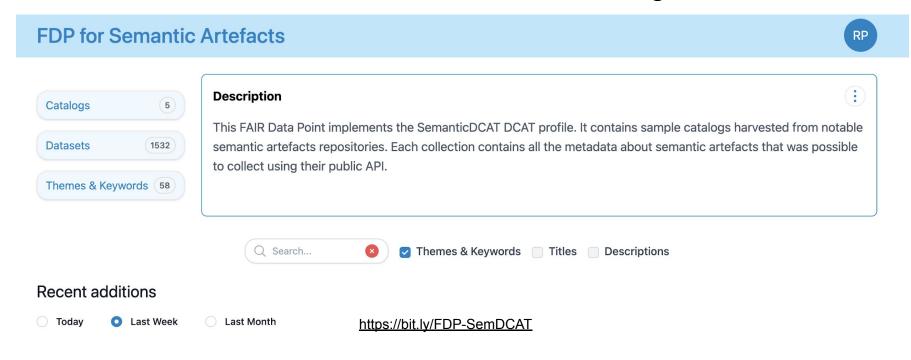




Mapping ontology repositories to SemDCAT-AP



A harmonised Semantic Artefact Metadata Catalog

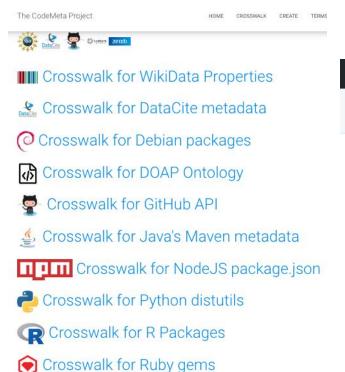


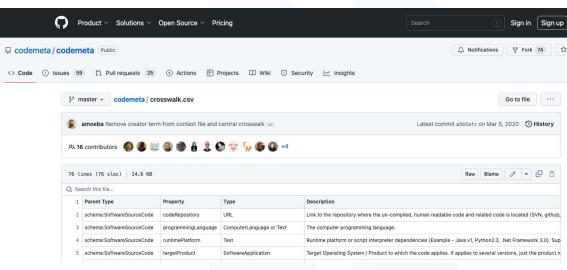
Clement Jonquet, Biswanath Dutta, Luiz O. Bonino da Silva Santos, Robert Pergl, Yann Le Franc. Common Minimum Metadata for FAIR Semantic Artefacts. Onto4FAIR 2023 - 2nd Workshop on Ontologies for FAIR and FAIR Ontologies, C. Trojahn; L. O. Bonino da Silva Santos; G. Guizzardi; C. Jonquet, Jul 2023, Sherbrooke, Canada. (hal-04106533v2)





CodeMeta Project: metadata for research software



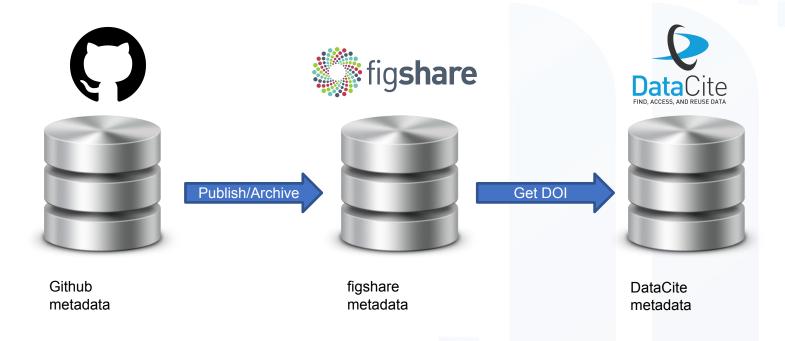








Reducing information loss across services









A predefined metadata model for software

Terms from Schema.org

Recognized properties for CodeMeta Code includes the following terms from https://schema.org. These terms are part of the CodeMeta specification and can be used without any prefer

