Refactoring CERIF and Mapping it to Research.fi

Jan Dvořák
Dragan Ivanović

https://orcid.org/0000-0001-8985-152X
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 (⇒ 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

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...
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)

  ![Class Diagram]

  Organisation_Unit_with_Members
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

[Diagram of resource management concepts including Agent, Resource_Request, Resource_Offer, Call_for_Applications, Application, Resource, Condition, Document, and Funding_Application.]
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)
CERIF Scholarly Publications Module

A module to represent scholarly publications in the refactored CERIF. Its secondary purpose is to demonstrate the liaison between the CERIF Core and a module.

Status
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

<table>
<thead>
<tr>
<th>class</th>
<th>property</th>
<th>data type</th>
<th>cardinality</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organization</td>
<td>Name</td>
<td>Text</td>
<td>1..*</td>
<td></td>
</tr>
<tr>
<td>Organization</td>
<td>Alternate name</td>
<td>Text</td>
<td>0..*</td>
<td></td>
</tr>
<tr>
<td>Organization</td>
<td>Description</td>
<td>Text</td>
<td>0..*</td>
<td></td>
</tr>
<tr>
<td>Organization</td>
<td>Organization status</td>
<td>Text</td>
<td>0..*</td>
<td></td>
</tr>
<tr>
<td>Organization</td>
<td>Organization's sector code</td>
<td>Literal</td>
<td>0..1</td>
<td></td>
</tr>
<tr>
<td>Organization</td>
<td>Visiting address</td>
<td>Text</td>
<td>1..*</td>
<td></td>
</tr>
<tr>
<td>Organization</td>
<td>Postal address</td>
<td>Text</td>
<td>1..1</td>
<td></td>
</tr>
</tbody>
</table>
Mapping of Research.fi elements to CERIF elements

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

<table>
<thead>
<tr>
<th>CERIF</th>
<th>property</th>
<th>data type</th>
<th>cardinality</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person</td>
<td>name.family names</td>
<td>Person</td>
<td>3 Open</td>
<td>Add Position_Title to the Activity diagram</td>
</tr>
<tr>
<td>Person</td>
<td>name.first names</td>
<td>Person</td>
<td>11 Closed</td>
<td>Add ′completed′ flag in Education</td>
</tr>
<tr>
<td>Person</td>
<td>alternate_names</td>
<td>Person</td>
<td>3 Open</td>
<td>Add alternate_names to Person</td>
</tr>
<tr>
<td>Person</td>
<td>alternate_names</td>
<td>Person</td>
<td>11 Closed</td>
<td>Add alternate_names to Person</td>
</tr>
<tr>
<td>Person</td>
<td>alternate_names</td>
<td>Person</td>
<td>3 Open</td>
<td>Pull ′name′ and ′acronym′ from Organisation_Unit to Group_or_C</td>
</tr>
<tr>
<td>Person</td>
<td>alternate_names</td>
<td>Person</td>
<td>11 Closed</td>
<td>Add Contact_Information datatype</td>
</tr>
<tr>
<td>Person</td>
<td>alternate_names</td>
<td>Person</td>
<td>3 Open</td>
<td>Add parent-child relationships between Organization_Units</td>
</tr>
<tr>
<td>Person</td>
<td>alternate_names</td>
<td>Person</td>
<td>11 Closed</td>
<td>Add Resource_Identifier hierarchy</td>
</tr>
<tr>
<td>Person</td>
<td>alternate_names</td>
<td>Person</td>
<td>3 Open</td>
<td>Add Local_Agent_Identifier into the Agent_Identifier hierarchy</td>
</tr>
</tbody>
</table>
Example of mapping

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

<table>
<thead>
<tr>
<th>Research.fi</th>
<th>Refactored CERIF RDF Triples</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;researcher&gt;</td>
<td>_:001 a <a href="https://w3id.org/cerif2/CERIF-Core/Person">https://w3id.org/cerif2/CERIF-Core/Person</a></td>
</tr>
<tr>
<td>&lt;persistent_identifier&gt;</td>
<td>_:001 <a href="https://w3id.org/cerif2/CERIF-Core/Agent#hasIdentifier">https://w3id.org/cerif2/CERIF-Core/Agent#hasIdentifier</a> _:002</td>
</tr>
<tr>
<td>&lt;identifier&gt;0000-1111-2222-3333&lt;/identifier&gt;</td>
<td>_:002 a <a href="https://w3id.org/cerif2/CERIF-Core/ORCID_Identifier">https://w3id.org/cerif2/CERIF-Core/ORCID_Identifier</a></td>
</tr>
<tr>
<td>&lt;identifier_type&gt;Orcid&lt;/identifier_type&gt;</td>
<td>_:002 <a href="https://w3id.org/cerif2/CERIF-Core/ORCID_Identifier#orcidId">https://w3id.org/cerif2/CERIF-Core/ORCID_Identifier#orcidId</a> &quot;<a href="https://orcid.org/0000-1111-2222-3333">https://orcid.org/0000-1111-2222-3333</a>&quot;</td>
</tr>
<tr>
<td>&lt;/persistent_identifier&gt;</td>
<td></td>
</tr>
</tbody>
</table>

|  |
| <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