FAIRCORE4EOSC

Added value for research communities and other stakeholders

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
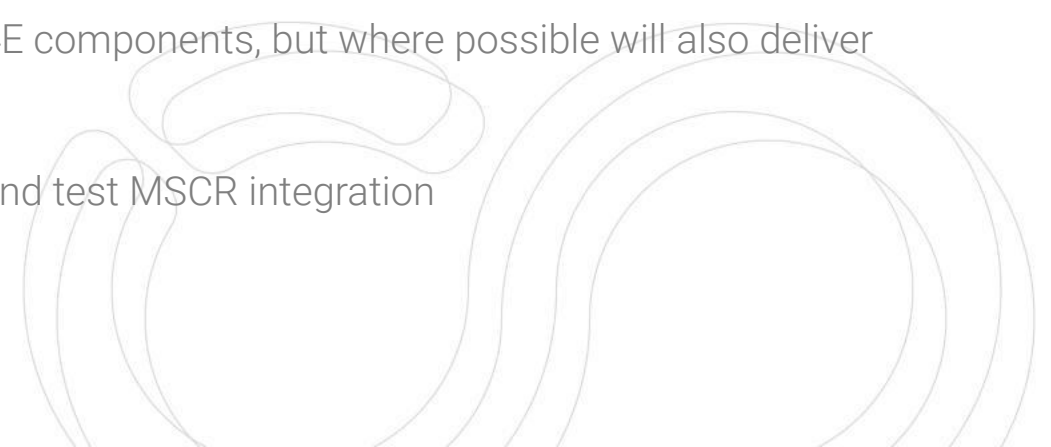
Examples of semantic mappings

**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. 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
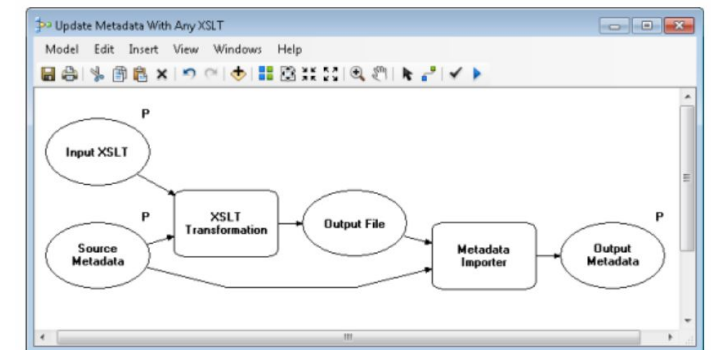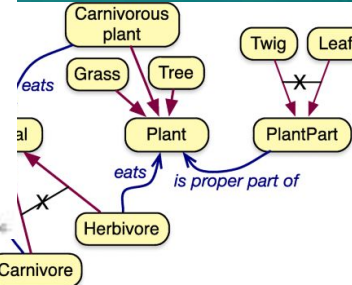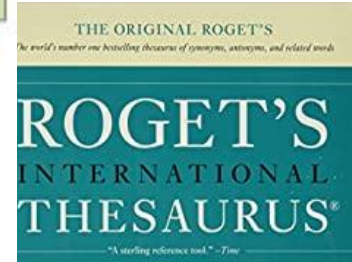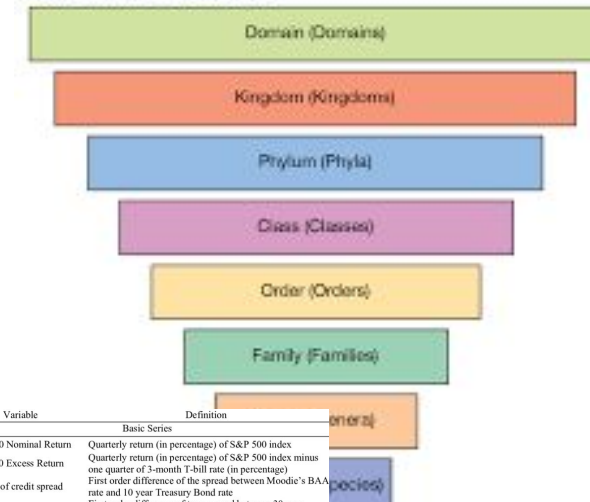
Semantic artefacts

