



**FAIRCORE4EOSC**  
Core Components Supporting a FAIR EOSC

# FAIRCORE4EOSC

Added value for research communities and other stakeholders



**Funded by  
the European Union**



# FC4E Semantic infrastructure components

Main purpose is to contribute to a Web of FAIR data and facilitate interoperability (structural and semantic) between different research communities, communities of practice and their infrastructures.

Metadata Schema and Crosswalk Registry (MSCR) – providing a registry and hosting facility for metadata schema including crosswalk/semantic mappings (between these), **publishing, discovery and access of metadata schemas**, and provide functions to operationalise metadata conversions by combining crosswalks and in combination with DTR functions

- Schema registries are already available either as general registries or discipline specific: [RDA metadata standards registry](#), [FAIRsharing](#), [Bioschemas](#), ...
- Crosswalks and semantic mapping registries are rare and mostly see mappings as subjugate of a schema or ontology
- Recommendations: [SEMAF report](#), [Fair Semantic recommendations report](#) call for semantic mappings as independent semantic artefacts



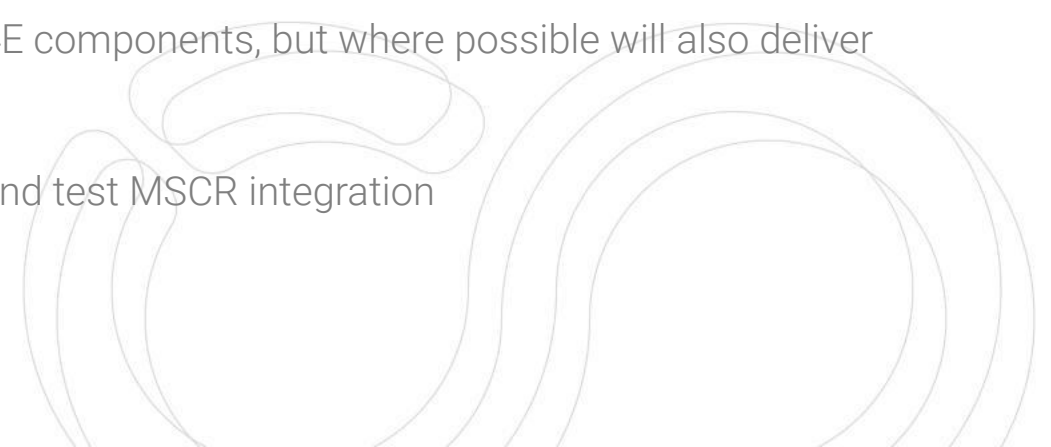
# Requirements for MSCR

Project proposal quotes and refers to existing reports and recommendations:

- FAIR Semantic Artefacts - recommendations from FAIRsFAIR project
- SEMAF report – proposes a flexible infrastructure for semantic mappings, making semantics mappings FAIR and inviting communities to share and reuse

Further requirements from:

- Community case-studies – where communities plan integrating & using the FC4E components in their workflow
- Demonstrators – intended to test and demonstrate the different FC4E components, but where possible will also deliver important new added value for users and infrastructure managers
- Where useful and possible other initiatives are invited to contribute and test MSCR integration



# Examples of semantic mappings

## Mapping between Darwin Core 1.4 concepts (DwC) and ABCD 2.06b

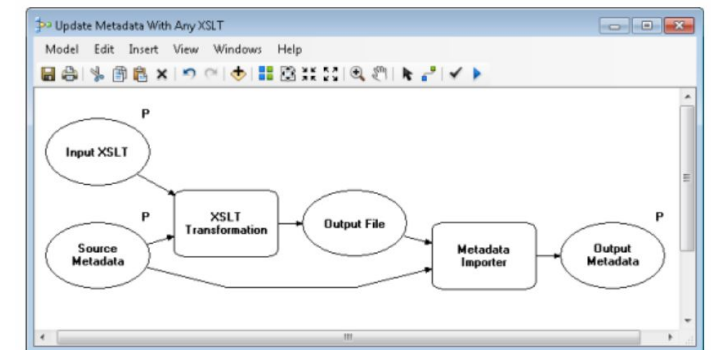
This document was originally based on a table dated August 4, 2005 and provided by Stan Blum to describe DwC "2" v. 1.4, see [http://darwincore.calacademy.org/Documentation/DarwinCore2Draft\\_v1-4\\_HTML](http://darwincore.calacademy.org/Documentation/DarwinCore2Draft_v1-4_HTML). It has been updated April 4, 2007 using the DwC Concept List r33 dated 17 Feb 2007 - 19:54:25 under <http://wiki.tdwg.org/twiki/bin/view/DarwinCore/DarwinCoreDraftStandard>. Deprecated DwC elements are still included (and marked accordingly) or commented on.