Property	Type	Description	
codeRepository	URL	Link to the repository where the un-compiled, human readable code and related code is located (SVN GitHub, CodePlex, institutional GitLab instance, etc.).	
programmingLanguage	ComputerLanguage or Text	The computer programming language.	
runtimePlatform	Text	Runtime platform or script interpreter dependencies (Example - Java v1, Python2.3, Net Framework 3.0). Supersedes runtime.	
targetProduct	SoftwareApplication	Target Operating System / Product to which the code applies. If applies to several versions, just the product name can be used.	
applicationCategory	Text or URL	Type of software application, e.g. 'Game, Multimedia'.	
applicationSubCategory	Text or URL	Subcategory of the application, e.g. 'Arcade Game'.	
downloadUrl	URL	If the file can be downloaded, URL to download the binary.	
fileSize	Text	Size of the application / package (e.g. 18MB). In the absence of a unit (MB, KB etc.), KB will be assumed.	
installUrl	URL	URL at which the app may be installed, if different from the URL of the item.	
memoryRequirements	Text or URL	Minimum memory requirements.	
operatingSystem	Text	Operating systems supported (Windows 7, OSX 10.6, Android 1.6).	
permissions	Text	Permission(s) required to run the app (for example, a mobile app may require full internet access or may run only on wift).	

Codemeta terms

The CodeMeta project also introduces the following additional properties, which lack clear equivalents in https://schema.org but can play an important role in software metadata records covered by the CodeMeta crosswalk.

Property	Туре	Description
softwareSuggestions	SoftwareSourceCode	Optional dependencies , e.g. for optional features, code development, etc.
maintainer	Person	Individual responsible for maintaining the software (usually includes an email contact address)
contintegration	URL	link to continuous integration service
buildInstructions	URL	link to installation instructions/documentation
developmentStatus	Text	Description of development status, e.g. Active, inactive, suspended. See repostatus.org
embargoDate	Date	Software may be embargoed from public access until a specified date (e.g. pending publication, 1 year from publication)
funding	Text	Funding source (e.g. specific grant)
ssueTracker	URL	link to software bug reporting or issue tracking system
referencePublication	ScholarlyArticle	An academic publication related to the software.
readme	URL	link to software Readme file

Please suggest additional terms or adjustments to this representation in the codemeta issues









Tabular mappings between Codemeta model and DataCite

Property	DataCite
Property	DataCite
programmingLanguage	Format
author	creators
dateCreated	date
dateModified	date
datePublished	publicationYear
fileFormat	Format
license	rights
publisher	publisher
version	version
description	description











Metadata mappings problems

- String to string mapping
- String to object mapping
- Object to object mapping
- Field to field mapping (usage of different vocabularies)



String to object mapping problem

Different character strings mean the same thing

Example:

Milan Ojsteršek, Ojsteršek Milan, Ojstersek Milan, Ojstersek M., M. Ojstersek....

Solution:

Author normalization with authoritative database (for example ORCID, DBLP, VIAF...). We need a normalization service.

Transformation from string (for example dc.creator) to object in other metadata schema (for

example datacite.creator)

dct:creator

creatorName nameType givenName familyName nameIdentifier (R) nameIdentifierScheme schemeURI



Field to field mapping problem

to

Different vocabularies is used in semantically similar field in different metadata schemas

Example:

resourceTypeGeneral in Datacite schema 4.3 Audiovisual Collection **DataPaper** Dataset Event Image InteractiveResource Model **PhysicalObject** Service Software Sound Text Workflow Other

COAR Resource types 3.1 vocabulary

<u>dataset</u>

- aggregated data
- clinical trial data
- compiled data
- encoded data
- experimental data
- genomic data
- geospatial data
- <u>laboratory notebook</u>
- measurement and test data
- observational data
- recorded data
- simulation data
- survey data



DataCite metadata schema mapping

DataCite Schema

Schema 4.4 Older Versions Contribute Support /

Documentation

DataCite Metadata Working Group. (2021). DataCite Metadata Schema Documentation for the Publication and Citation of Research Data and Other Research Outputs. Version 4.4. DataCite e.V. https://doi.org/10.14454/3w3z-sa82

Download PDF

On the occasion of the release of v4.4 of the DataCite Metadata Schema its Metadata Working Group has updated the mapping to Dublin Core. This replaces the mapping in the Appendix of the DataCite-MetadataKernel v2.1.

