

SEMAF

Flexible Semantic Mapping Framework

Daan Broeder – CLARIN ERIC
e-IRG 25-5-2021

Semantic Interoperability – a big challenge

- Science 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 a flexible Semantic Mapping Framework which should be FAIR and persistent (it includes part of our scientific knowledge to be captured)

Examples of semantic mapping

Entity 1	Entity 2
tectonic movement(ENVO:01001093)	Continental drift (SWEETPhenGeoTectonic: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)

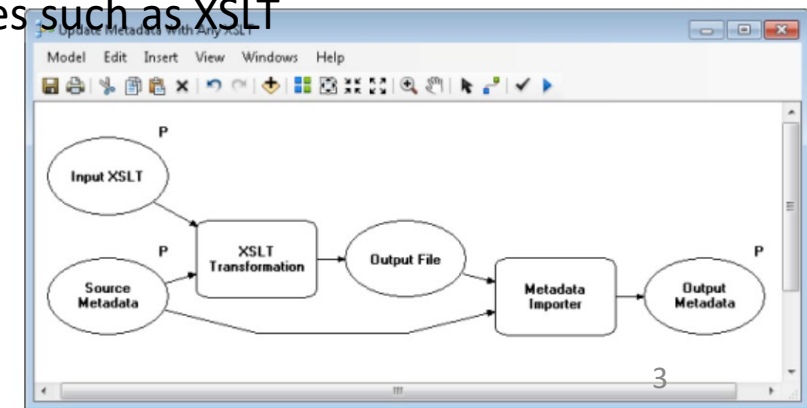
Mapping between Darwin Core 1.4 concepts (DwC) and ABCD 2.0

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 r33 dated 17 Feb 2007 - 19:54:25 under <http://wiki.tdwg.org/wiki/bin/view/DarwinCore/DarwinCoreDraftStandard>. Deprecated DwC elements 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

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 mapping specifications and conversion are combined in technologies such as XSLT



SEMAF – An EOSC co-creation project

- Collaboration Group European Data Experts (GEDE) and CLARIN ERIC to address inter- and cross-discipline semantic interoperability
- Goal to formulate a **”Proposal for a pragmatic flexible Semantic Mapping Framework”**
- SEMAF expert task-force conducted 25 interviews with community experts
- SEMAF final report delivered March 31 2021 proposes a 3-year follow-up project

Broeder, Daan, Budroni, Paolo, Degl'Innocenti, Emiliano, Le Franc, Yann, Hugo, Wim, Jeffery, Keith, Wiland, Claus, Wittenburg, Peter, Zwolf, Carlo Maria. (2021, March 31). SEMAF: A Proposal for a Flexible Semantic Mapping Framework (Version 1.0). Zenodo. <http://doi.org/10.5281/zenodo.4651421>

25 Interviewed experts & communities

Julian Richards	Director of the Archaeology Data Service	Humanities, Archaeology
Christian Ohmann, Steve Canham	ECRIN	Biomedical sciences
Ingemar Häggström, Carl-Fredrik Enell	EISCAT Scientific Association	Natural Science, Environmental Science
Alexandra Kokkinaki	BODC	Environmental Science, Oceanography
John Watkins	UK Centre for Ecology and Hydrology	Environmental Science, ecology, hydrology
Menzo Windhouwer	KNAW/HuC, CLARIN ERIC	Humanities, linguistics
Wolfgang Schmidle	DAI, Data scientist	Humanities, Archaeology
Johan Fihn Marberg	SND, CTO	General research data management, social sciences
Matej Durco	OEAW	Humanities
Dieter van Uytvanck	CLARIN ERIC, CTO	Humanities, linguistics
Baptiste Cecconi	OBSPM	Planetary science
Mathias Dillen	DiSSCo, MBG	Biodiversity, Environmental Science
David Fichtmüller	DiSSCo, BGBM	Biodiversity, Environmental Science
Carsten Thiel	CESSDA ERIC, CTO	Social Sciences
Herve L'Hours	UKDA	Social Sciences
Tobias Gradl	University of Bamberg	Research Data Management, Digital Humanities
Daniel Heydebreck, Anna-Lena Flügel, Claudia Martens	DKRZ	Climate science, general research data management
Ilaria Rosati, Nicola Fiore, Lucia Vaira, Pierfrancesco Tommasino	LifeWatch-ERIC, National Research Council of Italy	Biodiversity, Environmental Science
Dr. Helen Parkinson (Head of Molecular Archival Resources)	EMBL-EBI	Biomedical sciences, Bioinformatics
Margareta Hellström, Oleg Mirzov	ICOS Carbon Portal and Lund University	Environmental Science
Federica Spinelli, Alessia Spadi	RESTORE project, OVI, National Research Council of Italy	Humanities
Claudia Caliri	ISPC, National Research Council of Italy	Cultural heritage science
Carsten Baldauf	Nomad Project, FHI	Material Sciences
Lara Ferrighi	Data Manager, Norwegian Meteorological Institute	Meteorology
Chris Schubert	Head CCCA data center	Climate sciences