DwC 1.4 Record-level Element	ABCD 2.06b X-Path	DwC to ABCD	ABCD to DwC
	Datasets/Dataset/Units/Unit/...		
GlobalUniqueIdentifier	UnitGUID	Fully compatible.	Fully compatible.
DateLastModified	DateLastEdited	Fully compatible.	Fully compatible.
BasisOfRecord	RecordBasis	Fully compatible  DwC gives only recommendations for content. The examples given are the same as the restriction for ABCD, except that "StillImage" is used instead of "DrawingOrPhotograph", and that "MovingImage" and "SoundRecording" are listed, which should be mapped to "MultimediaObject" in ABCD	Fully compatible  ABCD is restricting content to values representing: "PreservedSpecimen", "LivingSpecimen", "FossileSpecimen", "OtherSpecimen", "HumanObservation", "MachineObservation", "DrawingOrPhotograph", "MultimediaObject" and "AbsenceObservation".
InstitutionCode	SourceInstitutionID	Fully compatible	Fully compatible
CollectionCode	SourceID	Fully compatible	Fully compatible

Entity 1	Entity 2
tectonic movement(ENVO:01001093)	Continental drift (SWEETPhenGeolTectonic:ContinentalDrift)
river bank (ENVO:00000143)	Riparian zone (SWEETRealmLandCoastal:RiparianZone)
marine benthic biome (ENVO:01000024)	Benthic zone (SWEETRealmOcean:BenthicZone)
leaf alternate placement(FLOPO:0001032)	Phyllotaxy (TO:0006014)
rhizome mass (FLOPO:0003190)	Rhizome dry weight (TO:0000556)
whole plant lifestyle (FLOPO:0980070)	Life cycle habit (TO:0002725)

Table 1: Example of pairwise mappings of ontologies from Biodiversity (Flora Phenotype Ontology/FLOPO and Plant Trait Ontology/TO) and Earth System Sciences (Environment Ontology/ENVO and Semantic Web for Earth and Environment Technology Ontology/SWEET). Mappings were created for the Biodiversity and Ecology track ([biodiv](#)) of the Ontology Alignment Evaluation Initiative (OAEI, [39]).

Often complex mapping specifications and conversion are combined in technologies such as XSLT, X3ML

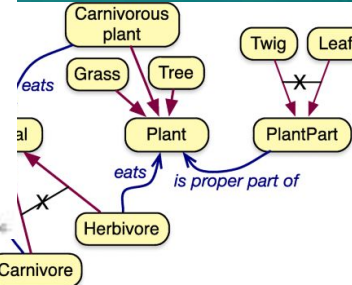
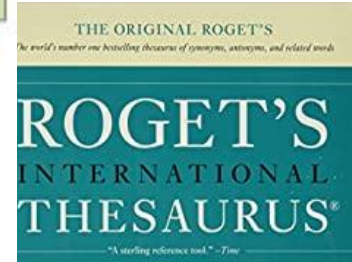
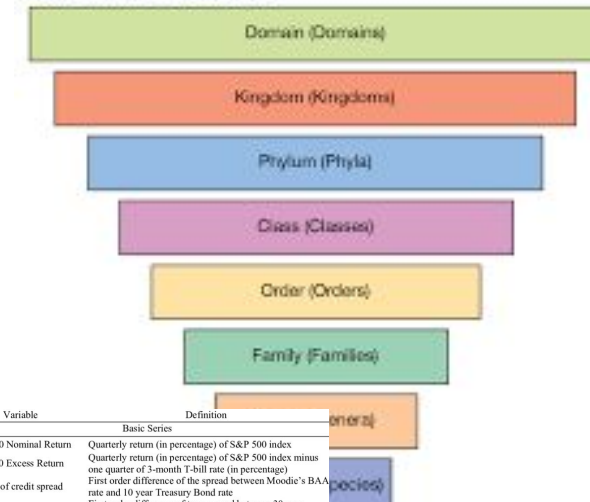


# Semantic artefacts

- Lists
- Glossaries
- XSD Schema
- *Thesauri*
- *Taxonomies*
- *Semantic mappings*
- *Ontologies*

W3C std. RDF, SKOS, OWL:  
widely accepted and supported formal way to represent vocabularies

