

# Refactoring CERIF and Mapping it to Research.fi

Jan Dvořák

 <https://orcid.org/0000-0001-8985-152X>

Dragan Ivanović

 <https://orcid.org/0000-0002-9942-5521>

euroCRIS 2023 Strategic Membership Meeting  
2023-11-22 @ Pamplona, Spain

# CERIF Refactoring Goals

1. Keep the strong features of current CERIF
  - a. Multilinguality
2. Take away the perceived complexity of CERIF
  - a. Emphasize the conceptual model
  - b. Improve documentation
3. Modernize CERIF
  - a. Change the modeling notation
  - b. Adapt CERIF for usage in APIs and for Linked Open Data
  - c. Modern serialization formats
  - d. Allow for systematic provenance tracking and verifiable credentials
4. Involve the community in further development of CERIF
  - a. Modularity
  - b. Open source sw development practices

# CERIF Refactoring Means

1. New modelling approach (and tool)
  - a. Improved documentation
  - b. Element addressability
2. Modularity ( $\Rightarrow$  scalability of further CERIF development)
  - a. The Core and pluggable modules
  - b. Datatypes to represent reusable functional groups
  - c. Developed in a decentralized way
3. Modern serialization formats (with appropriate schema expressions)
  - a. JSON-LD (JSON Schema, SHACL) for APIs
  - b. RDF (RDF Schema, SHACL) for LOD
  - c. previously: XML (XML Schema, Schematron)

# CERIF Core

← → ↻ https://github.com/EuroCRIS/CERIF-Core

EuroCRIS / CERIF-Core Public

Edit Pins Unwatch 4 Fork 1 Star 2

<> Code Issues 1 Pull requests Actions Projects Wiki Security Insights

main 2 branches 0 tags Go to file Add file Code

actions-user adding generated diag... d1e218d on Aug 26 350 commits

.github/workflows	Update plantuml.yml	17 months ago
datatypes	Aligning md files with example for UC1	5 months ago
diagrams	adding generated diagrams	3 months ago
entities	Add Person_Identifier as an entity	3 months ago
examples	link correction	13 months ago
guidelines	Let Attributes and Relationships sections ...	3 months ago
serializations/RDF	Rename Affiliation Statement -> Affiliation	5 months ago
tools	Correct the links to the parent entity	4 months ago
README.md	Fix the links to entities	3 months ago

☰ README.md

## CERIF Core

CERIF (=Common European Research Information Format) covers the domain of research information with a focus on the administrative and organizational aspects. It aims to provide machine-processable

**About**

The CERIF Refactoring Pilot project started in 2021 by euroCRIS.

- Readme
- 2 stars
- 4 watching
- 1 fork

**Releases**

No releases published  
[Create a new release](#)

**Packages**

No packages published  
[Publish your first package](#)

**Contributors 3**

- jdvorak001 Jan Dvorak
- chenejac Dragan Ivanovic
- actions-user

[https://github.com/  
EuroCRIS/  
CERIF-Core](https://github.com/EuroCRIS/CERIF-Core)

# CERIF Core

Contains:

- common data types
- abstract entities
- commonly used entities
- common constructs

The basis for:

- Particular research information interchange scenarios
- Extensions in terms of coverage

# CERIF Agent

Def: Anything that has the ability to perform actions. (following FOAF)