SEMAF Findings and Goals

Current mapping practices are fragmented, often hidden, insufficiently documented, inflexible.

Therefore SEMAF proposes a framework

- to register and host semantic mappings with proper metadata and provenance in **FAIR SEMAF registries** and
- some easy-to-use GUI tools for end-users to **create new mappings and reuse and extend** existing ones

leading principles should be:

- **pragmatic**: targeted at specific project data driven interoperability goals
- **flexible and inclusive**: integrate existing semantic components (making them FAIR) – don't enforce specific technologies but promote by offering attractive tools

SEMAF tool inspiration

OXO ONTOLOGY MAPPING

Home | Documentation | About

Term info
DOID:162 (cancer)
 URI: http://purl.obolibrary.org/obo/DOID_162
 Datasource: Human Disease Ontology DOID
 Type: ONTOLOGY

Network

Mapping Distance

Mapped id	Id source	Evidence	Distance
SNOMEDCT:188482002	SNOMEDCT	3	1
UMLS:C0006826	UMLS	7	1
SNOMEDCT:154577008	SNOMEDCT	3	1
SNOMEDCT:187597000	SNOMEDCT	3	1
ICD9CM:199	ICD9CM	7	1

Vocabulary Matching Tool

Source Concept	Label	Lang...	Match Type	Target Concept	Suggest	Delete
Identifier	filter column...			Filter column...		
http://purl.org/heritagedata/sche...	ABBEY	en	Exact Match	abbeys (monasteries)	Q	
http://purl.org/heritagedata/sche...	Abbey Church	en	Close Match	abbey churches	Q	
http://purl.org/heritagedata/sche...	Abbey Wall	en	Broad Match	precincts	Q	
http://purl.org/heritagedata/sche...	AGRICULTURAL BUILDING	en	Exact Match	agricultural buildings	Q	
http://purl.org/heritagedata/sche...	AGRICULTURAL DWELLING	en	Broad Match	agricultural buildings	Q	
http://purl.org/heritagedata/sche...	AGRICULTURAL HALL	en	Broad Match	agricultural buildings	Q	
http://purl.org/heritagedata/sche...	AGRICULTURE AND SUBSISTENCE	en	Broad Match	agriculture	Q	
http://purl.org/heritagedata/sche...	AIR RAID SHELTER	en	Exact Match	air raid shelters	Q	
http://purl.org/heritagedata/sche...	AIRCRAFT	en	Close Match	aircraft	Q	
http://purl.org/heritagedata/sche...	AIRFIELD	en	Exact Match	airfields	Q	
http://purl.org/heritagedata/sche...	AIRFIELD DEFENCE SITE	en	Broad Match	airfields	Q	
http://purl.org/heritagedata/sche...	AISLED BARN	en	Broad Match	barns	Q	
http://purl.org/heritagedata/sche...	AISLED BUILDING	en	Broad Match	hall houses	Q	
http://purl.org/heritagedata/sche...	AISLED HALL HOUSE	en	Close Match	hall houses	Q	
http://purl.org/heritagedata/sche...	Alan Williams Turret	en	Broad Match	field fortifications	Q	
http://purl.org/heritagedata/sche...	ALMSHOUSE	en	Exact Match	almshouses (buildings)	Q	
http://purl.org/heritagedata/sche...	ANCHORAGE (MARITIME)	en	Exact Match	anchorages	Q	
http://purl.org/heritagedata/sche...	ANGLE TOWER	en	Broad Match	defensive towers	Q	

390 rows

EXPORT JSON | EXPORT CSV | + ADD NEW ROW | CLEAR ROWS | SHOW HELP

Language | Login

Created by University of South Wales

This application retrieves some information originating from Getty Art & Architecture Thesaurus (AAT) which is made available under the ODC Attribution License. See <http://vocab.getty.edu/> for further details.