How animals are classified

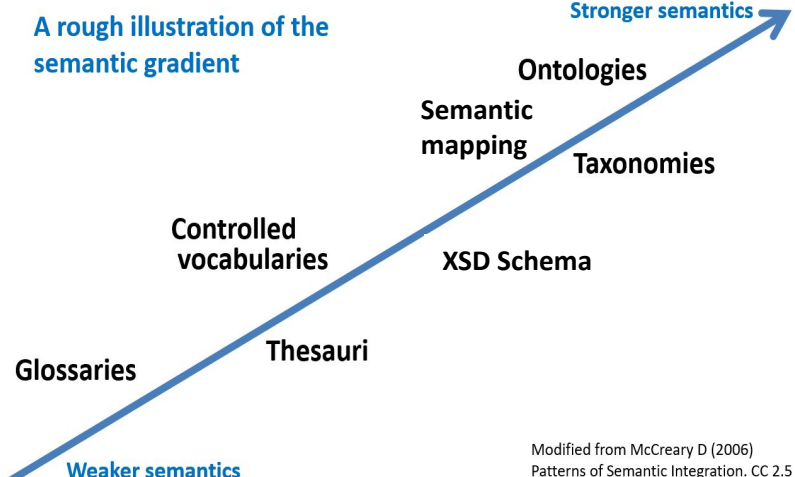


Symbol	Variable	Definition
Basic Series		
<i>NR</i>	S&P 500 Nominal Return	Quarterly return (in percentage) of S&P 500 index
<i>ER</i>	S&P 500 Excess Return	Quarterly return (in percentage) of S&P 500 index minus one quarter of 3-month T-bill rate (in percentage)
<i>ΔCSPR<sub>t</sub></i>	Change of credit spread	First order difference of the spread between Moody's BAA rate and 10 year Treasury Bond rate
<i>ΔTSPR<sub>t</sub></i>	Change of term spread	First order difference of term spread between 30 year Treasury Bond rate-3 month Treasury Bill rate
<i>ΔTBILL<sub>t</sub></i>	Change of Treasury bill rate	First order difference of 3-month Treasury Bill rate
Change of Trading Activity and Liquidity		
<i>ΔECM<sub>t</sub></i>	Exchange Commission	First order log difference of quarterly Exchange Commission Revenue
<i>ΔOCM<sub>t</sub></i>	OTC Commission	First order log difference of quarterly OTC Commission Revenue
<i>ΔMF<sub>t</sub></i>	Mutual Fund Commission	First order log difference of quarterly Mutual Fund Sales Revenue
<i>ΔNYSE<sub>t</sub></i>	NYSE Volume	First order log difference of quarterly NYSE share volume
<i>ΔMRG<sub>t</sub></i>	Margin Trading	Quarterly Margin Interest divided by quarterly T-bill rate, then take the first order log difference
<i>ΔECR<sub>t</sub></i>	Effective Commission Rate	Quarterly Exchange Commission Revenue divided by quarterly NYSE share volume, then taken first order log difference.
Variation of Trading Activity and Liquidity		
<i>VECM<sub>t</sub></i>	Variation of Exchange Commission	Logarithm squared distance to mean of Exchange Commission Revenue
<i>VOCM<sub>t</sub></i>	Variation of OTC Commission	Logarithm squared distance to mean of OTC Commission Revenue
<i>VMF<sub>t</sub></i>	Variation of Mutual Fund Commission	Logarithm squared distance to mean of Mutual Fund Sales Revenue
<i>NYSECV<sub>t</sub></i>	Variation of NYSE Volume	Coefficient of variation of daily NYSE trading volume in quarter t
<i>VECR<sub>t</sub></i>	Variation of ECR	Logarithm squared distance to mean of Effective Exchange Commission Rate



An authority record

- O'Brien, Flann, 1911-1966
- Na Gopaleen, Myles, 1911-1966
- Knowall, George
- Na gCopaleen, Myles, 1911-1966
- His At Swim-Two-Birds ... 1939
- His The best of Myles, 1983; CIP t.p. (Myles na Gopaleen (Flann O'Brien)
- His Myles away from Dublin, 1985; t.p. (Myles na Gopaleen (Flann O'Brien) selection written from the column written for ... under the name George Knowall)



Modified from McCreary D (2006) Patterns of Semantic Integration. CC 2.5

# Problem solving

Many metadata schema's, vocabularies and semantic mappings are not FAIR, not findable or accessible for non-initiated

As (metadata) schema and mappings are essential part of the scientific workflow they d need to be treated accordingly eg. first class citizens just as the research data itself

Formats for both schema and mappings are very diverse although RDF, OWL, SKOS offer much functionality table formats are still dominant in many disciplines

