

# The euroCRIS CERIF Refactoring Pilot Project

euroCRIS Webinar 2021-07-06

Moderator: Anna Guillaumet

 <https://orcid.org/0000-0002-1944-5259>

Presenters: Jan Dvořák

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

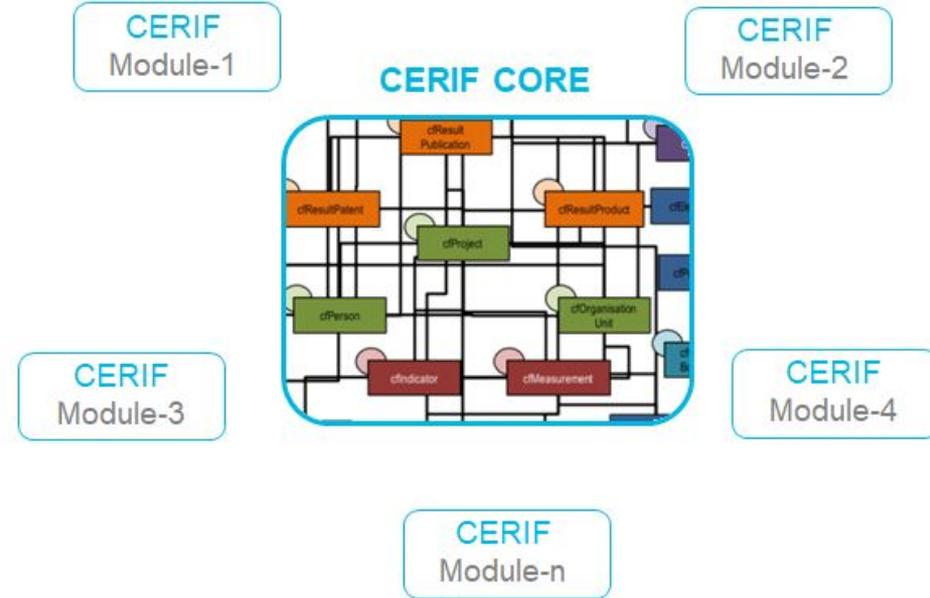
Dragan Ivanović

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

# The CERIF Refactoring Project

The CERIF model is large and complex, therefore a refactoring of the CERIF model has been planned first.

This refactoring proposes to **divide the CERIF model in a CORE model** that will contain the essential entities, and then the definition of specific areas of entities that completes the model and that should be defined by experts in the area.



# euroCRIS Technical Committee

euroCRIS has created a [Technical Committee for Interoperability and Standards \(TCIS\)](#), to lead and coordinate the CERIF refactoring project. Can find more information [here](#). Twelve members from eight different countries as shown below, five overseeing this project:

- Gavin Reddick (Researchfish Ltd - Interfolio UK Ltd) (UK)
- Alejandro Engelmann (Sveriges lantbruksuniversitet - SLU University Library) (SE)
- Daniele Bailo (EPOS/INGV) (IT)
- **Christian Hauschke (TIB Hannover/VIVO)** (DE)
- Brian Plauborg (Elsevier Pure) (DK)
- Sophie Biesenbender (DZHW) (DE)
- Jochen Schirrwagen (Uni Bielefeld, OpenAIRE) (DE)
- **Sebastian Herwig (Uni Münster, DINI FIS WG)** (DE)
- **Tommi Suominen (CSC)** (FI)
- **Anna Guillaumet (SIGMA AIE/VIVO)** (ES)
- Geert van Grootel (EWI) (BE)
- **Miguel Angel Sicilia (UAH)** (ES)

