

CRIS systems integration as a case study for the Metadata Schema and Crosswalk Registry

CRIS2024

15.5.2024

Joonas Kesäniemi, Tommi Suominen, Dragan Ivanović and Jan Dvořák



Funded by
the European Union



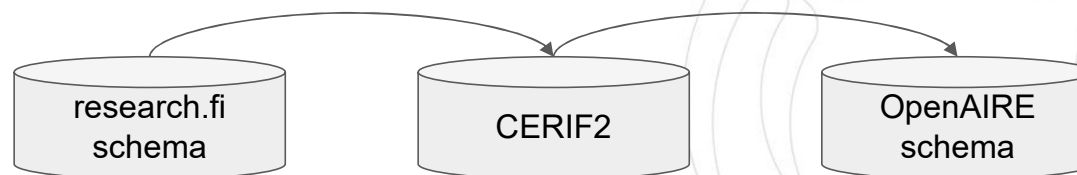
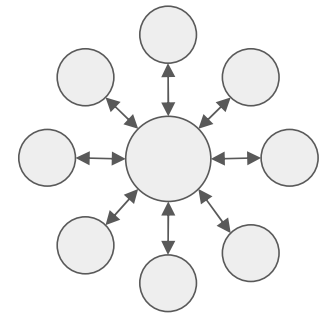
Agenda

- Motivation
- Crosswalk between research.fi and CERIF2
 - Schemas
 - Spreadsheet based mappings
 - Status
- Metadata Schema and Crosswalk Registry (MSCR)
 - Introduction
 - MSCR based mappings
 - Status



Motivation

- Case: Transfer data from different kinds of CRIS systems to OpenAIRE
 - research.fi (Finland) - focus of this presentation
 - Helix (Greece)
 - Publinova (Netherlands)
- Using CERIF2 as a core model
 - research.fi → CERIF2
 - Helix → CERIF2
 - CERIF2 → [OpenAIRE guidelines for CRIS managers?](#)
 - CERIF2 → SKG-IF?
- CERIF2 is comprehensive enough to support the full CRIS domain





FAIRCORE4EOSC
Core Components Supporting a FAIR EOSC

Schemas



**Funded by
the European Union**



Research.fi

The Finnish national CRIS

- Provides a landscape of Finnish research ecosystem and basis for science policy
- Supports the re-use of research information
- Merit researchers and make their research outputs and activities more visible