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

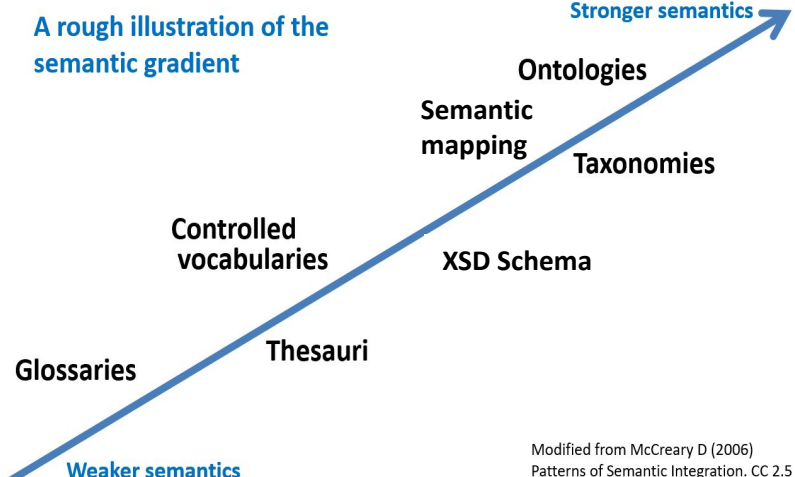W3C std. RDF, SKOS, OWL: widely accepted and supported formal way to represent vocabularies