DARIAH-DE Data Modeling Environment (DME)

Mapping-Editor

Data Modeling Environment (DME) / Datenmodelle und Mappings / Mapping-Editor

Quellmodell: goerep_ksw Zielmodell: MMM

Beispieltransformation

Input Ergebnisse

Quellmodell Zielmodell 3/3 < >

M. (Absendeort)
 - Spatial [Neuhof] (Empfangsort)
 - TechnicalInformations
 - Format [text/XML]
 - Type [Handschrift]
 - Subtype [Brief]
 - Identifier
 - Mww [ksw_goerep_10468]
 - External [ksw_goerep_10468]
 - LicenceInformations
 - Rights [Bitte beim Rechteinhaber anfragen]

Elementmodell

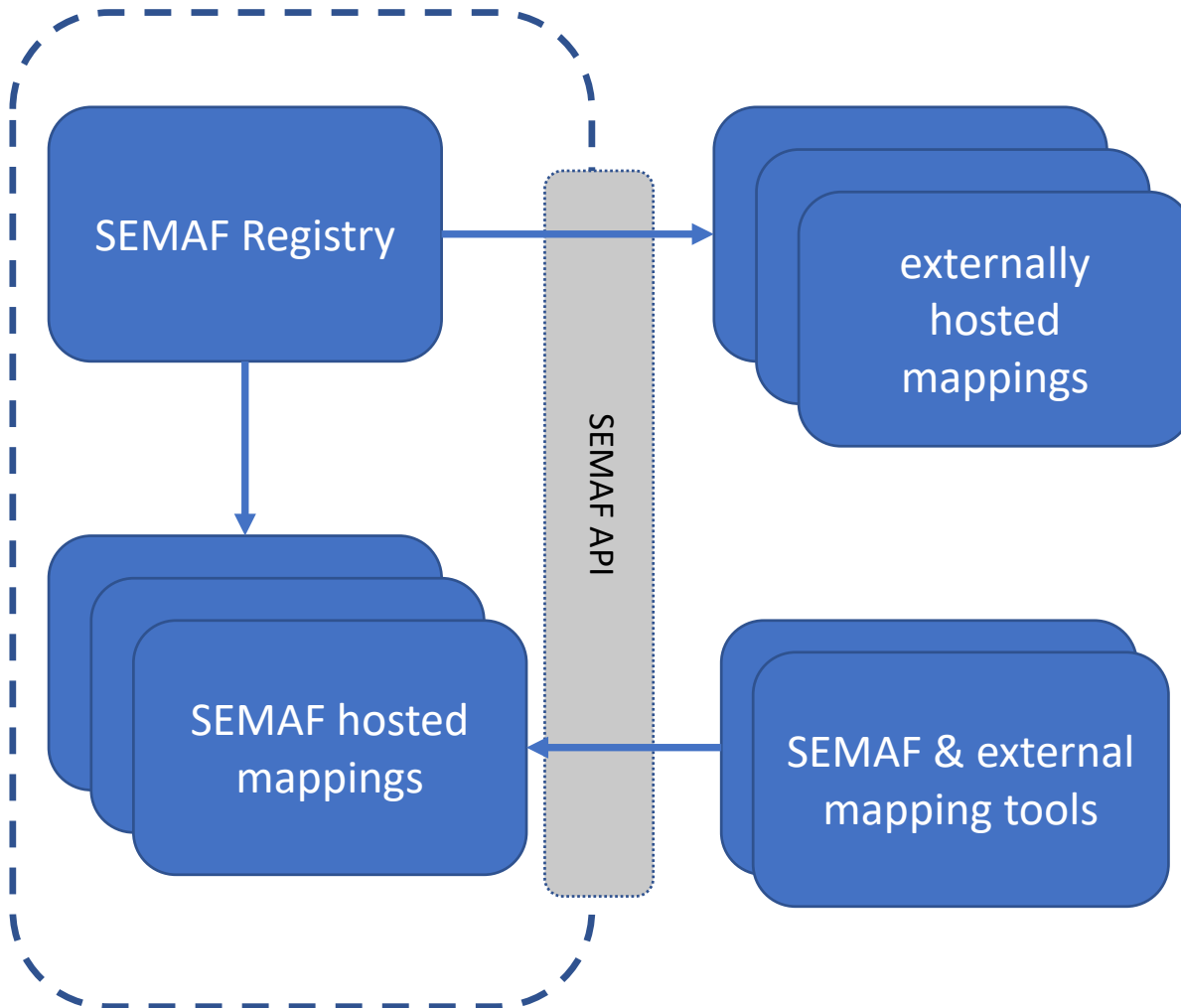
DOKUMENTENTYP
 EMPFANGSORT
 GBNR
 GNDID
 IDENT
 INCIPIT
 INDEXDATUM
 Jahr
 Monat