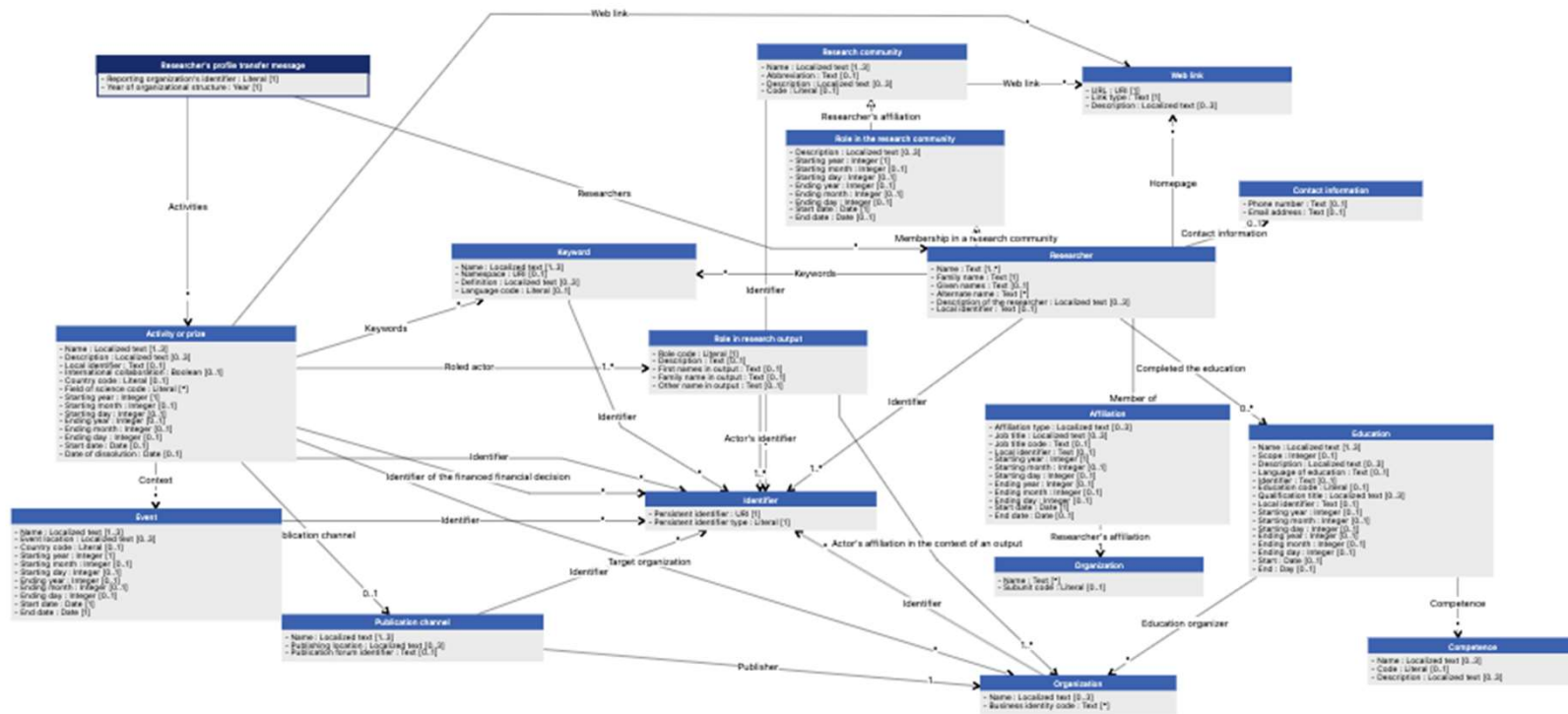


Research.fi schema

- [Researcher's profile transfer message](#)
 - Data exchange format for researcher related data
- Subset of the internal conceptual model that covers all aspects of research.fi data
- Serialized as XML Schema
- Hosted and maintained using the Finnish Interoperability Platform



Researcher's profile transfer message visualization



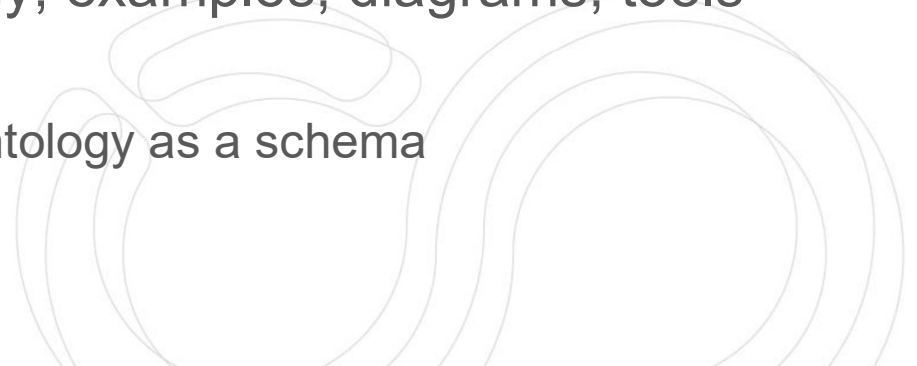
XML excerpt

```
<activity>
  <name lang="en">Test activity</name>
  <description></description>
  <countryCode>en</countryCode>
  <roledActor>
    <roleCode>11.1.1</roleCode>
    <description></description>
    <firstNameInOutput></firstNameInOutput>
    <familyNameInOutput></familyNameInOutput>
    <otherNamesInOutput></otherNamesInOutput>
    <actorsIdentifier>
      <identifier>0000-1111-2222-3333</identifier>
      <identifierType>Orcid</identifierType>
    </actorsIdentifier>
    <actorsAffiliationInTheContextOfAnOutput>
      <organization>
        <name></name>
        <businessIdentityCode></businessIdentityCode>
        <identifier>
          <identifier>0000-1111-2222-3333</identifier>
          <identifierType>Orcid</identifierType>
        </identifier>
      </organization>
    </actorsAffiliationInTheContextOfAnOutput>
  </roledActor>
</activity>
```