A rough illustration of the semantic gradient

Stronger semantics

Ontologies

Semantic mapping

Taxonomies

Controlled vocabularies

XSD Schema

Glossaries

Thesauri

Weaker semantics

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

Data Lake: different formats, different semantic spaces
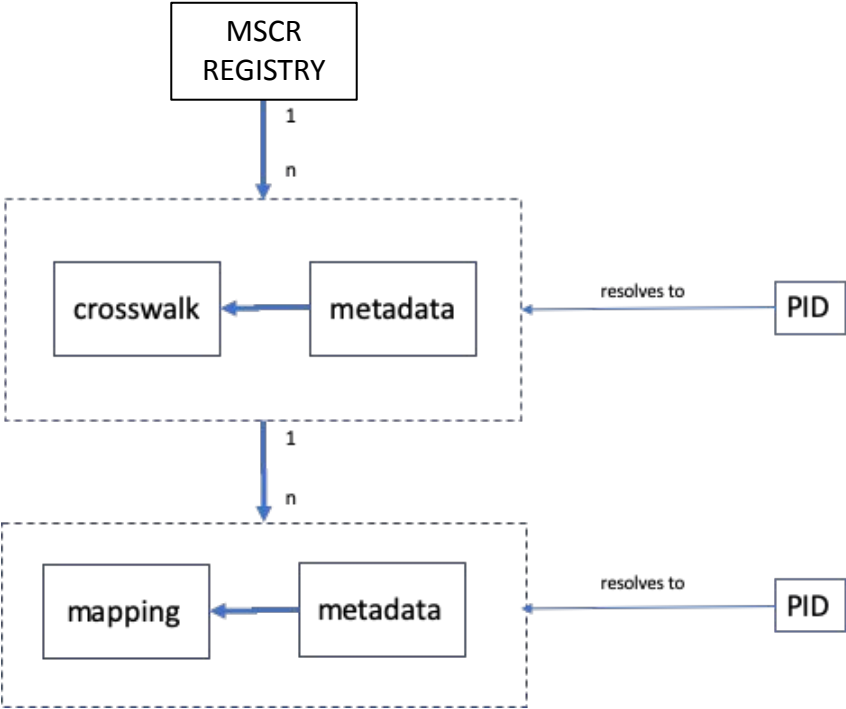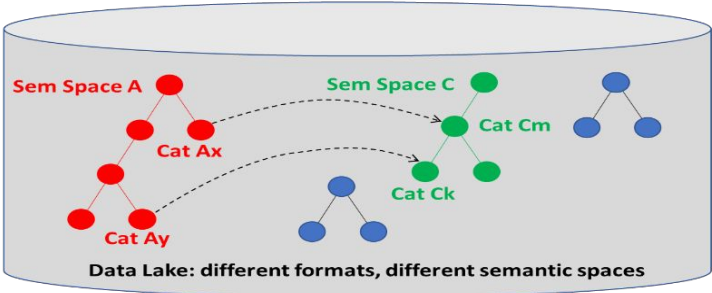
MSCR - managing semantic artefacts
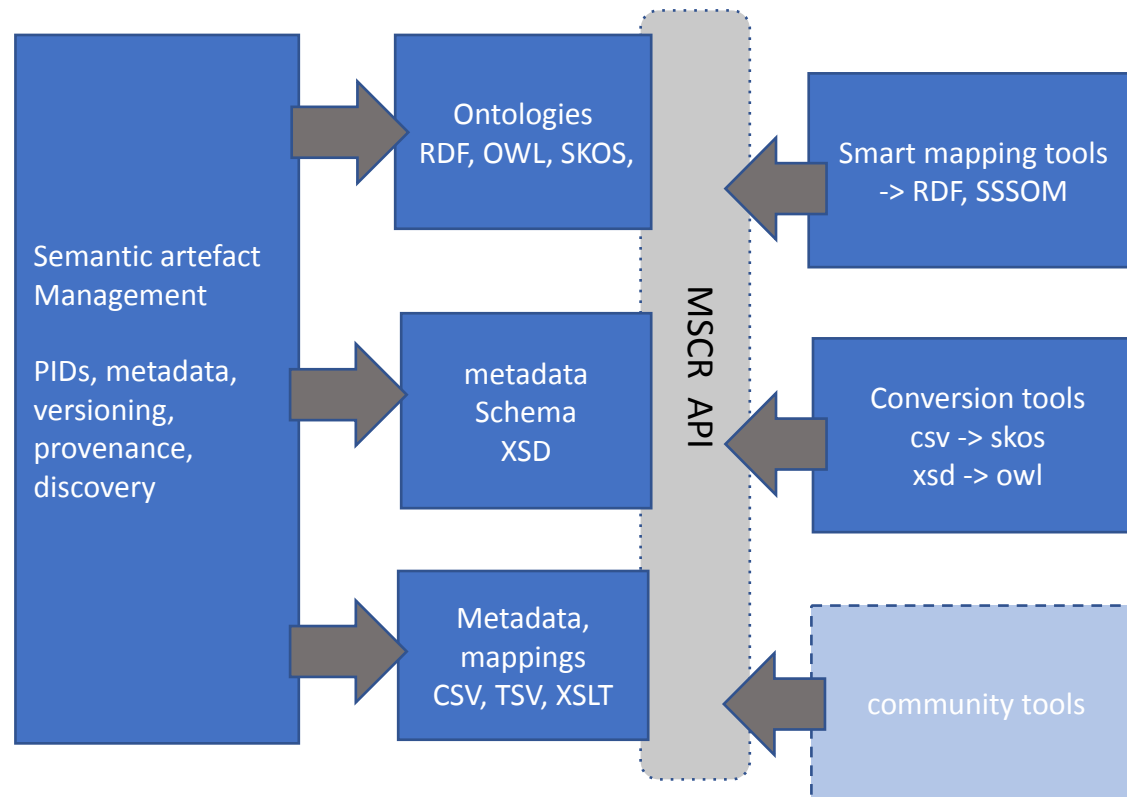
Basic data management 4ALL artefacts: PID issuing, core metadata, versioning, provenance info already increases FAIRness