# Outline

Current CERIF

- Features

- Shortcomings

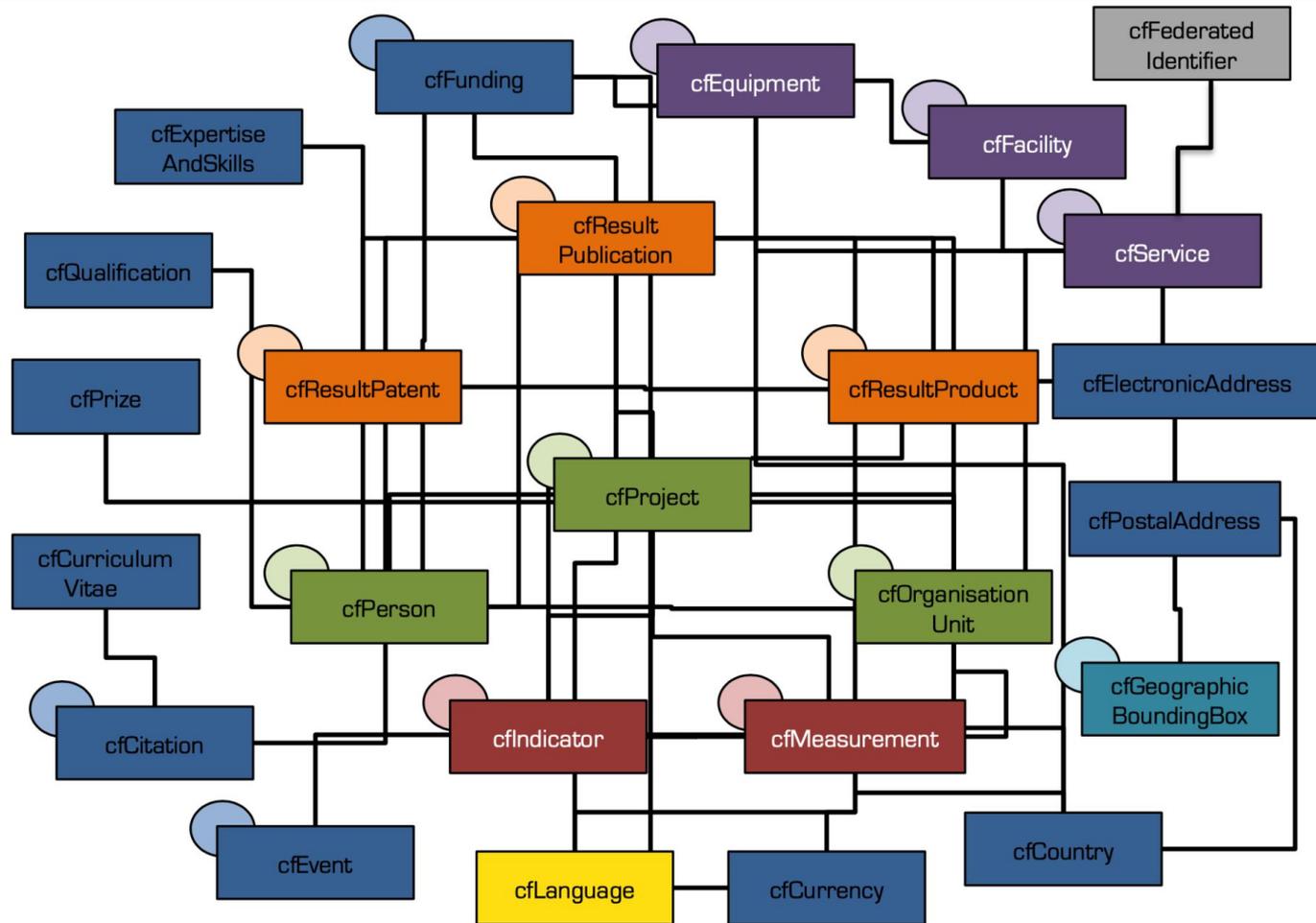
The CERIF Refactoring Pilot project

- The Goals and the Scope

- The progress so far

# Current CERIF

# Common European Research Information Format



# Current CERIF features

Entity-Relationship model

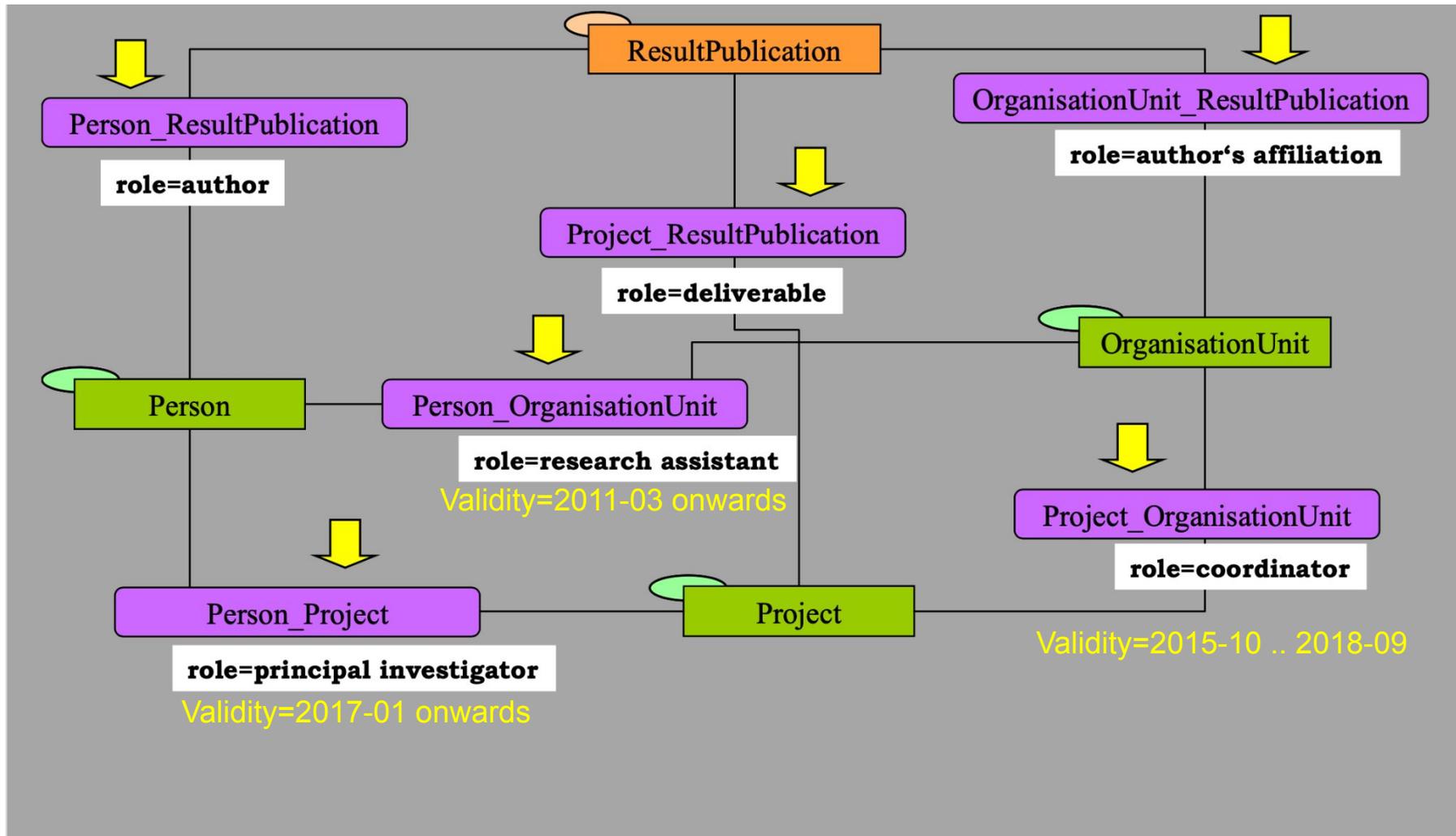
Basically a Graph structure:

- Entities ~ nodes
- Linking entities ~ edges

Supports multilingual free text attributes

Entities and links are classified:

- type, state, subject, relationship type, ...
- temporal validity ranges
- classifications stored in a central place, the Semantic Layer