CERIF2

- The new CERIF has been designed to take away the perceived complexity of CERIF, involve the community in further development of CERIF, and modernize serialization formats
- Designed to be extended with modules by e.g. domain and use case specific models
- Consists of entities, data types, ontology, examples, diagrams, tools
- All available on [GitHub](#)
 - MSCR will be using the generated OWL ontology as a schema



CERIF2 status of development

- CERIF Core (common data types, abstract entities, commonly used entities, common constructs)
 - <https://github.com/EuroCRIS/CERIF-Core>
- CERIF Scholarly Publication module
 - <https://github.com/EuroCRIS/CERIF-ScholarlyPublicationsModule>
- A tool for generating ontology from the conceptual description of entities (Markdown files)
 - <https://github.com/EuroCRIS/CERIF-Core/blob/main/serializations/RDF/core.owl>

[Code](#)[Issues](#) 16[Pull requests](#)[Discussions](#)[Actions](#)[Projects](#)[Wiki](#)

CERIF-Core

Public

[Edit Pins](#)[Unwatch](#) 5[Fork](#) 3[Star](#) 5[main](#)[Go to file](#)[Code](#)[About](#)

jdvorak001
and chenejac

The c...



b994b2e · 3 weeks ago

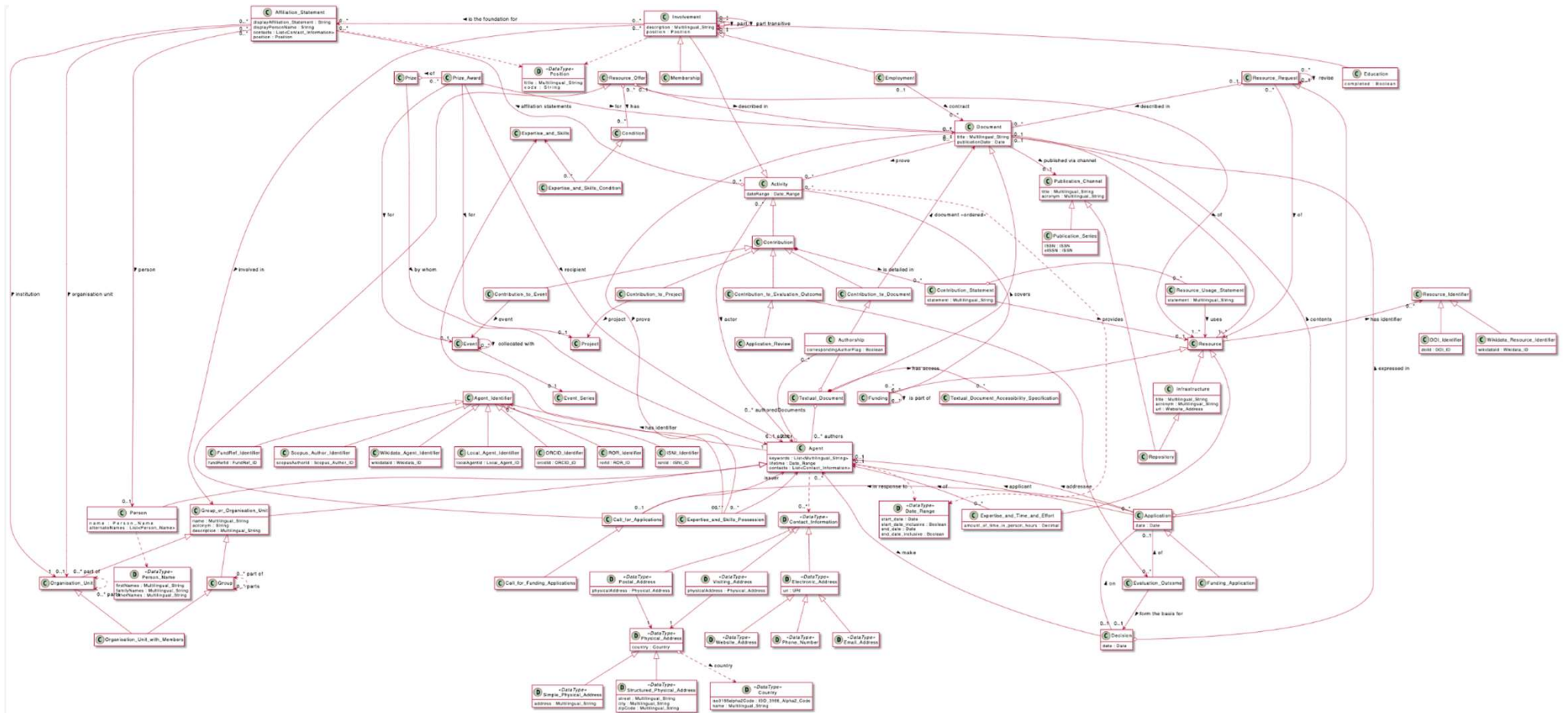
[514 Commits](#)