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
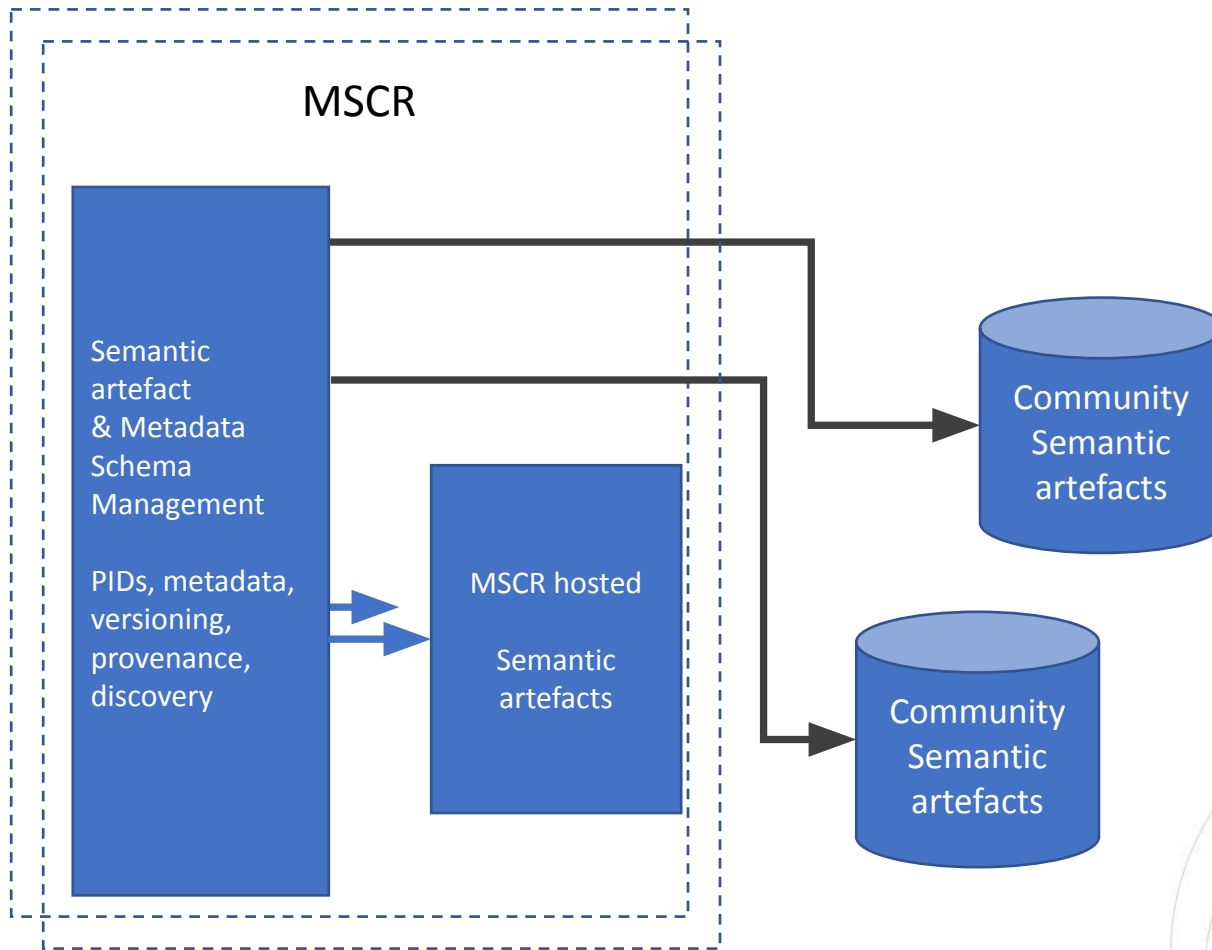
MSCR

Semantic
artefact
& Metadata
Schema
Management

PIDs, metadata,
versioning,
provenance,
discovery

MSCR hosted

Semantic
artefacts

Community
Semantic
artefacts

Community
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

8

Data structure correlation

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

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)