# CERIF-XML exchange formats (based on XML Schema)

## Original

1:1 with the ER structure

Only uses embedding for multilingual texts

→ Many foreign key relationships

→ Takes several API requests to get a presentable form of an object

Ex.: OpenAIRE Guidelines for CRIS Managers 1.0.  
DOI [10.5281/zenodo.17065](https://doi.org/10.5281/zenodo.17065)

## 2nd Generation

Template XML Schema → adaptation

Profiles: Useful subsets of CERIF for specific research information exchange scenarios

1. Specify a subset of CERIF entities & attributes
2. Fix semantic vocabularies to use
3. Add integrity constraints

Ex.: OpenAIRE Guidelines for CRIS Managers 1.1.  
DOI [10.5281/zenodo.1298649](https://doi.org/10.5281/zenodo.1298649)

# Current CERIF shortcomings (1/2)

## Complexity

- Vocabularies are not part of the model (Profiles help here)
- Time ranges are not used most of time, sometimes they can be confusing
- Generic constructs
  - Federated Identifiers
  - Measurements
  - Result Products

## Flexibility

Fine for producers

Very difficult for consumers

## Limitations of the current Result

### Publication/Patent/Product framework

- Fails to represent research inputs not originating from other research
- Leaves out research-related documents, such as administrative documents (applications, contracts, ...) which are useful to preserve in CRIS
- Open Science developments (datasets, software, ...) not represented in the model itself

# Current CERIF shortcomings (2/2)

## **Monolithic architecture**

→ Non-scalable development process

**ER modeling paradigm is not a good fit for information interchange**

XML is not the #1 format for APIs

# CERIF Refactoring

# 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 de-centralized 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)

# Motivating use-case #1

A researcher is willing to move from a country to other (e.g. France to Australia) to a different institution. It is expected that his **profile and vitae** may be **transferred electronically** to the new institution via a standardised model and format that can be **digitally signed** as a verifiable claim or verifiable credential (following the directions of the W3C working group of the same name).

## Motivating use-case #2

Partners in a multinational project funded under the EU H2020 programme need to **share project participation data electronically** with the coordinator **for reporting** under the particular rules of a scheme or call. The partners include non-EU members, e.g. from China.

## Motivating use-case #3

A researcher is **looking for a research infrastructure** for a very concrete physics experiment and wants to know if there are e-infrastructures available in other countries that he could eventually use, with or without paying for it.