Download PDF

Schema

DataCite Metadata Working Group. (2021). DataCite Metadata Schema for the Publication and Citation of Research Data and Other Research Outputs. Version 4.4. DataCite e.V. https://doi.org/10.14454/fxws-0523

View XSD

Changes

Addition of the new subproperty "classificationCode" in the Subject property.

ID	<u>DataCite-Property</u>	<u>Dublin Core</u>
1	Identifier	dcterms:identifier
1.a	identifierType	Not present in Dublin Core
2	Creator	dcterms:creator
2.1	creatorName	dcterms:creator
2.1.a	nameType	Not present in Dublin Core
2.2	givenName	Not present in Dublin Core
2.3	familyName	Not present in Dublin Core
2.4	nameldentifier	dcterms:identifier
2.4.a	nameldentifierScheme	Not present in Dublin Core
2.4.b	schemeURI	Not present in Dublin Core
2.5	Affiliation	dcterms:contributor
2.5.a	affiliationIdentifier	dcterms:identifier
2.5.b	affiliationIdentifierScheme	Not present in Dublin Core
2.5.c	SchemeURI	Not present in Dublin Core



DataCite metadata schema mapping

DataCite Schema

Schema 4.4 Older Versions Contribute Support /

Documentation

DataCite Metadata Working Group. (2021). DataCite Metadata Schema Documentation for the Publication and Citation of Research Data and Other Research Outputs. Version 4.4. DataCite e.V. https://doi.org/10.14454/3w3z-sa82

Download PDF

On the occasion of the release of v4.4 of the DataCite Metadata Schema its Metadata Working Group has updated the mapping to Dublin Core. This replaces the mapping in the Appendix of the DataCite-MetadataKernel v2.1.

Download PDF

Schema

DataCite Metadata Working Group. (2021). DataCite Metadata Schema for the Publication and Citation of Research Data and Other Research Outputs. Version 4.4. DataCite e.V. https://doi.org/10.14454/fxws-0523

View XSD

Changes

Addition of the new subproperty "classificationCode" in the Subject property.

ID	DataCite-Property	<u>Dublin Core</u>
1	Identifier	dcterms:identifier
1.a	identifierType	Not present in Dublin Core
2	Creator	dcterms:creator
2.1	creatorName	dcterms:creator
2.1.a	nameType	Not present in Dublin Core
2.2	givenName	Not present in Dublin Core
2.3	familyName	Not present in Dublin Core
2.4	nameldentifier	dcterms:identifier
2.4.a	nameIdentifierScheme	Not present in Dublin Core
2.4.b	schemeURI	Not present in Dublin Core
2.5	Affiliation	dcterms:contributor
2.5.a	affiliationIdentifier	dcterms:identifier
2.5.b	affiliationIdentifierScheme	Not present in Dublin Core
2.5.c	SchemeURI	Not present in Dublin Core



DataCite metadata schema mapping

DataCite Schema

Schema 4.4 Older Versions Contribute Support /

Documentation

DataCite Metadata Working Group. (2021). DataCite Metadata Schema Documentation for the Publication and Citation of Research Data and Other Research Outputs. Version 4.4. DataCite e.V. https://doi.org/10.14454/3w3z-sa82

Download PDF

On the occasion of the release of v4.4 of the DataCite Metadata Schema its Metadata Working Group has updated the mapping to Dublin Core. This replaces the mapping in the Appendix of the DataCite-MetadataKernel v2.1.

Download PDF

Schema

DataCite Metadata Working Group. (2021). DataCite Metadata Schema for the Publication and Citation of Research Data and Other Research Outputs. Version 4.4. DataCite e.V. https://doi.org/10.14454/fxws-0523

View XSD

Changes

Addition of the new subproperty "classificationCode" in the Subject property.