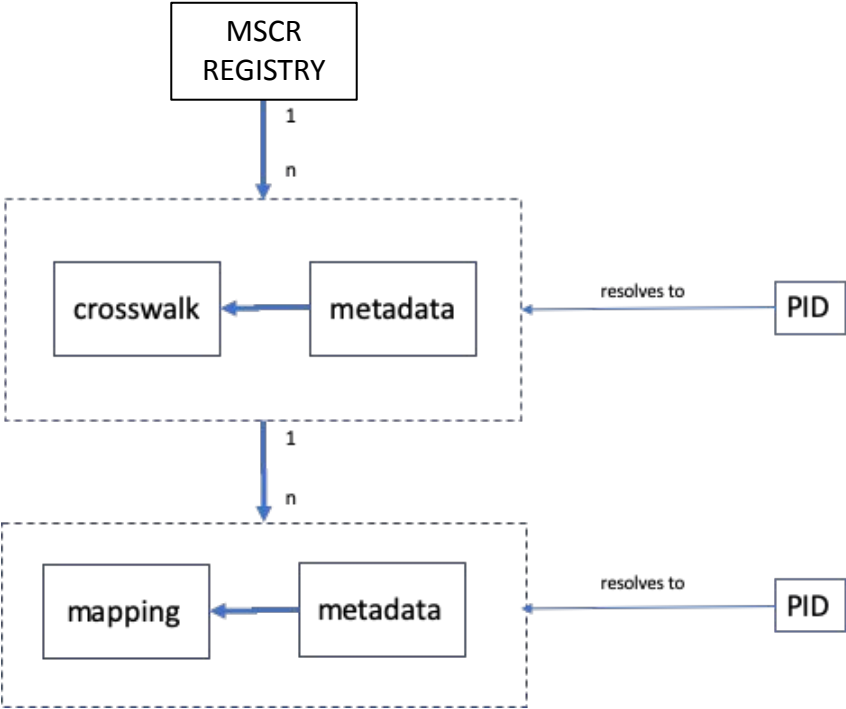
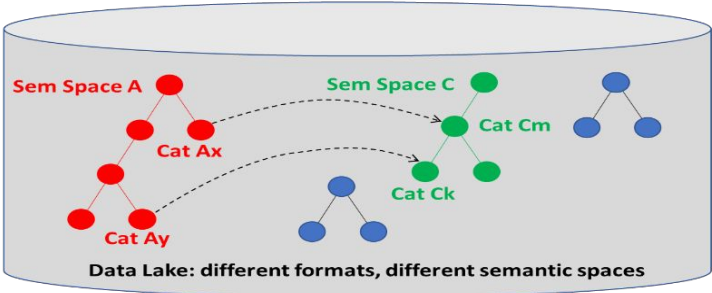
Creating and managing schema and esp. mappings and crosswalks is very labor intensive, existing efforts should be preserved, reused and build upon

For this FC4EOSC builds a Metadata Schema and Crosswalk registry

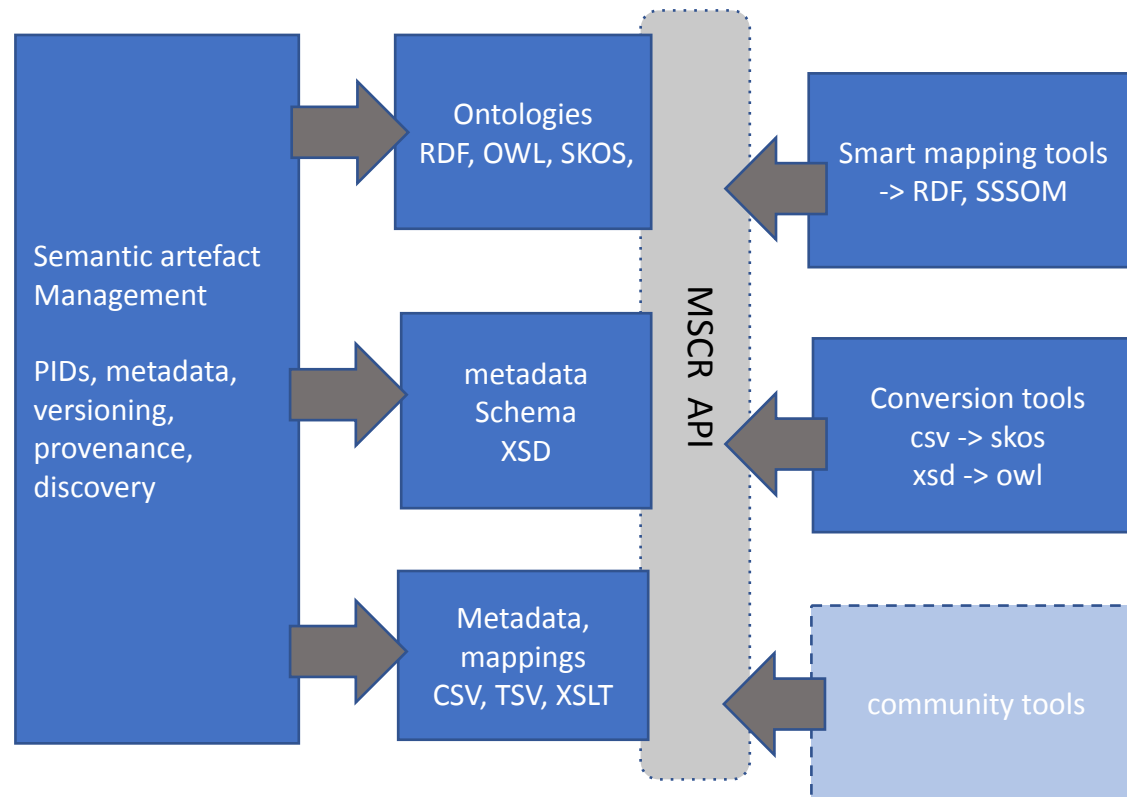
- Registration of elsewhere safely hosted schema, semantic mappings and crosswalks
- Offer basic schema and mapping management: PIDs, metadata, provenance, versioning etc.
- Offer attractive tools to create semantic mappings and crosswalks and publish and share these with colleagues