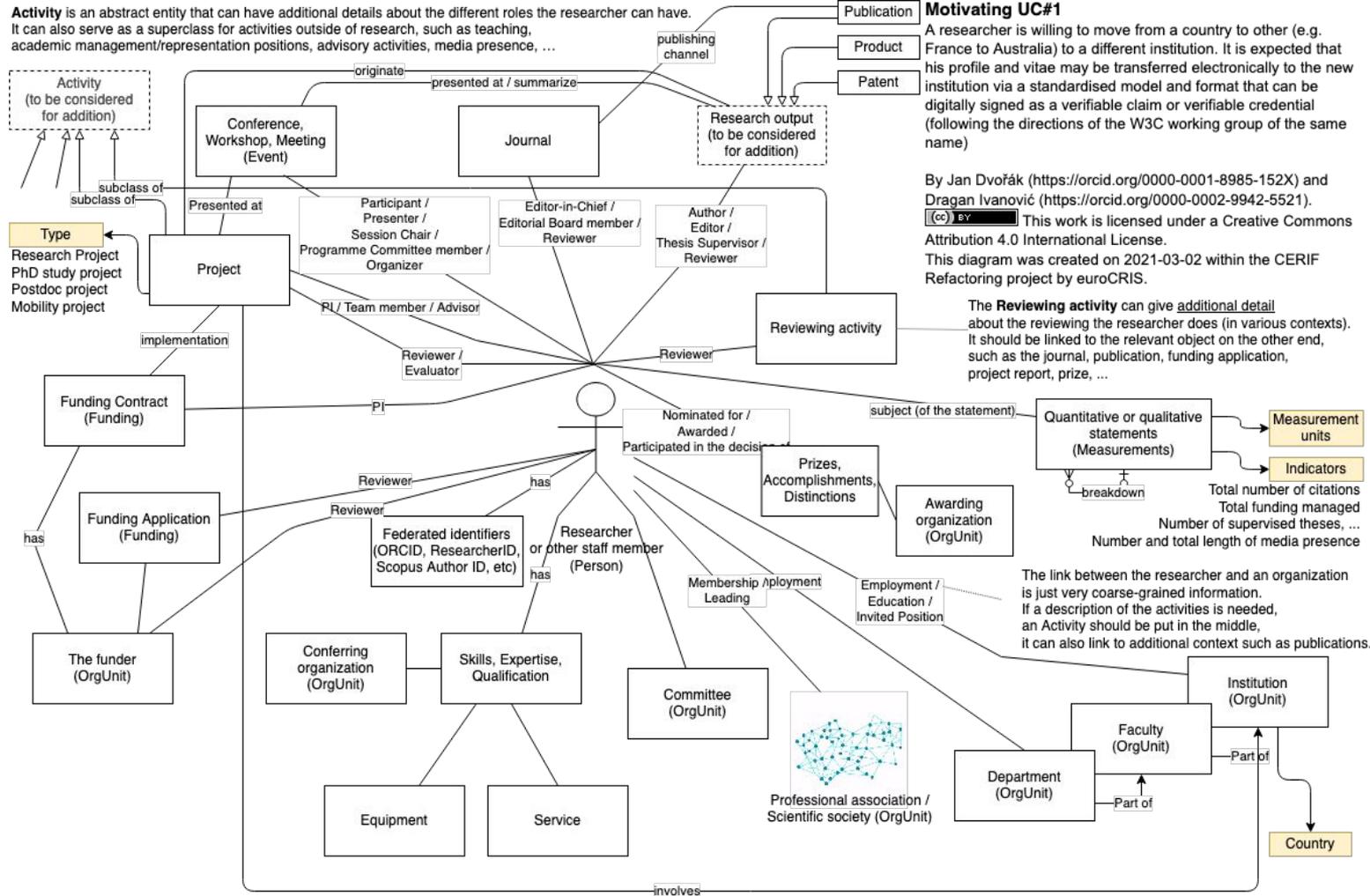
# CERIF Refactoring Pilot Project

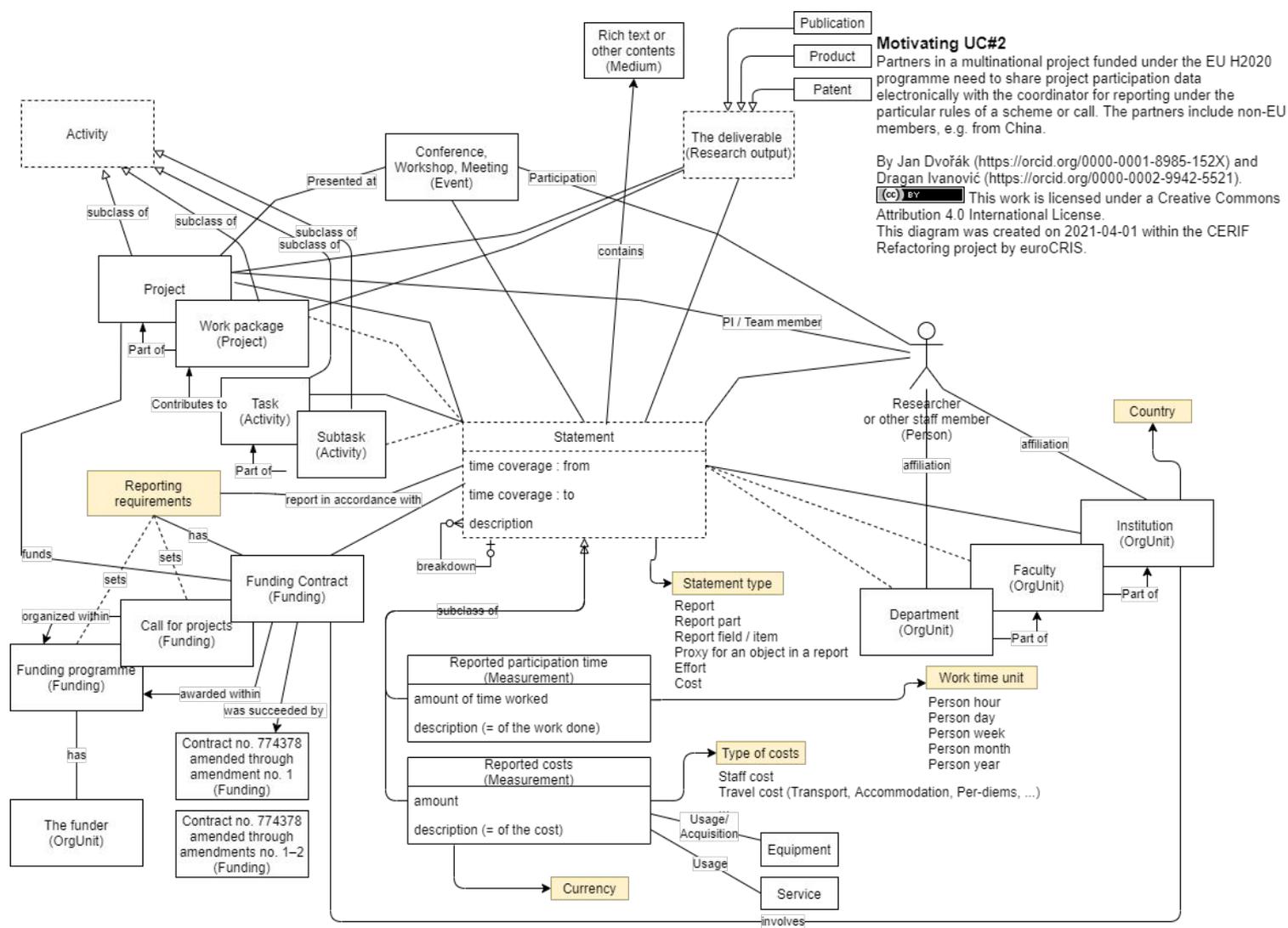
1. Separating the specification of data types from the main entities in the model and preparing them for multilinguality and localization, using best practices from other data interchange standards.
2. Extracting the “CERIF core model” as a subset of CERIF that represents the minimum entities, properties and relations that are essential to a set of previously agreed CRIS core use cases.
3. Selecting the best modeling notation or language for expressing the data types and the CERIF core model, and for future models layered on top of the core. The notation needs to have both textual, formal (or semi-formal) and graphical representations.
4. Documenting the core model and data types using formats and templates that generate documentation in several formats and support translations.
5. Specify potential modules to be layered on top of the core model for covering at least the use cases mentioned, and that could be independently reused. Draft the contents that may be included in those modules for the use cases, including selecting best practices in models that could be reused, e.g. in the sub-domain of bibliographic information.
6. Specifying a concrete mapping in at least JSON and RDF of the core model and the extensions to demonstrate the use cases, including some form of constraint checking for validation of the data transferred in both cases.
7. Specifying how provenance information can be expressed in a data interchange to support the concept of verifiable claims or credentials.
8. Creating a proof of concept implementation in which some basic test data is interchanged according to the abovementioned use cases or others of similar complexity and scope.