.github/workflows	Run 'generate plantu...	7 months ago
datatypes	Correct the relationsh...	3 months ago
diagrams	adding generated ...	3 weeks ago
entities	Standardize also the ...	3 weeks ago

The CERIF Refactoring Pilot project
by euroCRIS.

[Readme](#)[Activity](#)[Custom properties](#)[5 stars](#)[5 watching](#)[3 forks](#)[Report repository](#)

CERIF2-core visualization



Mapping Researcher's profile transfer message to CERIF2



Funded by
the European Union



Mapping spreadsheet

Mapping between conceptual models

Research.fi					CERIF	
class	property	data type	cardinality	description	class	property
Role in research output	Role code	Text	1..1		CRedit Contribution Statement	Role type based on mapping from Role code
Role in research output	Description	Text	0..1		Contribution Statement	statement
Role in research output	First names in output	Text	0..1		Affiliation Statement	displayPersonName
Role in research output	Family name in output	Text	0..1		Affiliation Statement	displayPersonName
Role in research output	Other name in output	Text	0..1		Affiliation Statement	displayPersonName

Mapping logic is included as a comment



Status of mapping work

- Analysis of Research.fi
 - Done
- Mapping of Research.fi data model elements to CERIF elements
 - Agents
 - Done
 - Publications, Research Dataset, Channels
 - Partially done, some improvements of CERIF needed
 - Funding, Infrastructures
 - Postponed for later due to pending changes in the research.fi models
- Operational mapping
 - Started in the MSCR crosswalk tool



Metadata Schema and Crosswalk Registry



Funded by
the European Union



Metadata Schema and Crosswalk Registry

- Allows for registration of existing schemas and crosswalks in multiple formats
- Includes a tool for creating crosswalks between registered schemas
- Provides a harmonized view to schemas expressed in different formats
- Facilitates the operationalization of crosswalks
- Handles versioning and assigns handles

MSCR glossary

Schema

Structural definition of data (cf. Ontology and complex data types). Should be owned or maintained by someone or something. Consists of properties

Vocabulary schema

List of things. Can be hierarchical or flat.

Property

Data typed. Can be simple or complex. Has cardinality.

Mapping

Description of a link between two property sets. Mainly manually curated (cf. automated mapping)

Crosswalk

Set of mappings. Binds together two registered schemas

Personal workspace

Schemas

Crosswalks

Group workspaces

MSCR testers

Test org

CERIF2

Metadata Schema

MSCR allows registered users and co-facilitate the transformation of data

+ Register Schema

Schemas

Name

Status 1

EUDAT Core

Codemeta test

CERIF2 -test

DataCite

Status 2

OKM research fields

OPENAIRE test

MSCR internal metadata schema

Simple DC XML schema, version 20

MSCR internal metadata schema - M

OpenAIRE Guidelines for Literature

Register Schema

You must provide a schema file to register a schema. Supported schema file formats are JSON, CSV, TTL, PDF and XSD.