# Semantic mappings and crosswalks

category	Sem Space	category	Sem Space	Relation Type	metadata	collection
Ax	A	Cm	C	synonym	ref	A2C
Ay	A	Cm	C	part_of	ref	A2C



# MSCR - managing semantic artefacts



Basic data management 4ALL artefacts: PID issuing, core metadata, versioning, provenance info already increases FAIRness

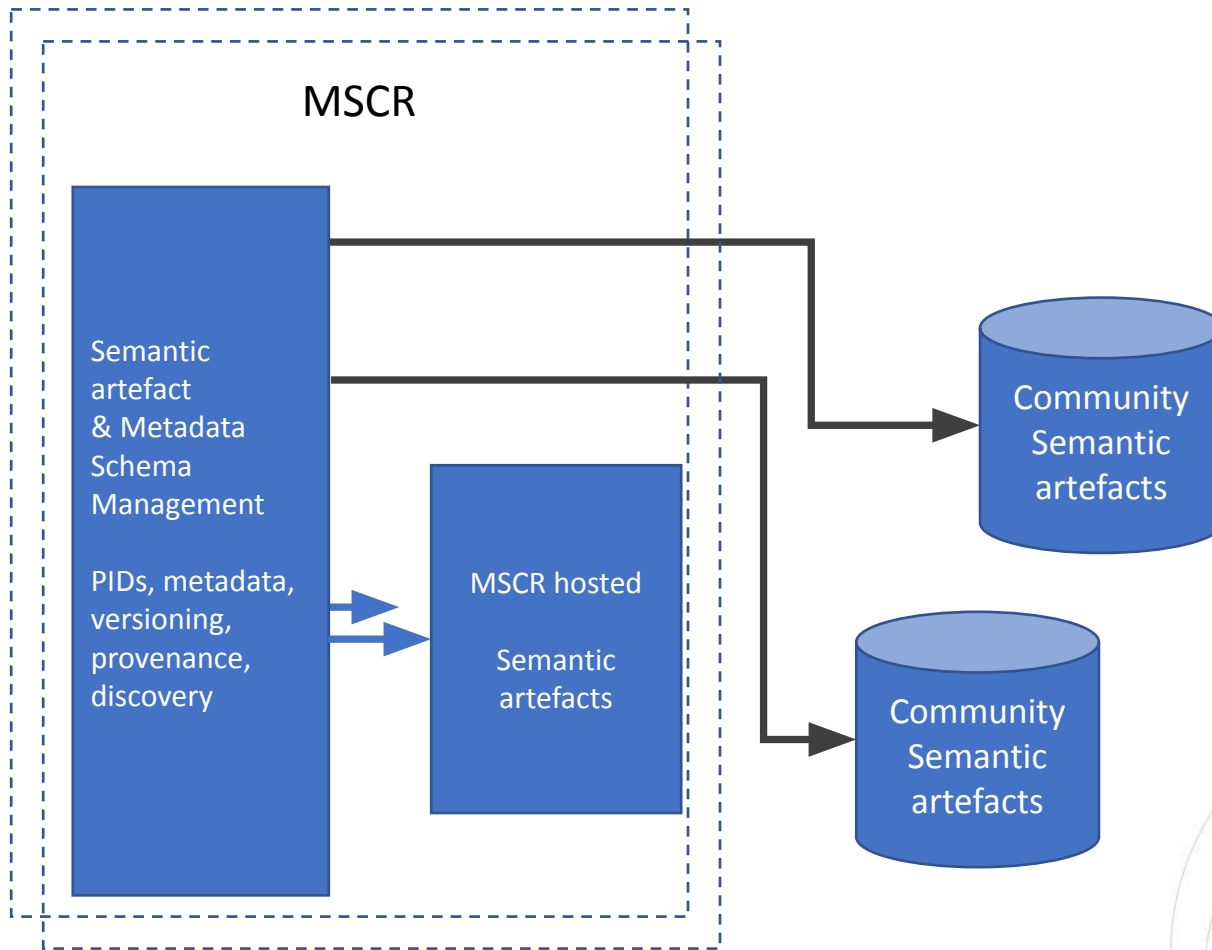
Both hosting and registration of artefacts for communities already managing their own hosting