# Phases of the Project

1. Reorganization of CERIF entities into modules
2. Selecting notations and language for modelling
3. Setting up the infrastructure
4. Modelling of the core and the auxiliary entities
5. Modelling of the selected subset of modules
6. Interchange formats

# Progress so far



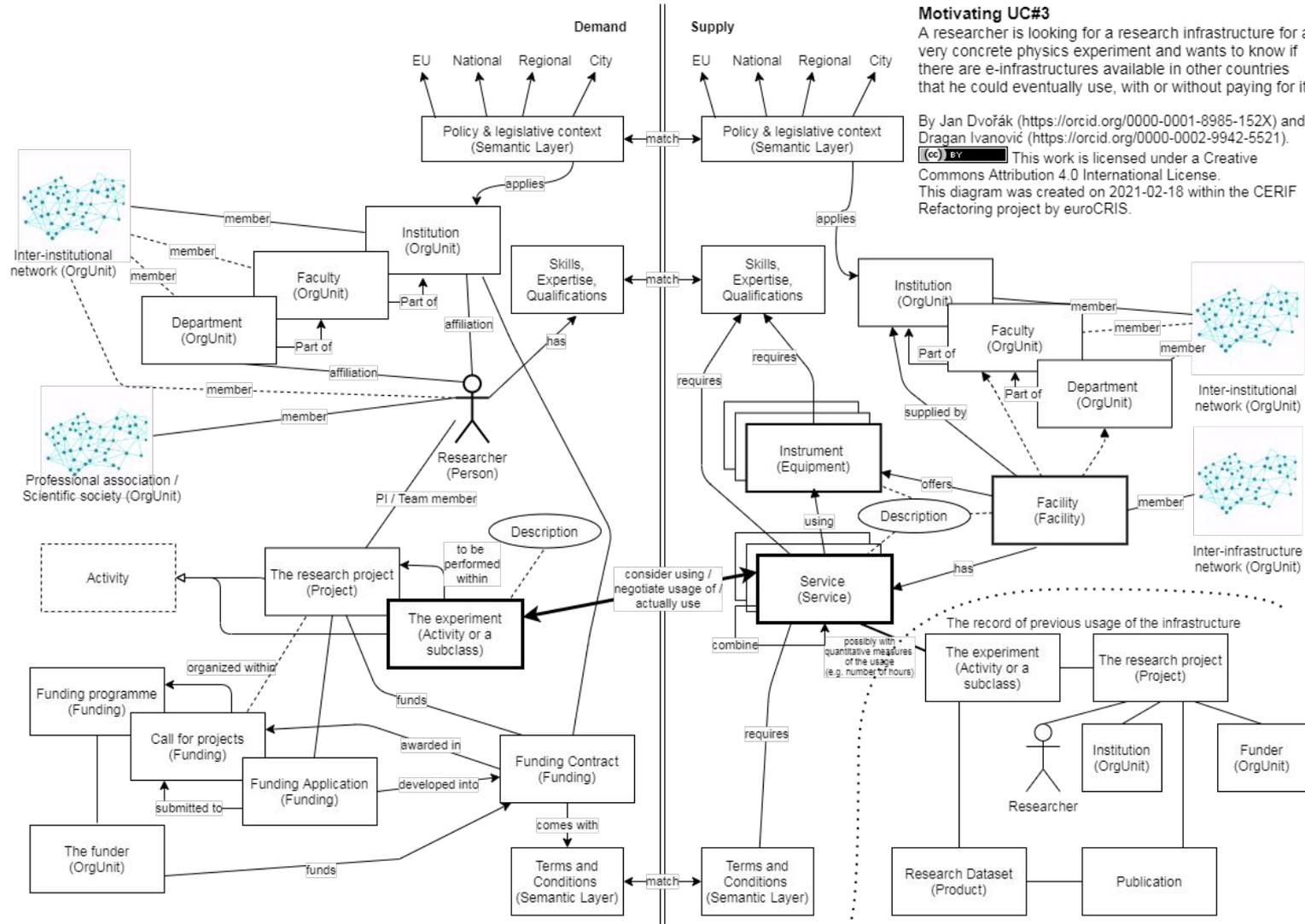


### Motivating UC#2

Partners in a multinational project funded under the EU H2020 programme need to share project participation data electronically with the coordinator for reporting under the particular rules of a scheme or call. The partners include non-EU members, e.g. from China.

By Jan Dvořák (<https://orcid.org/0000-0001-8985-152X>) and Dragan Ivanović (<https://orcid.org/0000-0002-9942-5521>).

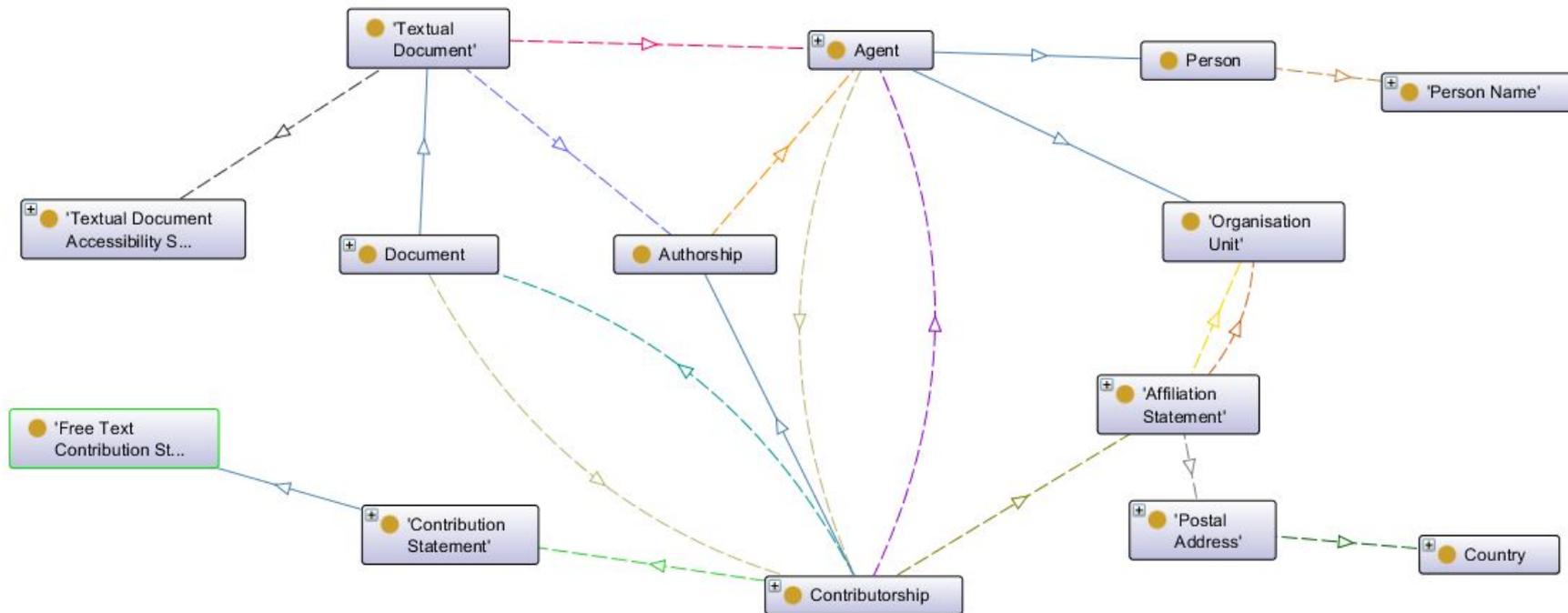
This work is licensed under a Creative Commons Attribution 4.0 International License. This diagram was created on 2021-04-01 within the CERIF Refactoring project by euroCRIS.