ID	DataCite-Property	<u>Dublin Core</u>
1	Identifier	dcterms:identifier
1.a	identifierType	Not present in Dublin Core
2	Creator	dcterms:creator
2.1	creatorName	dcterms:creator
2.1.a	nameType	Not present in Dublin Core
2.2	givenName	Not present in Dublin Core
2.3	familyName	Not present in Dublin Core
2.4	nameldentifier	dcterms:identifier
2.4.a	nameIdentifierScheme	Not present in Dublin Core
2.4.b	schemeURI	Not present in Dublin Core
2.5	Affiliation	dcterms:contributor
2.5.a	affiliationIdentifier	dcterms:identifier
2.5.b	affiliationIdentifierScheme	Not present in Dublin Core
2.5.c	SchemeURI	Not present in Dublin Core



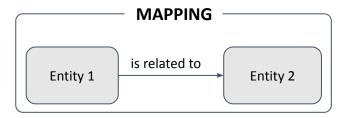
There are more types of mappings

Unit Conversion

- Mapping Nanomaterials shapes
- Knowledge graphs mappings
- Data Mappings
- API mappings



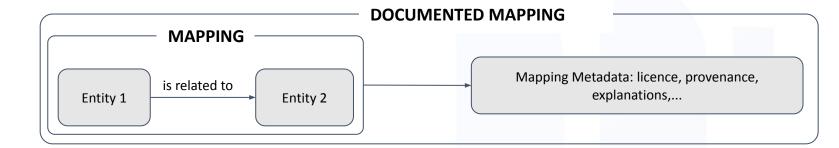
What do we mean by mappings?







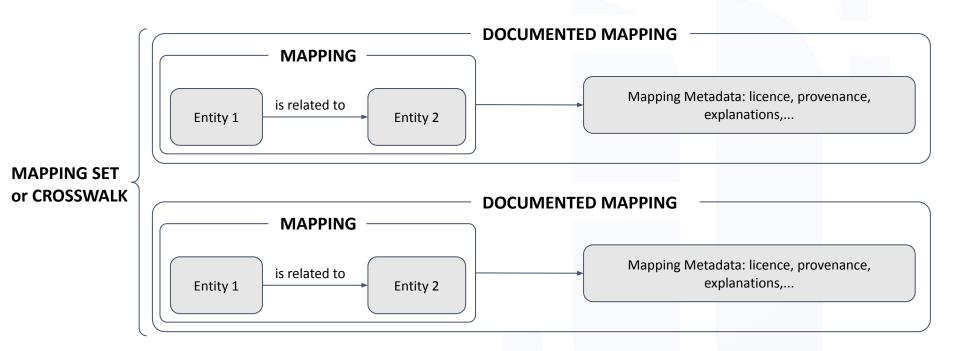
What do we mean by mappings?







What do we mean by mappings?





How are mapping stored?

Within the ontology or as machine readable format

In tabular format for simple correspondances shared on github.

• In research papers



Can I reuse mappings made by others?

- In few exemption, yes (e.g. Biomedical,...)
- Most of the time, no
 - Hard to find (need to know a guy who knows a guy who knows where to find them)
 - In some cases, not published
 - Not interoperable: relation is often not explicit and format varies.
 - Not reusable
 - No context explaining the reason for the mapping
 - No information about who created the mapping
 - Based on community standard practice and format requiring additional mappings (mappings of mappings)