Provide schema file URI reference

OR

Upload documents



Drag files to upload
OR
[Browse files](#)

Format

CSV

Name (required)

Description

Status

Select status

Register Schema

Cancel

searched, browsed and downloaded without restrictions. MSCR provides an API to

Revisions

PID

1 Not available [View](#)

1 21.T13999/EOSC-202404000253357 [View](#)

1 Not available [View](#)

1 Not available [View](#)

2 21.T13999/EOSC-202404000253393 [View](#)

1 Not available [View](#)

1 21.T13999/EOSC-202404000253384 [View](#)

1 Not available [View](#)

1 Not available [View](#)

1 Not available [View](#)

1 Not available [View](#)

1 Not available [View](#)

Simple DC XML schema, version 2002-12-12

Search from source schema

Find an attribute... 

▼ ROOT

- contributor
- coverage
- rights
- format
- creator
- publisher
- language
- source
- subject
- identifier
- date
- description
- type
- relation
- title

Selected node info

 Nothing to show just now. Select an entity to view information.

☒ Show node titles



OpenAIRE Guidelines for Literature Repositories

Actions 

Search from target schema

Find an attribute... 

▼ ROOT

- file
- citationTitle
- any
- version
- citationVolume
- publisher
- > creators
- language
- > titles
- citationConferencePlace
- identifier
- > alternateIdentifiers
- resourceType
- > geoLocations
- > contributors
- fundingStream
- coverage

Selected node info

 Nothing to show just now. Select an entity to view information.

☒ Show node titles

Mappings

Source	Target	Filter from mappings 
rights	rights	▼

Mapping examples supported by the MSCR

- research.fi activities → CRediT taxonomy
 - Many-to-one type of mapping
 - supervisor of doctoral candidate, supervisor of licentiate candidate, supervisor of master student, etc. -> Supervision
- Role in research output → Affiliation statement
 - First names in output, Family name in output, Other name in output → displayPersonName
- Mapping between different ways of expressing DOI identifier
 - <identifierLink><https://doi.org/xxxx></identifierLink> → <doi>xxxx</doi>
 - <doi>xxxx</doi> → <identifier>doi:xxxx</identifier>
 - <identifier type="doi">xxxx</identifier> → <identifierLink><https://doi.org/xxxx></identifierLink>
- Mapping vocabulary values
 - <role_code>11.1.1</role_code> →
<role vocab="credit" vocab-identifier="https://credit.niso.org/" vocab-term="Supervision" vocab-term-identifier="https://credit.niso.org/contributor-roles/supervision/" >