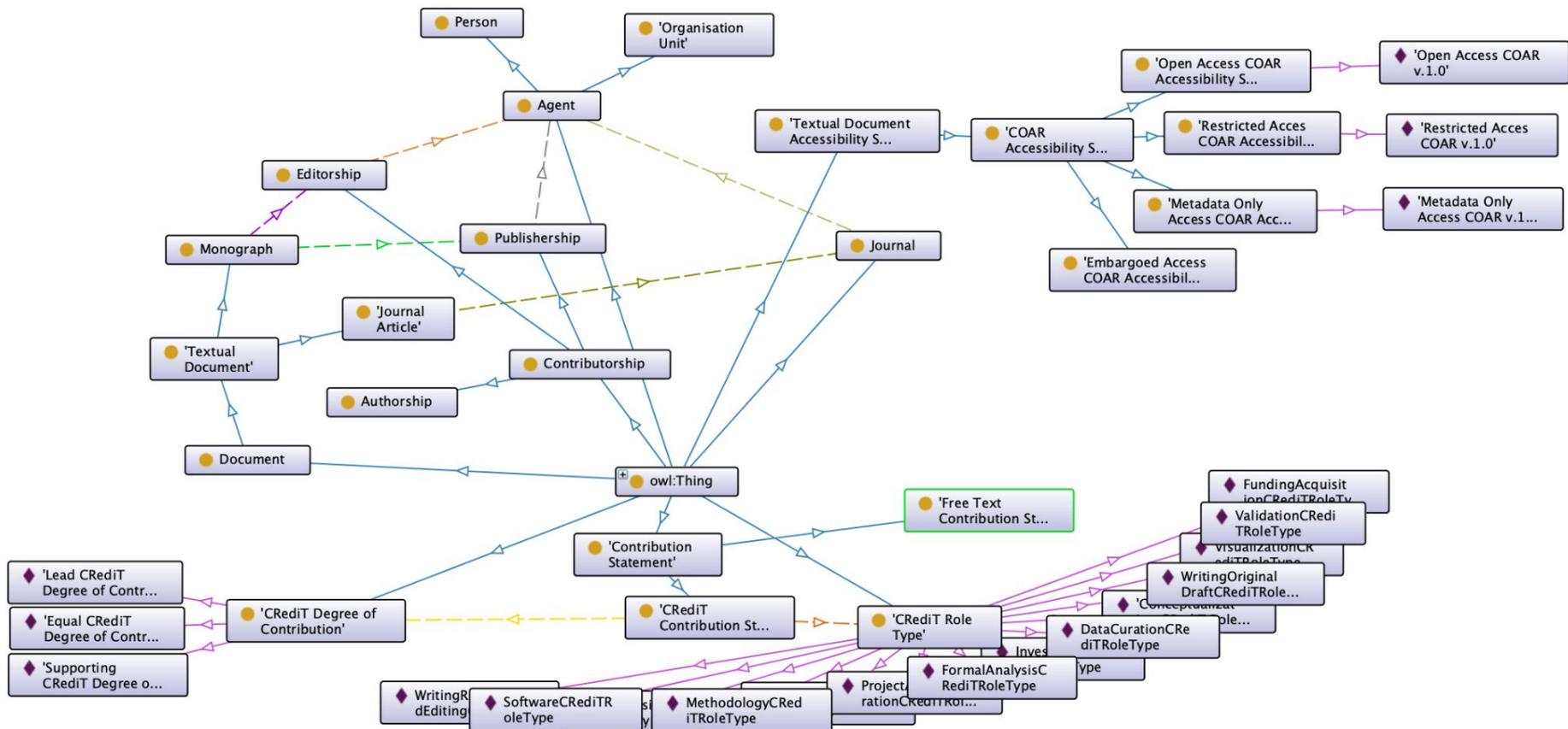
# Core & Module Examples



# The prototype CERIF Core as OWL



# Prototype Scholarly Publication Modules with parts of the prototype Core as OWL



# Example publication

<https://doi.org/10.1016/j.ejor.2020.09.020>



European Journal of Operational Research

Volume 291, Issue 3, 16 June 2021, Pages 906-917

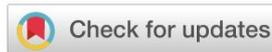


## The responsibility of social media in times of societal and political manipulation

Ulrike Reisach 

Department of Information Management, Prof. Dr. Ulrike Reisach, Neu-Ulm University of Applied Sciences, Wiley-Street 1, D-89231 Neu-Ulm, Germany

Received 17 December 2019, Accepted 16 September 2020, Available online 22 September 2020.