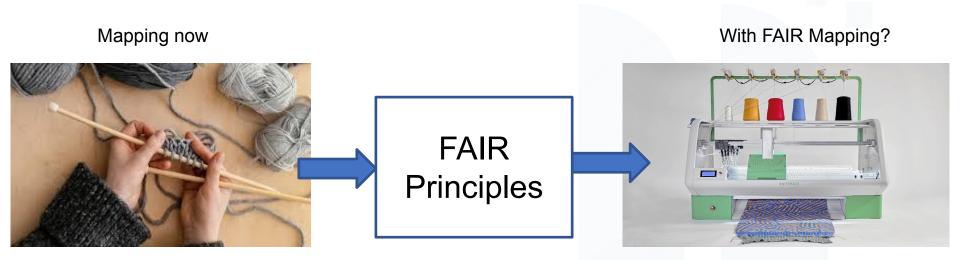
Cost/Benefits of FAIR Mappings

Mapping now





Cost/Benefits of FAIR Mappings





Mappings in FAIR Impact

Our objectives

- Establish guidelines on how to make mappings FAIR
- Propose a machine-actionable common exchange model for sharing the diversity of FAIR mappings
- Engage with communities to co-create, test and adopt the model for FAIR mappings and to identify methodologies and practices around mappings
- Establish a **governance framework** for mapping in collaboration with T4.1



Our ongoing work

Two sides of the same coins:

 Analysis of the requirements and technical recommendations for making mappings FAIR

 Practical aspects of mappings from creation to maintenance: understanding and documenting community practices



How do we work? Workshops and community engagement



Contribute to the survey on mappings and practices

https://bit.ly/fairmapping_survey







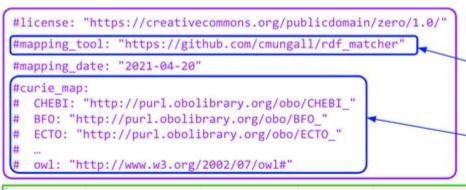
What about the FAIR Recommendations?

Grouped the 15 individual FAIR principles into 4 categories:

- Model and Format: Interoperability (I1, I2, I3) and Reusability (R1, R1.1, R1.2, R1.3)
- Metadata: Findability (F2, F3) and Reusability (R1, R1.1, R1.2, R1.3)
- **PID**: Findability (F1, F3)
- Service and API: Accessibility (A1, A1.1, A1.2, A2) and Findability (F4)



Model and format: 1st Recommendation





Header (metadata)

Source of mappings

CURIEs (registered prefixes) and their PURLs

Mappings

subject_id	subject_label	predicate_id	object_id	object_label	match_type	subject	object	mapping_tool	confidence	subject_match_fi
CHEBI:33282		owl:equivalentClass	XCO:0000483	antibacterial agent	Lexical	CHEBI	XCO	rdf_matcher	0.500024776	dc:identifier
ECTO:9001190	exposure to nitromethane	owl:equivalentClass	NCIT:C44412	Nitromethane	Lexical	ECTO	NCIT	rdf_matcher	0.500097687	olo:hasExactSyno
XCO:0000408	metformin	owl:equivalentClass	ECTO:9000391	exposure to metformin	Lexical Stemming	xco	ECTO	rdf_matcher	0.5000903614	rdfs:label
XCO:0000408	metformin	owf:equivalentClass	GO:1901558	response to metformin	Lexical	xco	GO	rdf_matcher	0.500042380	rdfs:label
NCIT:C44462	UV Radiation Exposure	owl:equivalentClass	XCO:0000042	ultraviolet ray exposure	Lexical	NCIT	XCO	rdf_matcher	0.500075951	olo:hasExactSyno
ECTO-9000308	exposure to fathy acid	multiproductions Class	GO:0070542	response to fathy acid	Levical/Stemming	ECTO.	GO	rdf matcher	0.500050796	nin-hasEvactSunn



#licen

Model and format: 1st Recommendation



- Use SSSOM as model to represent mappings and crosswalks
- Use SSSOM metadata schema
- SSSOM as common exchange model will require extensions to cover the different types of mappings.