Offer benefits for 'upgrading' weak semantic formats to state-of-the-art ones by smart tools, not force it

Offer access to the artefacts via open API inviting creation of new tools by communities



# Registering and hosting semantic artefacts



Both hosting and registration of artefacts for communities already managing their own hosting

Benefits for registering community schema, vocabularies etc. with basic metadata management

NOTE the SEMAF report suggested enabling potential federating of semantic (mapping) registries



# Data structure correlation

DARIAH Data Modelling Environment - DME – a general data structure specification and correlation tool

The screenshot displays the DARIAH-DE Mapping-Editor interface. At the top, it shows the DARIAH-DE logo and the title 'Mapping-Editor'. Below the header, the source model is identified as 'goerep\_ksw' and the target model as 'MMM'. The interface is divided into two main panels: 'Beispieltransformation' on the left and 'Elementmodell' on the right. The 'Beispieltransformation' panel shows a tree view of the source model's structure, including elements like 'Spatial', 'TechnicalInformations', 'Format', 'Type', 'Subtype', 'Identifizier', 'Mww', 'External', 'LicenceInformations', and 'Rights'. The 'Elementmodell' panel shows a hierarchical tree of the target model's structure, including elements like 'DOKUMENTENTYP', 'EMPFANGSORT', 'GBNR', 'GNDID', 'IDENT', 'INCIPT', 'INDEXDATUM', 'Jahr', 'Monat', 'Role', 'Reference', 'Forename', 'Surname', 'Source', 'Type', 'Reference', 'Extent', and 'Date'. A complex network of blue lines connects the elements between the two models, with yellow arrows indicating the direction of the mapping. At the bottom of the interface, there is a log showing successful transformation results for three example inputs.

© DARIAH-DE [Datenschutz](#) [Impressum](#) [Kontakt](#)

# Vocabulary / Ontology alignment

Several vocabulary mapping tools available:

- Cocoda (gbv)
- Vocabulary matching tool - VMT (Ariadne project by Univ. South Wales)
- Ontology Xref service - Oxo (EBI)

The screenshot displays the Cocoda Mapping Tool interface. The main window is titled 'Cocoda Mapping Tool' and features a navigation bar with options like 'Imprint', 'Accessibility', 'Privacy Policy', 'Feedback', 'Manual', 'Account', and 'Settings'. The central area is labeled 'Mapping Editor' and shows a mapping between 'LCSH' (source) and 'DFG' (target). The source is 'Avionics' and the target is 'Avionics', with the status 'no concordance'. Below this, there is a search bar and a 'Search' button. A table lists various mapping sources: 'Local' (no results), 'Concordance Registry' (no results), 'Wikidata-Mappings' (no results), and 'KonSys Mappings' (no results). The 'Mapping Details' section provides metadata for the mapping, including the source scheme (LCSH), target scheme (DFG), mapping type (exact match), creation and modification dates, and a unique identifier.

# Disciplinary Vocabulary and Ontology platforms

Many existing platforms for registering vocabularies or ontologies or both but discipline specific:

Many localized versions of mature software packages:  
SKOSMOS, BioPortal



**Term info**  
**DOI:162** (cancer)  
 URI: <http://purl.obolibrary.org/obo/DOI:162>  
 Datasource: Human Disease Ontology Development  
 Type: ONTOLOGY  
[View in OLS](#)

Mapping Distance

1 2 3

Mapped id

- SNOMEDCT:188482002
- UMLS:C0006826
- SNOMEDCT:154577008
- SNOMEDCT:187597000
- ICD9CM:199

## Browse

Browse the library of semantic resources

Showing 25 of 25 Sort: Popular

**Publish New Semantic Resource**

**Entry Type**  
 **Ontology** (25)  
 **Ontology View** (0)

**Uploaded in the Last**

**Category**

- Aquatic Biodiversity (4)
- Aquatic Ecology (4)
- Biodiversity (11)
- Biodiversity Conservation (1)
- Earth Sciences (3)
- Ecology (5)
- Ecosystem Diversity (0)
- Environmental Sciences (5)