Generalizes the former cfPerson + cfOrgUnit

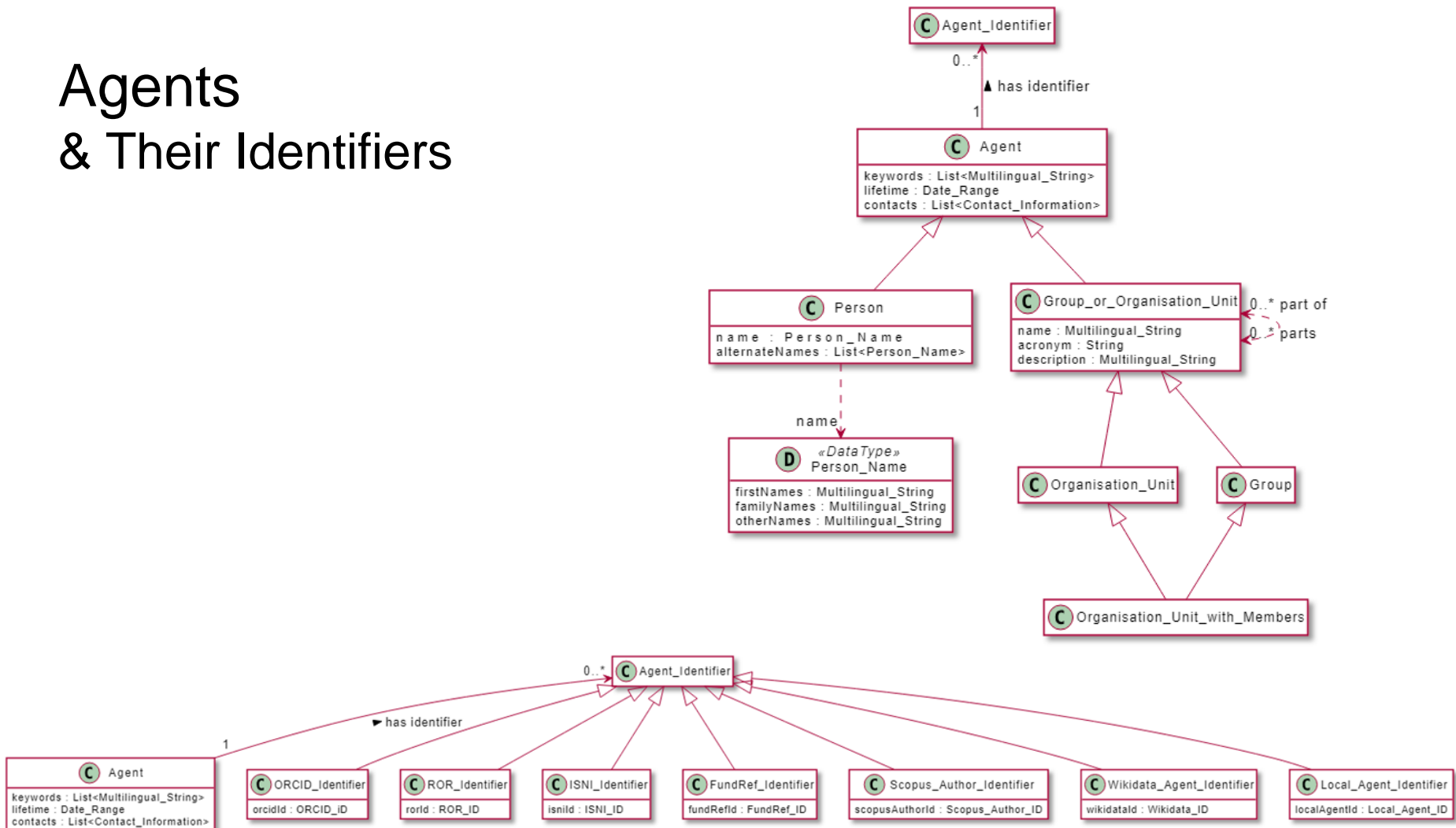
Subclassed as:

- Person
- Group\_or\_Organisation\_Unit
  - Organisation\_Unit (formalized)
  - Group (has members)