XCO:0000408	metformin	owl:equivalentClass	ECTO:9000391	exposure to metformin	Lexical Stemming	xco	ЕСТО	rdf_matcher	0.5000903614 rdfs:label
XCO:0000408	metformin	owl:equivalentClass	GO:1901558	response to metformin	Lexical	xco	GO	rdf_matcher	0.500042380f rdfs:label
NCIT:C44462	UV Radiation Exposure	owl:equivalentClass	XCO:0000042	ultraviolet ray exposure	Lexical	NCIT	xco	rdf_matcher	0.500075951; oio:hasExactSynor
ECTO-9000308	exposure to fatty acid	nut-equivalentClass	GO:0070542	response to fathy acid	Levical/Stemming	ECTO.	GO	rdf matcher	0.500050796' olo:hasEvactSunor



Service and API

Services such as mapping repository/registry support the implementation of the FAIR principles

Recommendations provide additional requirements for these services.



coeosc FAIRCORE4EOSC

Enabling a FAIR EOSC ecosystem

Expected impact of the **Metadata Schema and Crosswalk Registry (MSCR):**

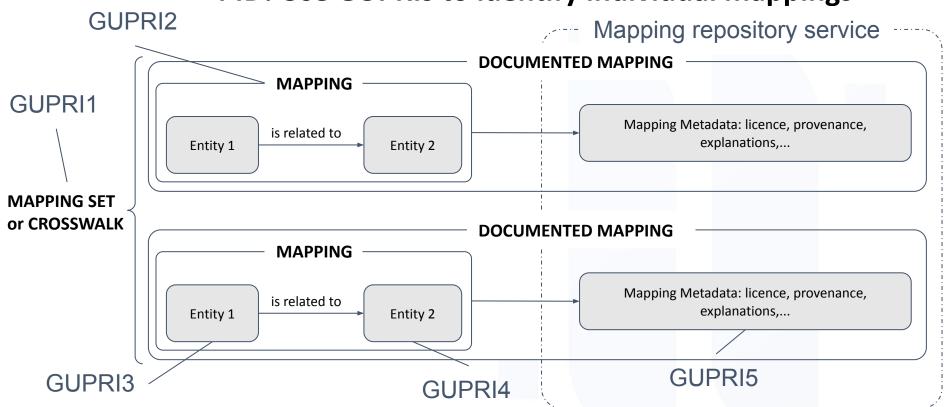








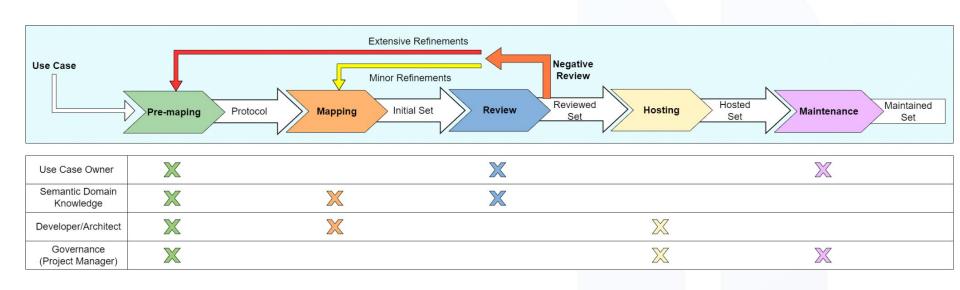
PID: Use GUPRIs to identify individual mappings







Mapping Framework: Summary





Framework: template

						•
_	Context	Are there existing mappings for this use case or similar?	▼	link (optional)	use case link	
	Context	If so, why are they unsuitable?	text		documented use case link	
	Nature	Is it a one-off solution or will it be maintained?	•		Protocol	
	Nature	If applicable, where will the solution be maintained	•	hosting resource link	Protocol	
	Nature	Have you decided on a license?	▼	link	Protocol	
	Nature	How will you share the protocol	▼	link	Protocol	
	Sustainability	If applicable, how will you incorporate feedback, refine & update	text or n/a		Protocol	
	Nature	Does the mapping involve any standard or established vocabulary as source?	•	vocabulary link	Protocol	
	Nature	Does the mapping involve any standard or established vocabulary as target?	▼	vocabulary link	Protocol	
	Scope	What will the solution look like?	▼		Protocol	
	Scope	What are the relationships between source and target terms? (semantic terms to be used)	•	terms link	Protocol	
	Scope	Will the mapping be manual or automated?	•		Protocol	
	Scope	identified source	URI		Protocol	
	Scope	Source - identify the concepts to be mapped	text or link(s)		Protocol	
	Scope	identified target	URI		Protocol	
	Scope	Target - identify the concepts to be mapped	text or link(s)		Protocol	