FAIRCORE4EOSC
Core Components Supporting a FAIR EOSC

(Some) Mapping issues with MSCR



Funded by
the European Union



Mapping issues - modelling approaches

Subclassing vs. codelists

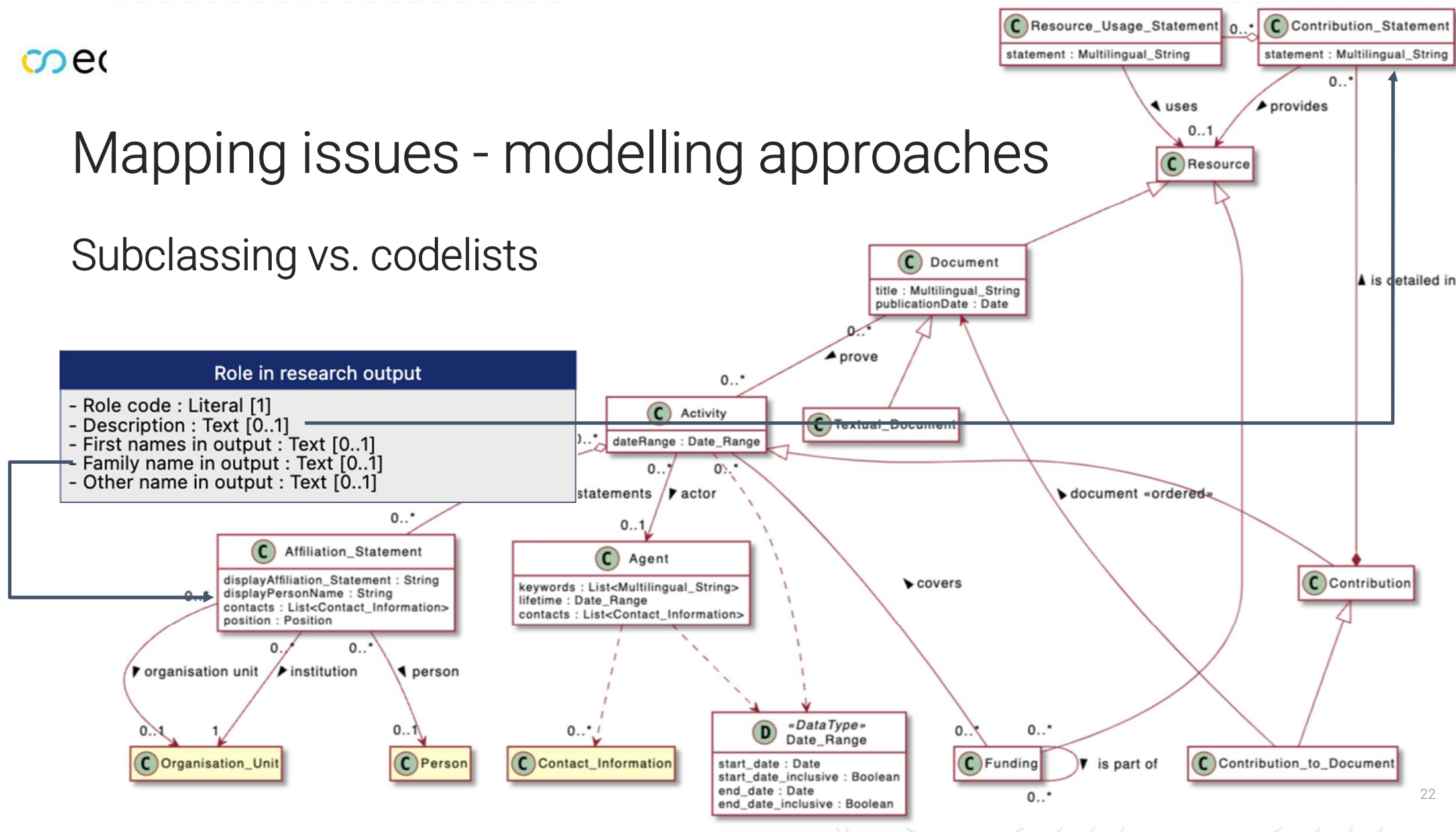
Role in research output
- Role code : Literal [1] - Description : Text [0..1] - First names in output : Text [0..1] - Family name in output : Text [0..1] - Other name in output : Text [0..1]

- 11 - supervising Valid
- 11.1 - supervision of postgraduate students Valid
 - 11.1.1 - supervisor of doctoral candidate Valid
 - 11.1.2 - supervisor of licentiate candidate Valid
- 11.2 - supervision of undergraduate students Valid
 - 11.2.1 - supervisor of bachelor student Valid
 - 11.2.2 - supervisor of master student Valid