Organisation\_Unit\_with\_Members



# Agents & Their Identifiers





# CERIF Activity

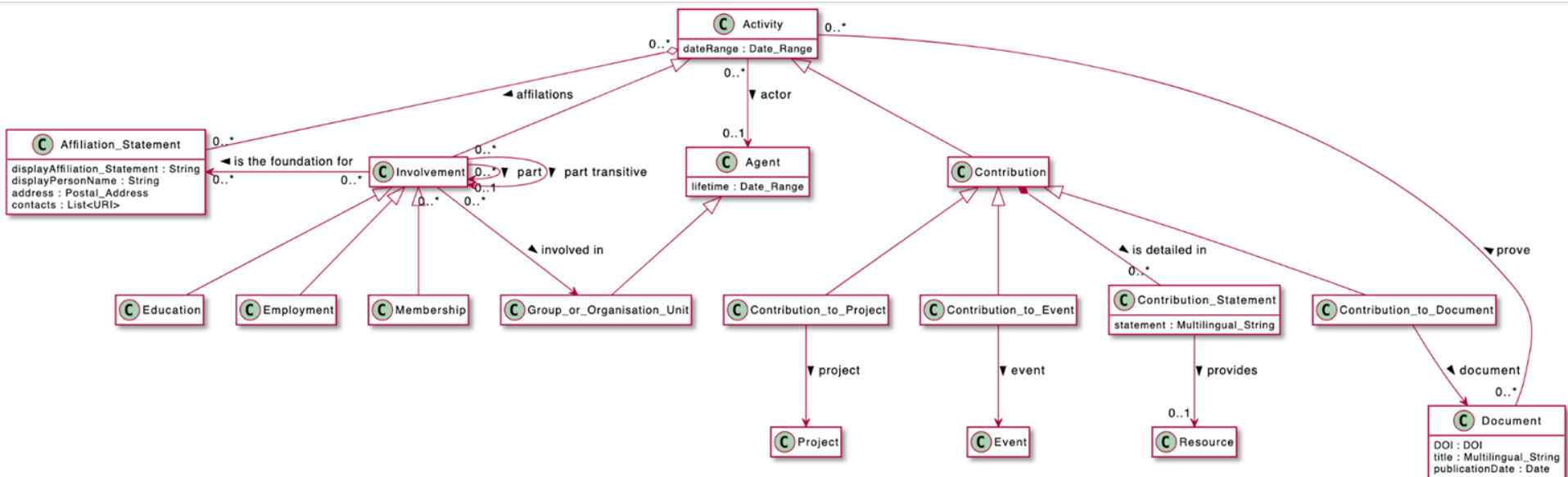
Def: Anything an Agent does.

Subclassed as:

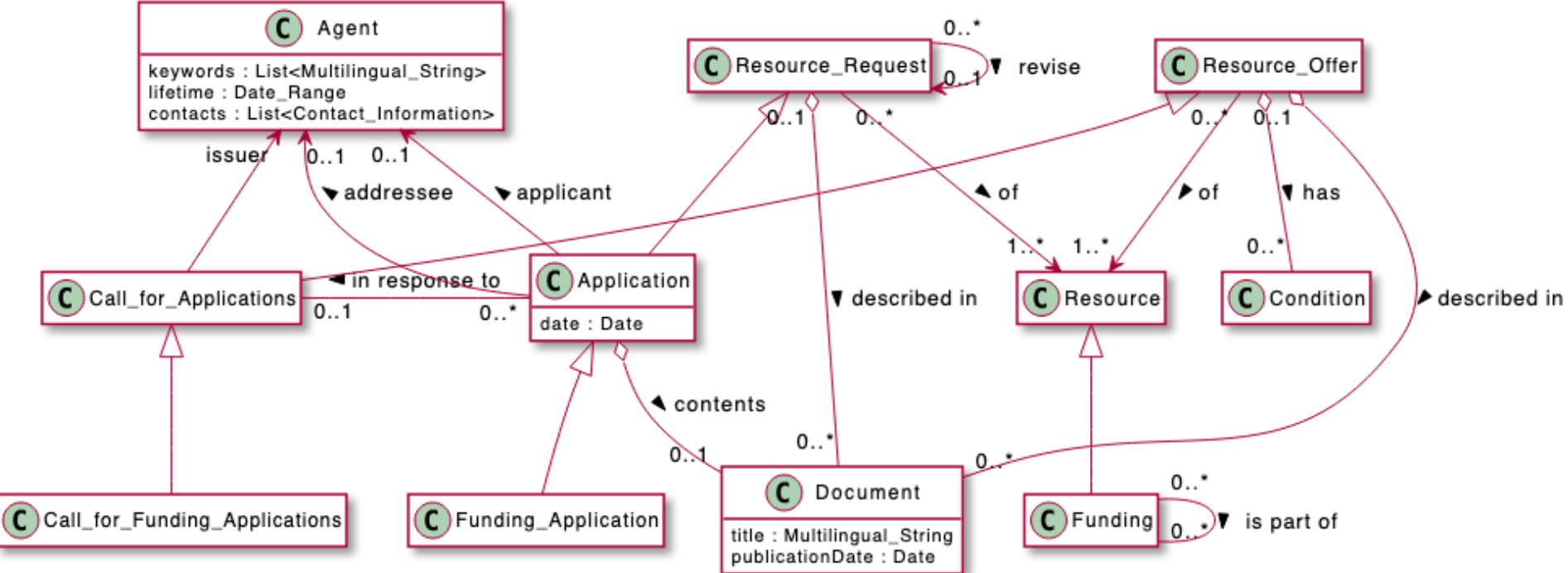
- Involvement (Membership, Employment or Education)  
in a Group or Organisation Unit
- Contribution to a Document, a Project, an Event, ...

