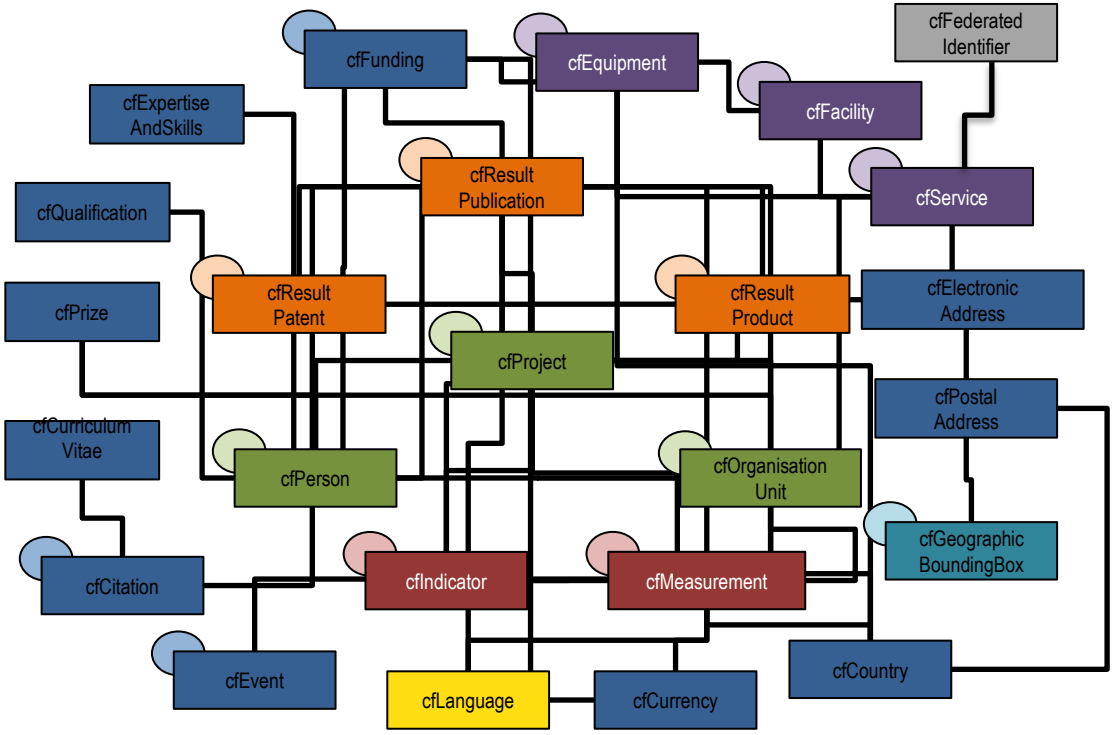
CERIF Tutorial



Dragan Ivanović

May 19th, 2026

CRIS 2026, Azores, Portugal

*This deck of slides is based on the CERIF Tutorial by **Brigitte Jörg**, CERIF TG Leader 2004–2012, and contains slides by **Jan Dvořák**, CERIF TG Leader 2013-2025, and **Valérie Brasse**, euroCRIS Executive for projects 2011–2015.*