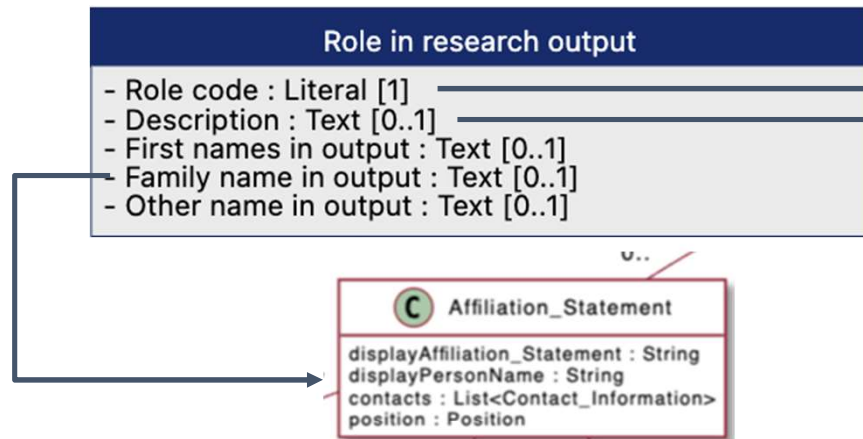
Mapping issues - modelling approaches

Subclassing vs. codelists

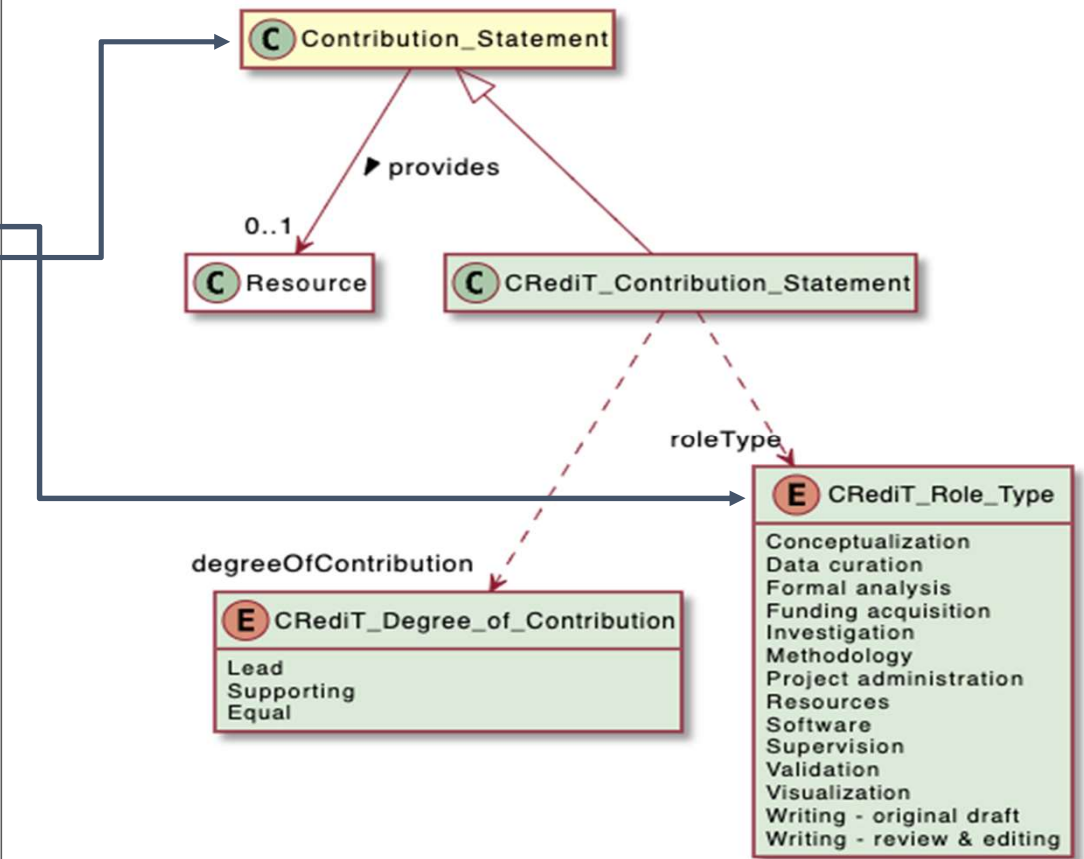


Mapping issues - modelling approaches

Subclassing vs. codelists



CERIF Scholarly Publications Module



Mapping issues - data model visualization

research.fi

Search from source schema

Find an attribute... 