Template

Top sheet

+ Sheet per phase Reusing categories Indicate inputs/outputs

• E.g. protocol 'Guided'



RDA Working Group on FAIR Mappings

- Co-created with FAIRCORE4EOSC
- Current charter version: https://bit.ly/fair_mapping_wg_charter
- Planned submission for Community Review by the end of July
- Initial activities
 - Bird of a Feather session during RDA P22
 - Bird of a Feather session submitted for RDA P23



Planned WG outputs

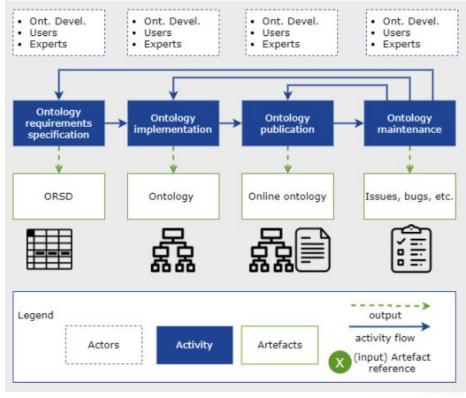
- Technical recommendations to make FAIR Mappings and FAIRness evaluation grid
- Practical mapping framework and guidelines
- Harmonized mapping use case collection
- Mapping classification/ontology
- Generic exchange model and the associated metadata
- FAIR Mapping Knowledge Base



Some useful resources to consider for putting together a research data management vocabulary



Linked Open Terms: methodology

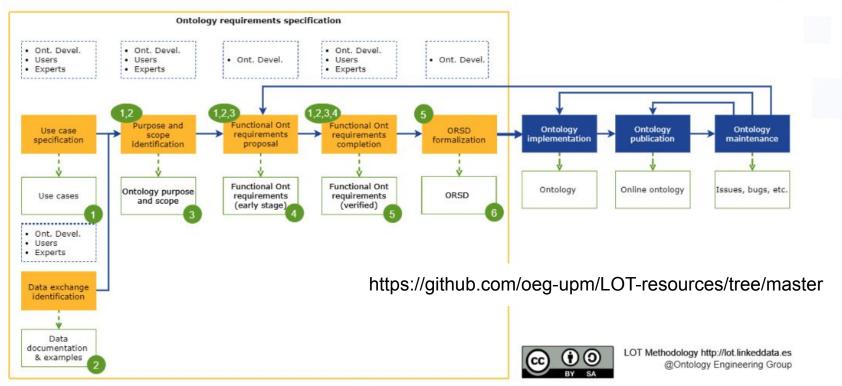


https://lot.linkeddata.es/

María Poveda-Villalón, Alba Fernández-Izquierdo, Mariano Fernández-López, Raúl García-Castro, LOT: An industrial oriented ontology engineering framework, Engineering Applications of Artificial Intelligence, Volume 111,2022, 104755, https://doi.org/10.1016/j.engappai.2022.104755.



Requirement specification





Other resources from FAIR IMPACT

- Processes & tools to engineer FAIR semantic artefacts
 - https://zenodo.org/records/10551054
- Semantic artefact governance models and disciplinary approaches for inclusion within EOSC
 - https://zenodo.org/records/13142842



Other resources for RDM Ontology

- Terms4FAIRSKills
 - https://zenodo.org/records/4772741

 - https://github.com/terms4fairskills/FAIRterminology CASRAI/RDM-Glossary (Codata): https://codata.org/new-machine-actionable-rdm-terminology-launched/
- DAMA-DM Book
 - https://www.dama.org/cpages/body-of-knowledge
- DCSO Ontology:
 - Cardoso, J., Castro, L.J., Ekaputra, F.J. et al. DCSO: towards an ontology for machine-actionable data management plans. J Biomed Semant 13, 21 (2022). https://doi.org/10.1186/s13326-022-00274-4