This is a refinement of the Linking Entities of current CERIF starting from cfPerson or cfOrgUnit (dateRange is an attribute of Activity)

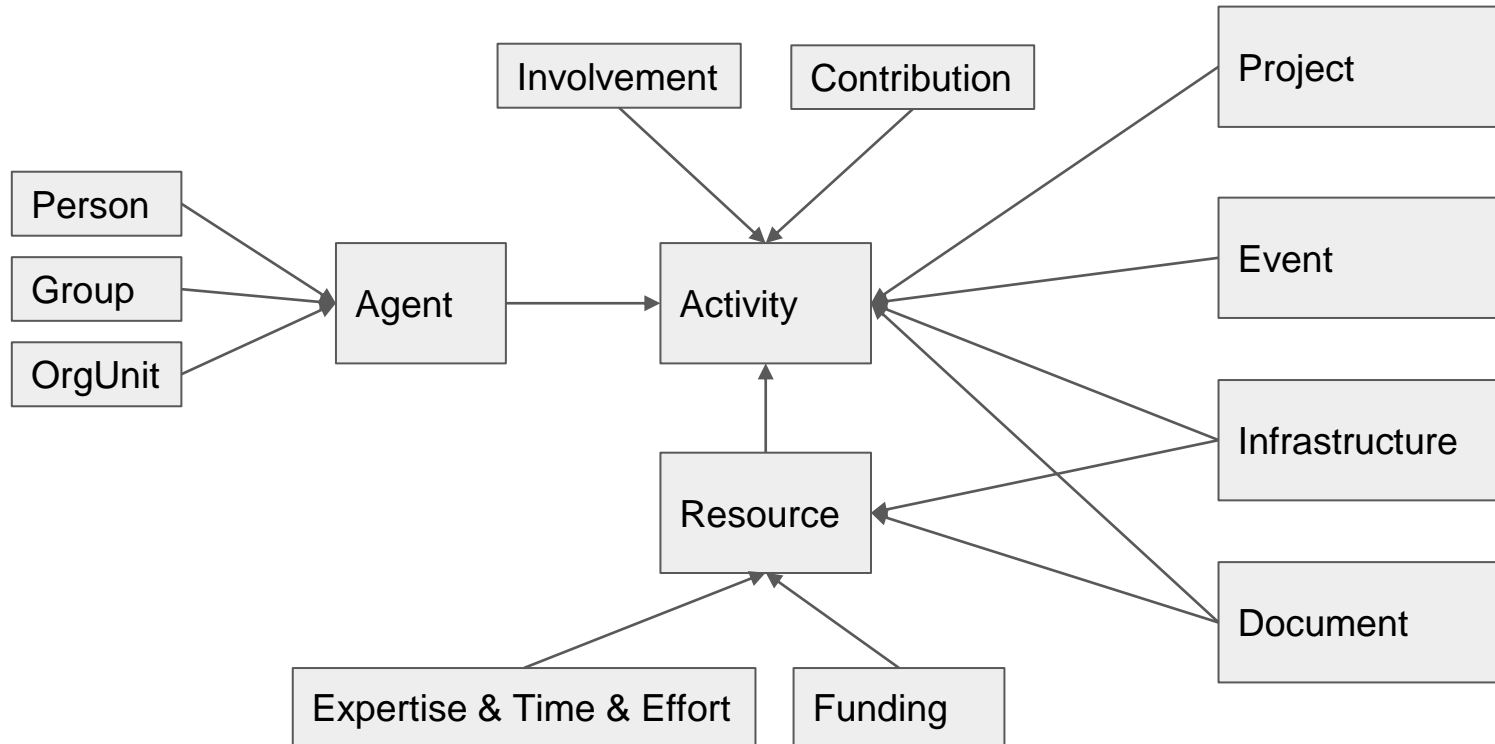
# Activity hierarchy



# Resources, Offers, Requests



# CERIF Core “View from the Top”



# Modules

# CERIF Modules

Add functionality

- E.g. the Scholarly publication module

Customize the model for a specific scenario

- E.g. institutions to report publications produced with Horizon Europe funding to the EC (tentative)

Contributed by another group of people

Reuse & Extend stuff from the Core and other modules

Candidate modules:

- Dataset metadata
- Funding & Projects
- Research Infrastructure description
- Research software metadata
- Bibliometric indicators

# CERIF Scholarly Publication Module

1st module started along with the Core

Scope:

- Journal Article
- Journal
- Monograph
- (Book Chapter)
- (Proceedings Paper)

main 1 branch 0 tags

Go to file Add file Code

jdvorak001 Add human-readable anchors to the journal-articles relationship ✓ 7fe3b7a on 10 Mar 87 commits

.github/workflows	Update plantuml.yml	6 months ago
datatypes	The initial creation of datatypes ISBN and ISSN, and update of Editor...	10 months ago
diagrams	🚀 adding generated diagrams	2 months ago
entities	Add human-readable anchors to the journal-articles relationship	2 months ago
examples	link correction	6 months ago
serializations/RDF	adding examples directory	6 months ago
README.md	The Development section now references the Core	6 months ago

☰ README.md ✎

# CERIF Scholarly Publications Module

A module to represent scholarly publications in the refactored CERIF. Its secondary purpose is to demonstrate the liason between the [CERIF Core](#) and a module.