**EuroSciVoc (EUROSCIVOC)** concepts 10,439  
 European Science Vocabulary (EuroSciVoc) is the taxonomy of fields of science based on OECD's 2015 Frascati Manual taxonomy  
 Uploaded: 12/17/21

**Endemisms Thesaurus (ENDEMISMS)** concepts 15  
 Thesaurus on endemisms  
 Uploaded: 5/27/19

**Darwin Core Degree of Establishment Controlled Vocabulary (DWCDOE)** concepts 24  
 The Darwin Core term degreeOfEstablishment provides information about degree to which an Organism survives, reproduces, and expands its range at the given place and time  
 Uploaded: 1/4/22



## AGROVOC Multilingual Thesaurus

Alphabetical **Hierarchy**

A Á Â Ã Ä Å Æ Ç È É Ê Ë Ì Í Î Ï Ñ Ò Ó Ô Õ Ö × Ø Ù Ú Û Ü Ý Þ ß à á â ã

A horizons  
 Aaptosyax grypus  
 Aaron's rod → Verbascum  
 ABA  
 abaca  
 abachi → Triplochiton scleroxylon  
 Abalistes stellaris  
 abalone culture  
 abalone fisheries → gastropod fisheries  
 Abalones  
 abamectin  
 abandoned land  
 abattoir by-products  
 abattoirs  
 Abbottina rivularis  
 abbreviations  
 abdomen  
 abdominal cavity  
 abdominal fat  
 abdominal pregnancy  
 Abelmoschus  
 Abelmoschus esculentus  
 Abelmoschus moschatus  
 Abergelle goat  
 Aberia → Dovyalis  
 Abies  
 Abies alba  
 Abies amabilis  
 Abies balsamea  
 Abies balsamea lasiocarpa → Abies lasiocarpa  
 Abies borisii regis  
 Abies cephalonica

## Vocabulary information

TITLE	AGROVOC Multilingual Thesaurus
LAST MODIFIED	Thursday, November 3, 2022 08:24:42
TYPE	<a href="http://www.w3.org/2004/02/skos/core#ConceptScheme">http://www.w3.org/2004/02/skos/core#ConceptScheme</a>
VOID:INDATASET	<a href="http://aims.fao.org/aos/agrovoc/void.ttl#Agrovoc">http://aims.fao.org/aos/agrovoc/void.ttl#Agrovoc</a>
URI	<a href="http://aims.fao.org/aos/agrovoc">http://aims.fao.org/aos/agrovoc</a>