▼ ROOT

> Funding Call

> Publication channel

> Parallel publication

> Ontology

> Web link

> Role in a research community

> Identifier

> Contact information

> Event

> Funding Decision

> Research community

>

> Research data

▼ Affiliation

Starting day

End date

Start date

Selected node info

@type:

tutkimus:Affiliaatio

qname:

ttv:Affiliaatio

description:

A person's relationship with the organisation includes a role (e.g. employment, docent, visiting researcher) and, in particular, in an employment relationship, also a job title and it has a certain duration, although it may still be open. Other affiliations with a fine difference in meaning (for which there is also a use case in the research): 1. Author — organization affiliation 2. Funding decision — applicant's affiliation

@id:

☒ Show node titles

CERIF2 - generated OWL

Search from target schema

Find an attribute... 

> :ISNI_Identifier

> :Call_for_Funding_Applications

> :Country

> :Funding

> :Contribution_Statement

> :Position

> :Person_Name

> :Electronic_Address

> :Expertise_and_Skills_Condition

> :Evaluation_Outcome

▼ :Involvement

:isCoveredBy

:isEvidencedBy

:dateRange

:in

:actor

:hasAffiliationStatements

:Event

Selected node info

@type:

:Involvement

qname:

:Involvement

description:

The act or process of taking part in something,^[1](<#fn1>)</sup> namely in a [Group or Organisation Unit] (../entities/Group_or_Organisation_Unit.md).

@id:

https://w3id.org/cerif2/Involvement

title:

Involvement

type:

object

properties:

☐ Show node titles



Mapping issues - support for complex mappings

Edit mapping

Source: A

Type: string

Description: N/A

Predicate

Exact match ▾

Notes:

No notes set. Add free form notes here.

Target: 1

Type: string

Description: N/A

Save

Cancel



Type mismatch between source and target

Source: [AffiliationIdentifier](#)

Source type: `String`

Description:

Example value: xx-yy-xx-yyxx

Source filter

Target ▼

[filter](#)(

Operator ▼

)

Value

Source operation

StringToInt ▼

Mapping operation

FormatDate ▼

pattern

inputFormat

Mapping type

Exact match ▼

Target: [dcterms:indentifier](#)

Target type: `Int`

Description:

Example value: xyxyxyxyxy

Target operation

addStringPrefix ▼

Prefix

Save

Cancel

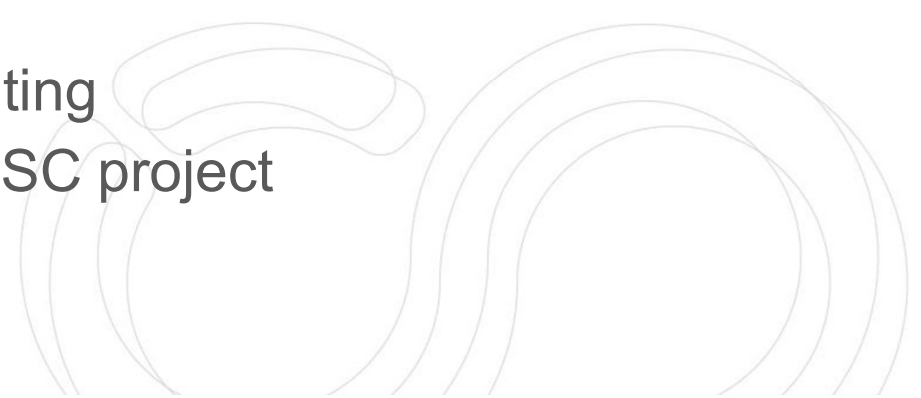
Benefits of using MSCR for crosswalks

- Mappings are stored in structured, machine-readable format
 - Possibility to export in current and upcoming de facto/standard formats (e.g. [SSSOM](#))
- Source and target schemas are available for validation of both input and output data
- Experts from “both sides” of the crosswalk can work together with a common format and tool (MSCR-UI)
- Utilization of linked data related to schemas
 - Vocabulary links
 - Ontology links



Timeline

- Summer 2024
 - First operationalization implementations available
 - XML to XML / XML to RDF
- Fall 2024
 - Support for more complex mappings (many-to-one, value transformations)
- Winter 2025
 - Release environment and public testing
- May 2025 - End of the FAIRCORE4EOSC project



faircore4eosc.eu

Twitter: @FAIRCORE4EOSC

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

Youtube: FAIRCORE4EOSC

Thank you



**Funded by
the European Union**