Show less 

+ Add to Mendeley  Share  Cite

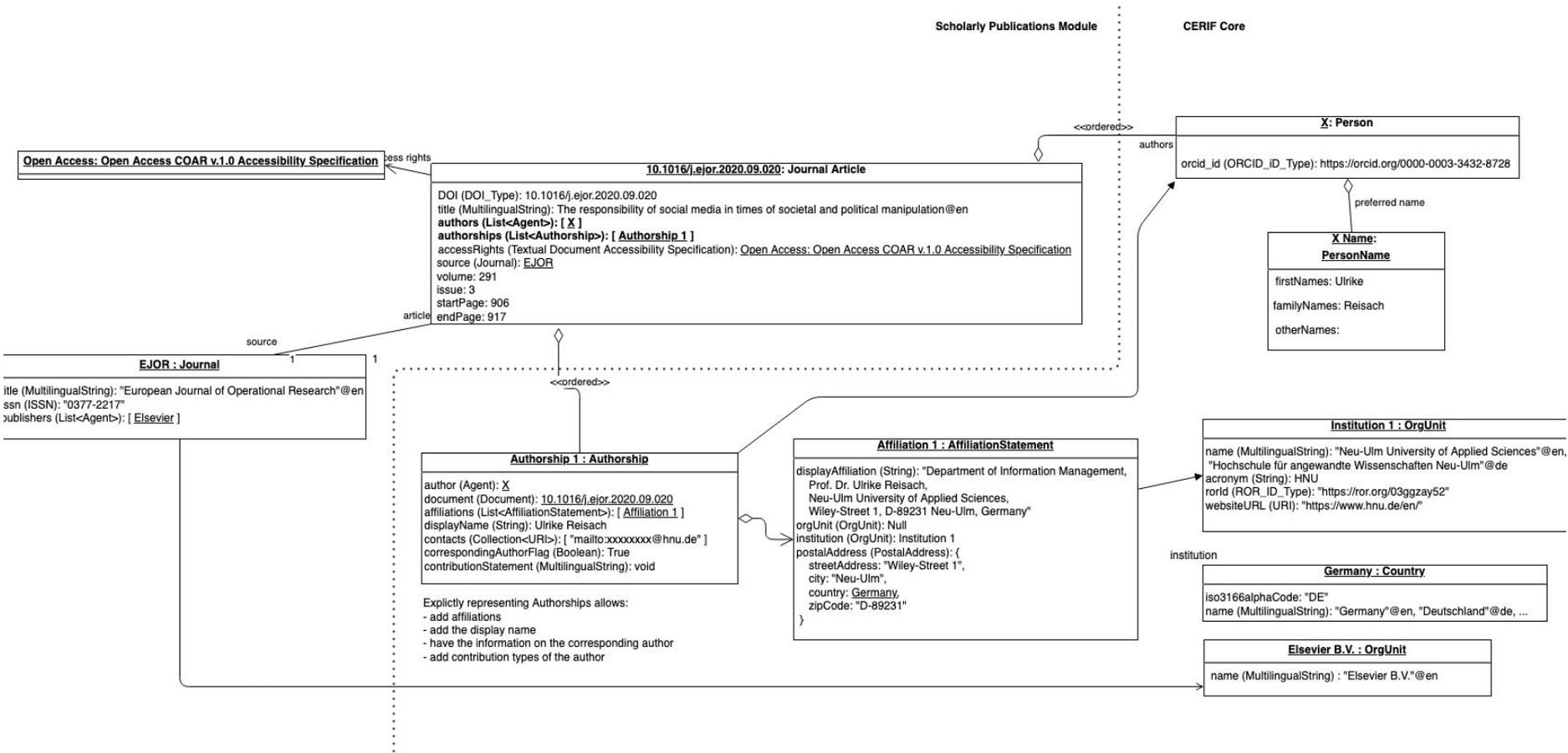
<https://doi.org/10.1016/j.ejor.2020.09.020>

Under a Creative Commons [license](#)

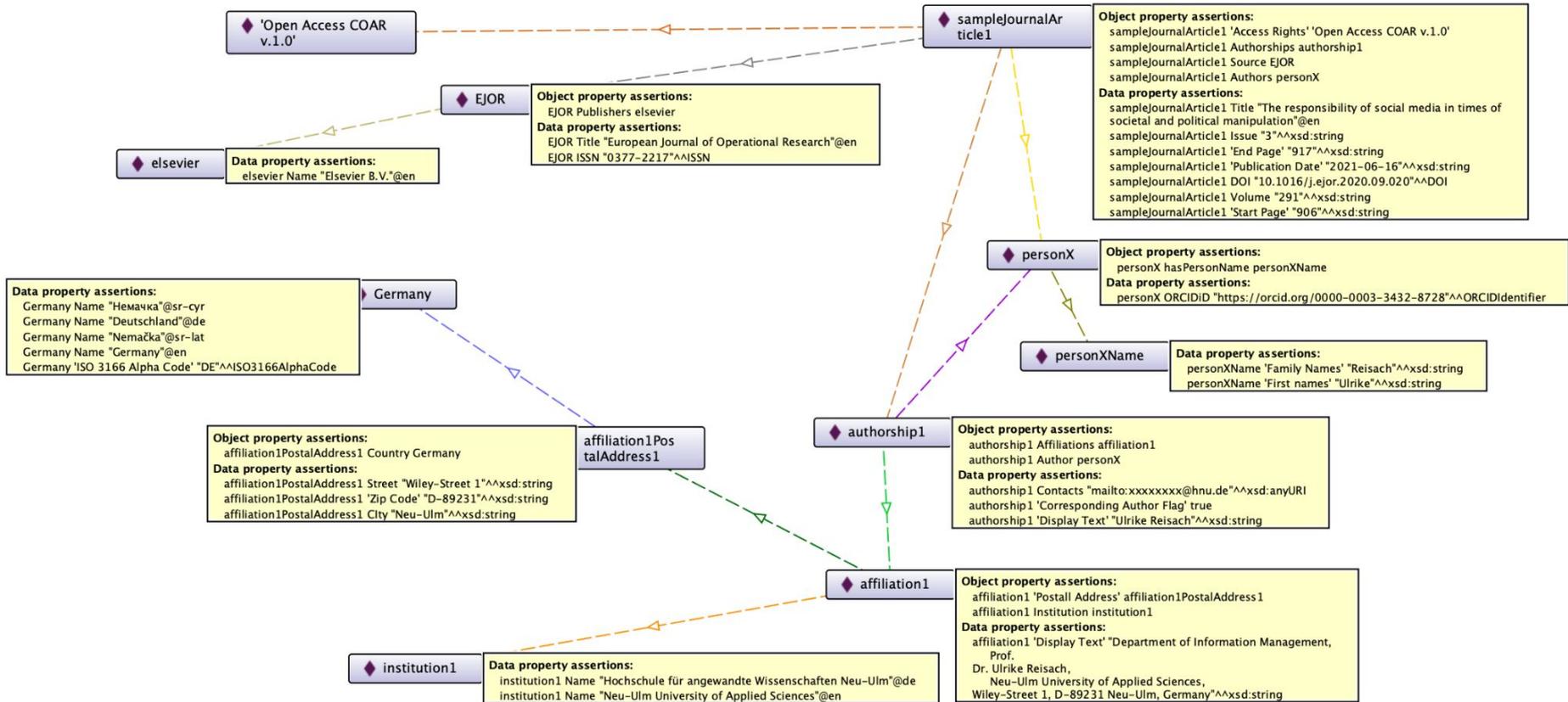
[Get rights and content](#)