### Resource counts by type

Type	Count
Concept	40615

### Term counts by language

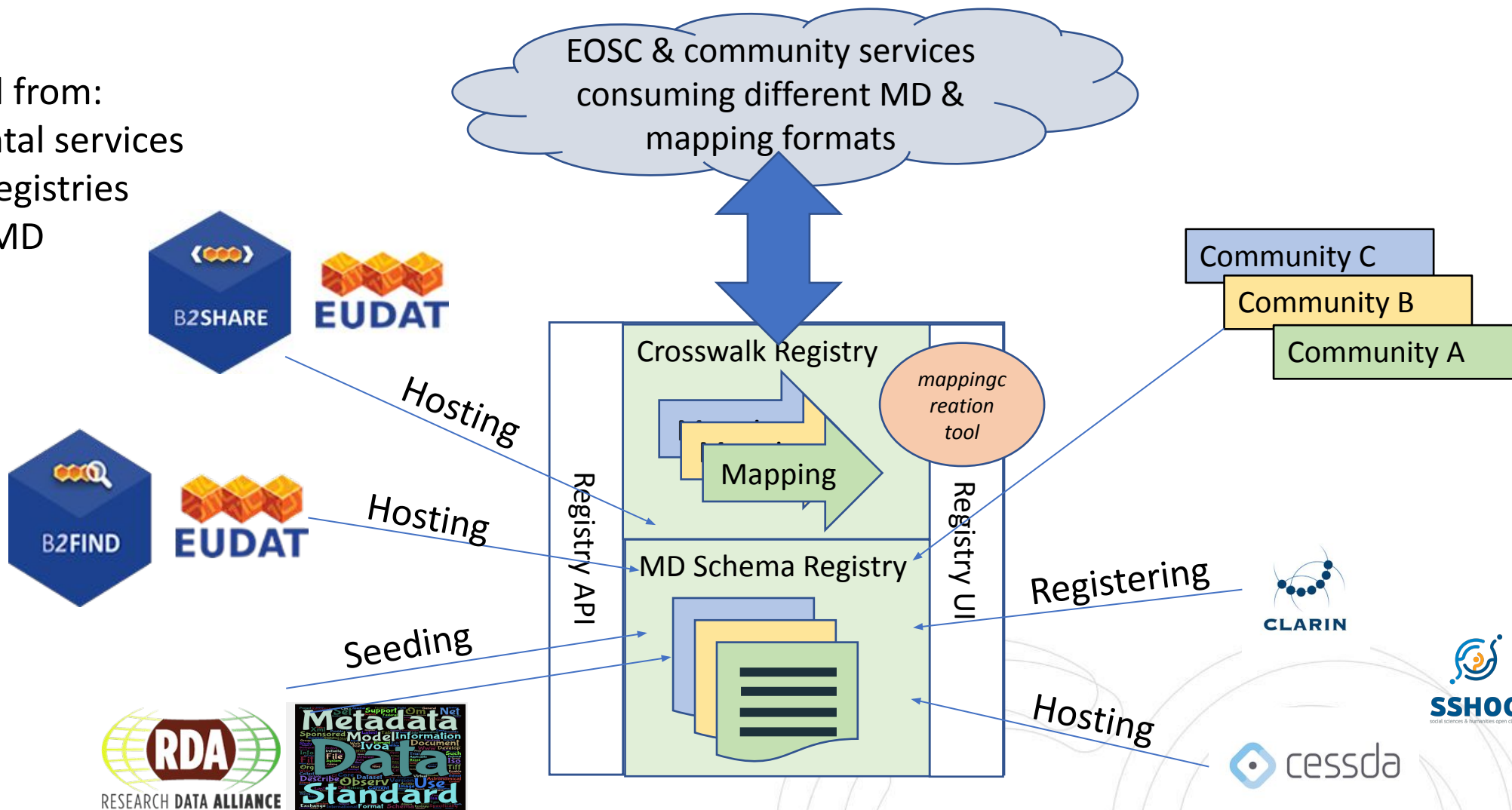
Language	Preferred terms	Alternate terms	Hidden terms
Arabic	35789	1387	0
Catalan	369	4	0
Czech	37099	8759	0
Danish	529	7	0
German	38443	7822	0
Greek	181	2	0
English	40558	11636	0
Spanish	38138	11427	0
Estonian	322	4	0

# Summary

- Offer registration & hosting off a broad class of semantic artefact formats: allow tables formats, xsd,
- Offer essential management of the hosted semantic artefacts: PID, metadata, versioning, provenance
- Publish and share semantic artefacts in and between communities allowing further updates and improvements
- Semantic mappings and crosswalks to be first class citizens eg. proper metadata, PID, etc.
- Offer smart tools for creating semantic mappings and crosswalks, motivating users to upgrade to better representations
- Target audiences:
  - Research communities, projects and individual researchers
  - Institutions & organizations interested in creating and maintaining semantic mappings and crosswalks for their internal processes: research related as for B2FIND, CRIS, ... but other domains are conceivable

MSCR populated from:

- EOSC horizontal services
- Established registries
- Community MD schemas



[faircore4eosc.eu](https://faircore4eosc.eu)

Twitter: @FAIRCORE4EOSC

LinkedIn: [company/faircore4eosc](https://www.linkedin.com/company/faircore4eosc)

Youtube: FAIRCORE4EOSC



**Funded by  
the European Union**



**Discussion:**  
**Does semantic interoperability get enough attention?**





# Some thoughts on semantic artefacts & interoperability

There are many registries of semantic artefacts , metadata schema, some crosswalks both general and discipline specific, different levels of maturity, different governance & sustainability expectations, different missions... but also overlap

In general choice is a good thing , but does there exist, or should there be some coordination towards collaboration and interoperability of the interoperability providers?

Otherwise, is there anything obvious missing from the landscape wrt semantic interoperability?

....



[faircore4eosc.eu](https://faircore4eosc.eu)

Twitter: @FAIRCORE4EOSC

LinkedIn: [company/faircore4eosc](https://www.linkedin.com/company/faircore4eosc)

Youtube: FAIRCORE4EOSC



**Funded by  
the European Union**



# Semantic Interoperability – a big challenge

- Science is for a large part based on measuring and describing phenomena using schemas and concepts that are often discipline specific
- For Open Science, sharing and transparency these schemas and concept/vocabulary definitions need to be as FAIR (open registries) as the data itself
- When integrating data from different communities – also mappings are required!
  - spending many resources & time on deep ontological work does not make sense
  - mapping needs to be driven by concrete data cases and purposes only a pragmatic approach will help
  - researchers already do this, but mappings are hidden in texts, software, spreadsheets, etc. thus, they are not explicit, reusable (not FAIR)

Therefore propose a flexible Semantic Mapping Framework which should be FAIR and persistent