Role
 Reference
 Forename
 Surname
 Source
 Type
 Reference
 Extent
 Date

2017-08-30 17:01:10.695 SUCCESS Beispielinput transformiert (Dauer 127ms): 3 Ressourcen erkannt
 2017-08-30 17:01:10.526 SUCCESS Beispielinput verarbeitet (Dauer 48ms): 3 Ressourcen erkannt
 2017-08-30 17:01:10.223 SUCCESS Beispielinput transformiert (Dauer 145ms): 3 Ressourcen erkannt

© DARIAH-DE Datenschutz Impressum Kontakt

SEMAF Semantic Mappings Registry

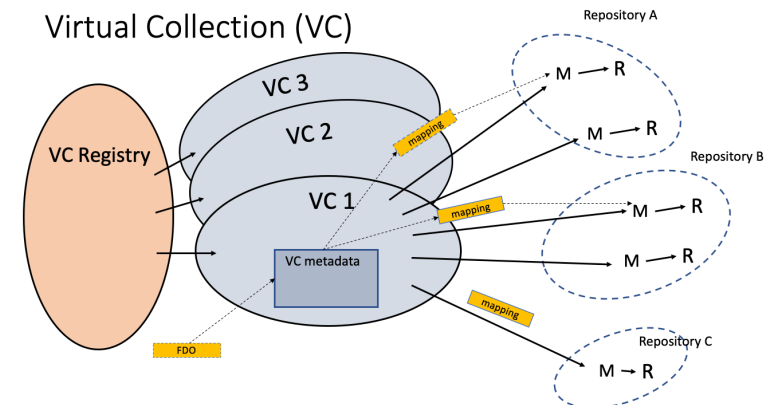
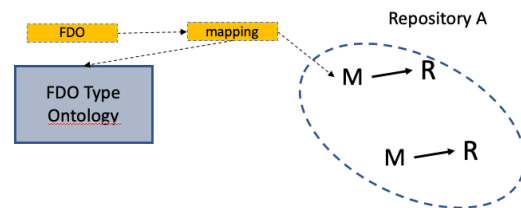


- Registry with proper metadata and provenance **considered very useful by all**
- Registry to be inclusive and support registering external hosted mappings
- Added value of SEMAF hosting should be better provenance tracking and uniform accessibility management
- Potentially support several such SEMAF Registries (federation)

Semantic Mapping and FDOs

- Semantic Mappings are part of our scientific knowledge and should be persistent and reusable FDOs
- FDOs give access to the rich context of mappings
- FDOs need Semantic Mapping/SEMAF
 - community metadata needs to be mapped to widely agreed FDO type ontology
 - persistent registries for the FDO type ontology and for mappings
- Virtual collections of heterogeneous data being used in interdisciplinary research etc. are FDOs – mappings between them need to be linked to the collection

FDO & descriptive metadata



Recommendations

- Flexible pragmatic Semantic Mapping is essential for semantic interoperability, within disciplines and between disciplines
- Semantic mappings are everywhere in data and metadata are part of our scientific knowledge, processes and key for reproducibility
 - need to become FAIR
 - need to be stored, shared, managed

SEMAF proposal essentials:

- Invest in a framework with as basis an open registry specification, an API and a reference implementation to create interoperability
- Invest in smart tools created by smart young developers that make use of specs and offer high usability
- SEMAF integrated in EOSC for sustainability and governance

SEMAF requirements and construction proposal available in the final report at <http://doi.org/10.5281/zenodo.4651421>

Thank you!

Please see also the SEMAF final report

<http://doi.org/10.5281/zenodo.4651421>