## Status

### About

No description, website, or topics provided.

Readme

0 stars

3 watching

0 forks

---

### Releases

No releases published

[Create a new release](#)

---

### Packages

No packages published

[Publish your first package](#)

---

### Contributors 3

chenejac Dragan Ivanovic

jdvorak001 Jan Dvorak



# OWL Ontology for the Refactored CERIF

# OWL Ontology generated using a java tool

Markdown descriptions → Ontology

(parsed by [flexmark](#))

(constructed and output by [owlapi](#))

Features used:

- owl:DatatypeProperty
- owl:ObjectProperty
- owl:FunctionalProperty
- owl:InverseProperty
- rdfs:domain
- rdfs:range
- ...

Currently for the Core

TODO:

- Support also Modules
- Check consistency with the PlantUML diagrams

# Where we are

The Core considered complete

The Scholarly publication module will need expansion with publication types

- We'd like to rely on the community here

Next developments:

- Dataset module
- Funding module
- Serialization formats (JSON+RDF | JSON-LD)

## Summary

Working towards a new CERIF that is:

- Better positioned for real-world research information interchange scenarios
- More customizable
- Engaging the community

Currently piloted in the FAIRCORE4EOSC project

- Mapping national research information services

# Mapping Research.fi to the Refactored CERIF

# Goal

- To make interoperable Research.fi with any platform which could map its data to the CERIF format
  - Mapping over proxy
- To develop/validate infrastructure for defining crosswalks on top of this mapping example
- To validate the Refactored CERIF model