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

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
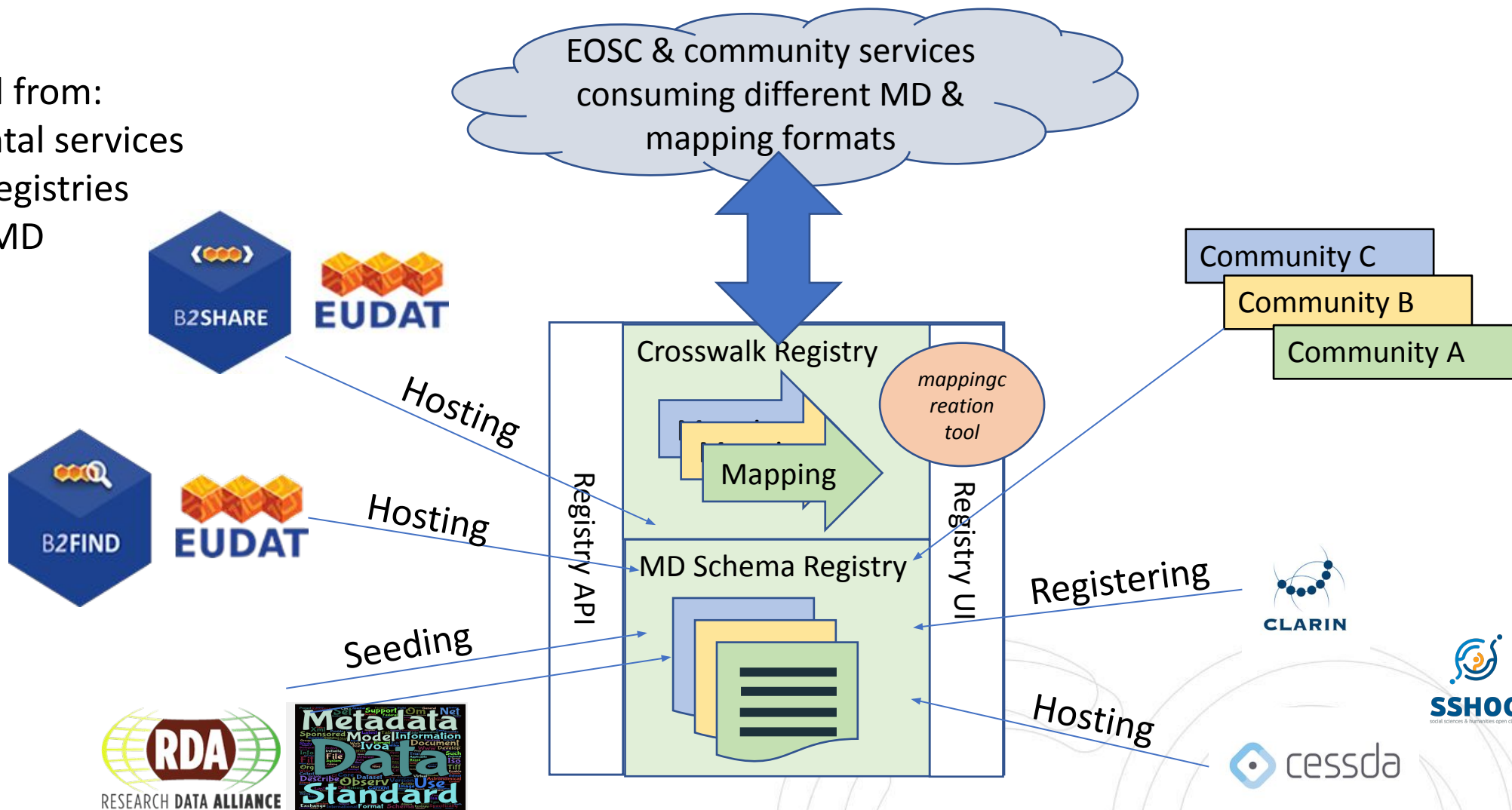
EOSC & community services consuming different MD & mapping formats

Crosswalk Registry

mapping creation tool

Mapping

Registry API

Registry UI

MD Schema Registry

Community C
Community B
Community A

Hosting

Hosting

Seeding

Registering

Hosting

eosc | **FAIRCORE4EOSC**
Core Components Supporting a FAIR EOSC

faircore4eosc.eu

Twitter: @FAIRCORE4EOSC

LinkedIn: company/faircore4eosc

Youtube: FAIRCORE4EOSC

**Funded by
the European Union**

Discussion:
Does semantic interoperability get enough attention?

Funded by
the European Union

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 sematic interoperability?

….

eosc | **FAIRCORE4EOSC**
Core Components Supporting a FAIR EOSC

faircore4eosc.eu

Twitter: @FAIRCORE4EOSC

LinkedIn: company/faircore4eosc

Youtube: FAIRCORE4EOSC

**Funded by
the European Union**

17

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