[open access](#)

# Example publication expressed in UML

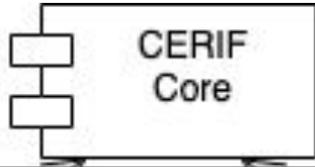


# Example publication expressed in RDF

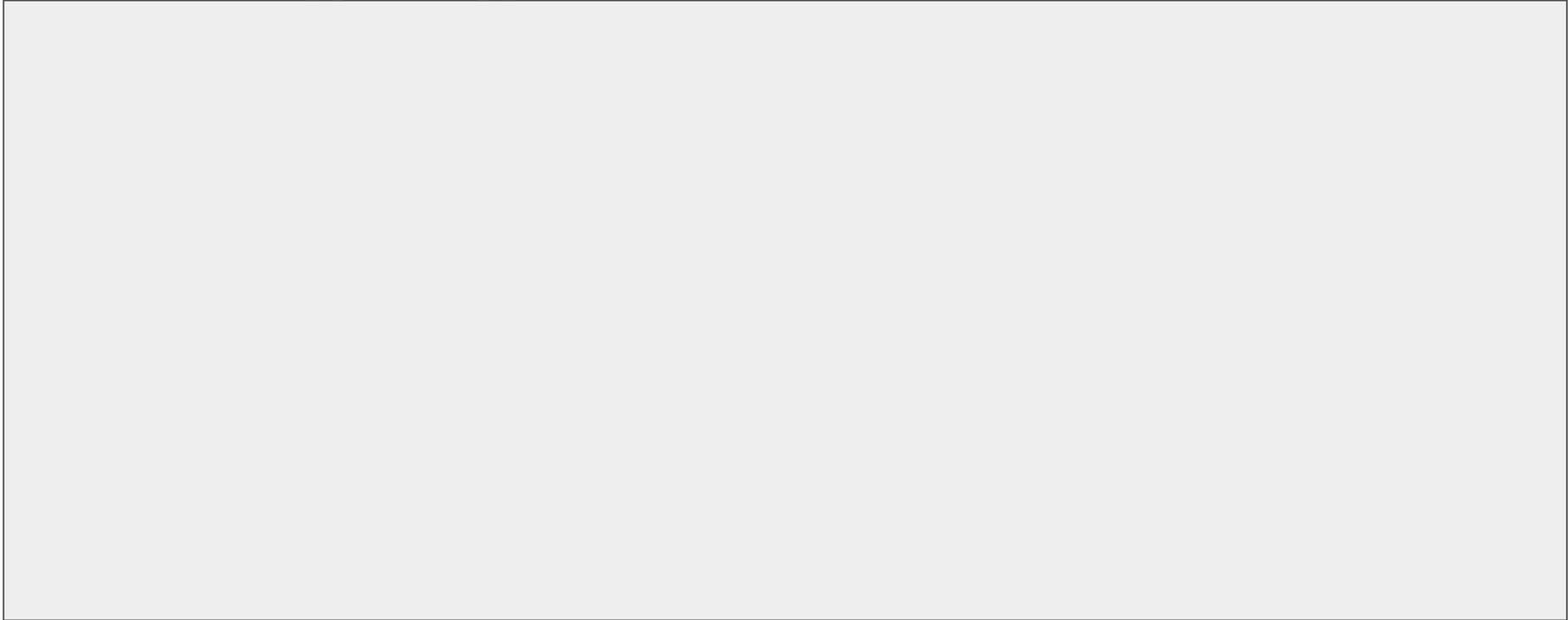


# Modular Architecture

# CERIF Core and Modules Architecture



Always present



# A prototype CERIF Core repository on Github

Currently for internal review

The screenshot shows the GitHub interface for the repository 'EuroCRIS / CERIF-Core'. The repository is private and has 4 unwatchers, 0 stars, and 0 forks. The main branch is 'main' with 2 branches and 0 tags. The repository was last updated 6 days ago with 80 commits. The file list includes 'datatypes', 'entities', 'tools', and 'README.md', all updated 6 days ago. The README.md file is open, showing the title 'CERIF Core' and the text: 'This is the Core of CERIF (=Common European Research Information Format), the result of the CERIF Refactoring Pilot project started in 2021 by euroCRIS.' The status section indicates '(2021-06-24) This is experimental.' The right sidebar shows the repository is about the CERIF Refactoring Pilot project, started in 2021 by euroCRIS. There are no releases or packages published yet. Two contributors are listed: chenejac and jdvorak001 (Jan Dvorak).

https://github.com/EuroCRIS/CERIF-Core

<https://github.com/EuroCRIS/CERIF-Core>