# Phases

- Analysis of [Research.fi data model](#)
  - Getting domain knowledge
  - Making subset which should be mapped
- Mapping of Research.fi data model elements to CERIF elements
  - Agents, Publications, Research Dataset, Channels, Funding, Infrastructures
  - Aligning/creation of CERIF Core and modules
- Operational mapping
  - Expression of mapping in the crosswalk tool
- Validation
  - Testing of operational mapping and improving the crosswalk tool

# Analysis of Research.fi

- CERIF experts got access to Research.fi model
- Missing translation into English has been added
- Entities and attributes extracted to a spreadsheet document
- A couple of meetings between Research.fi model experts and CERIF experts

<a href="#">Research.fi</a>				
class	property	data type	cardinality	description
Organization				
Organization	Name	Text	1..*	
Organization	Alternate name	Text	0..*	
Organization	Description	Text	0..*	
Organization	Organization status	Text	0..*	
Organization	Organization's sector code	Literal	0..1	
Organization	Visiting address	Text	1..*	
Organization	Postal address	Text	1..1	



# Mapping of Research.fi elements to CERIF elements

- Mapping expressed in a spreadsheet document
- Alignment/extension of the CERIF core

CERIF				
class	property	data type	cardinality	description
Person				
Person	name.family names			
Person	name.first names			
Person	alternate_names.???			

## Add alternate\_names to Person #10

 Closed jdvorak001 opened this issue on Sep 6 · 0 comments



jdvorak001 commented on Sep 6

Member ...

A collection of Person\_Names. Semantics will be defined by the provenance of each such name.



chenejac added a commit that referenced this issue 3 weeks ago




Add alternate names into Person (#10)

970832c


 3 Open  11 Closed

 Add Position\_Title to the Activity diagram


#12 by jdvorak001 was closed 1 minute ago

 Add 'completed' flag in Education


#11 by jdvorak001 was closed 2 minutes ago

 Add alternate\_names to Person


#10 by jdvorak001 was closed 3 minutes ago

 Pull 'name' and 'acronym' from Organisation\_Unit to Group\_or\_C


#9 by jdvorak001 was closed 1 minute ago

 Add Contact\_Information datatype


#8 by jdvorak001 was closed now

 Add parent-child relationships between Organization\_Units

#7 by jdvorak001 was closed 1 minute ago

 Add Resource\_Identifier hierarchy

#6 by jdvorak001 was closed 1 minute ago

 Add Local\_Agent\_Identifier into the Agent\_Identifier hierarchy

#5 by jdvorak001 was closed now

# Example of mapping

- Manually created example of data mapping (not structure mapping)
  - For the purpose of specification, implementation and validation of the crosswalk tool

<a href="#">Research.fi</a>	Refactored CERIF RDF Triples
<researcher>	_:001 a https://w3id.org/cerif2/CERIF-Core/Person
<persistent_identifier>	_:001 https://w3id.org/cerif2/CERIF-Core/Agent#hasIdentifier _:002
<identifier>0000-1111-2222-3333</identifier>	_:002 a https://w3id.org/cerif2/CERIF-Core/ORCID_Identifier
<identifier_type>Orcid</identifier_type>	_:002 https://w3id.org/cerif2/CERIF-Core/ORCID_Identifier#orcidId "https://orcid.org/0000-1111-2222-3333"
</persistent_identifier>	
<member_of>	_:003 a https://w3id.org/cerif2/CERIF-Core/Employment

# Status

- Analysis of Research.fi
  - Done
- Mapping of Research.fi data model elements to CERIF elements
  - Agents, Publications, Research Dataset, Channels, Funding, Infrastructures
  - Aligning/creation of CERIF Core and modules
- Operational mapping
  - Expression of mapping in the crosswalk tool
- Validation
  - Testing of operational mapping and improving the crosswalk tool