Dragan Ivanović

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

dragan.ivanovic@uns.ac.rs

euroCRIS

- CERIF TG Deputy Leader since 2020
- euroCRIS Board member since 2023

University of Novi Sad, Faculty of Technical Sciences

- Researcher & Lecturer & Research software engineer

Lyrasis

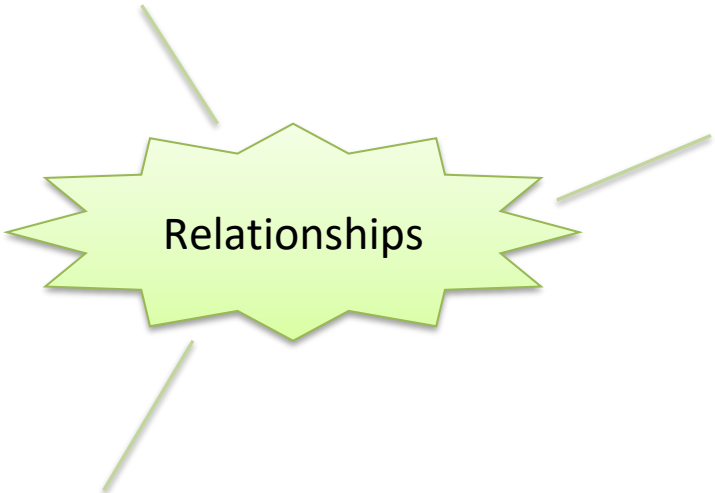
- VIVO tech lead since 2021

RESEARCH INFORMATION

Research Information

The process of research

- Research projects
- Funding
- Research infrastructures



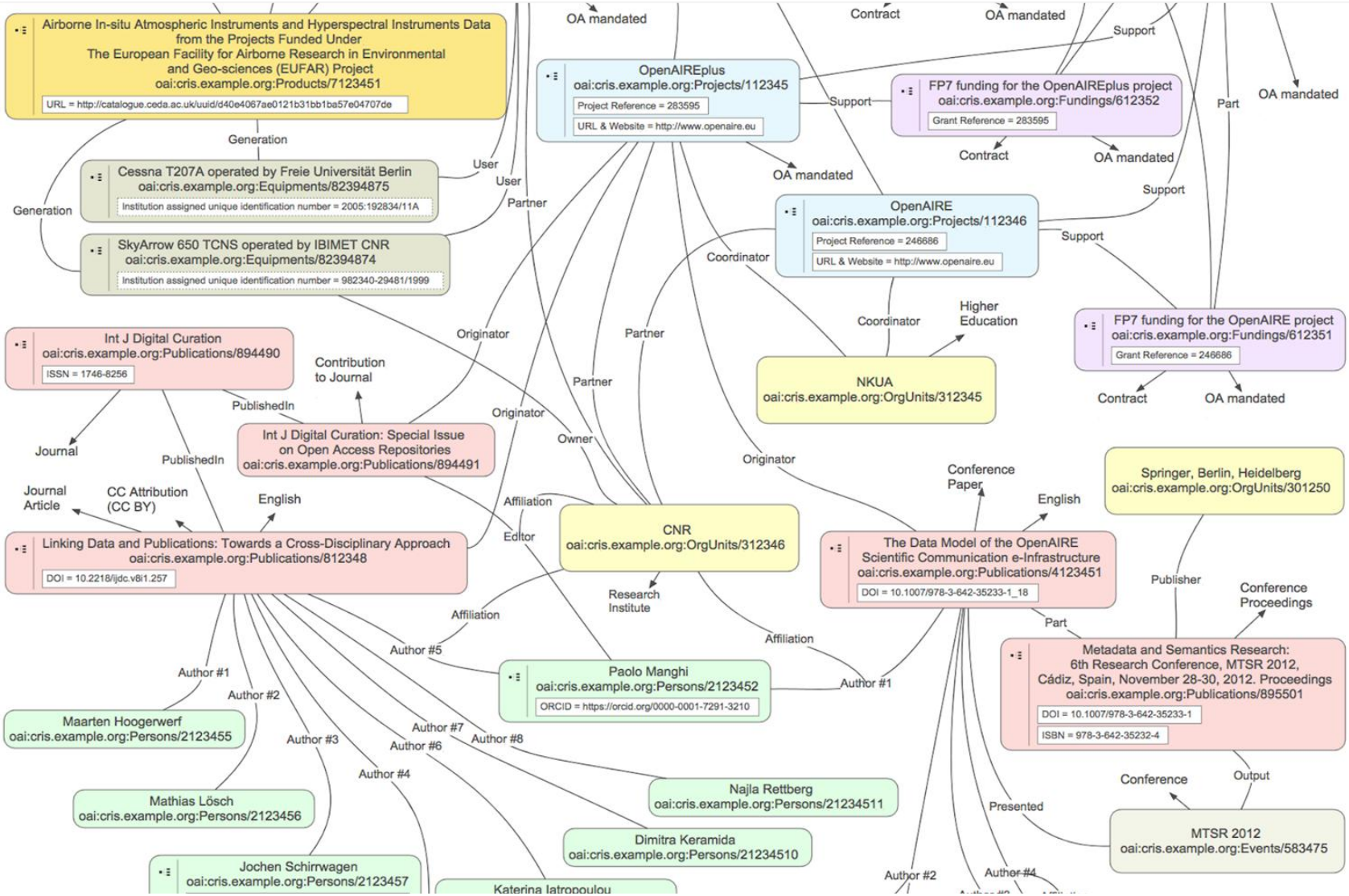
The research actors

- Researchers
- Institutions
- Funders
- Publishers
- Facility operators
- Associations

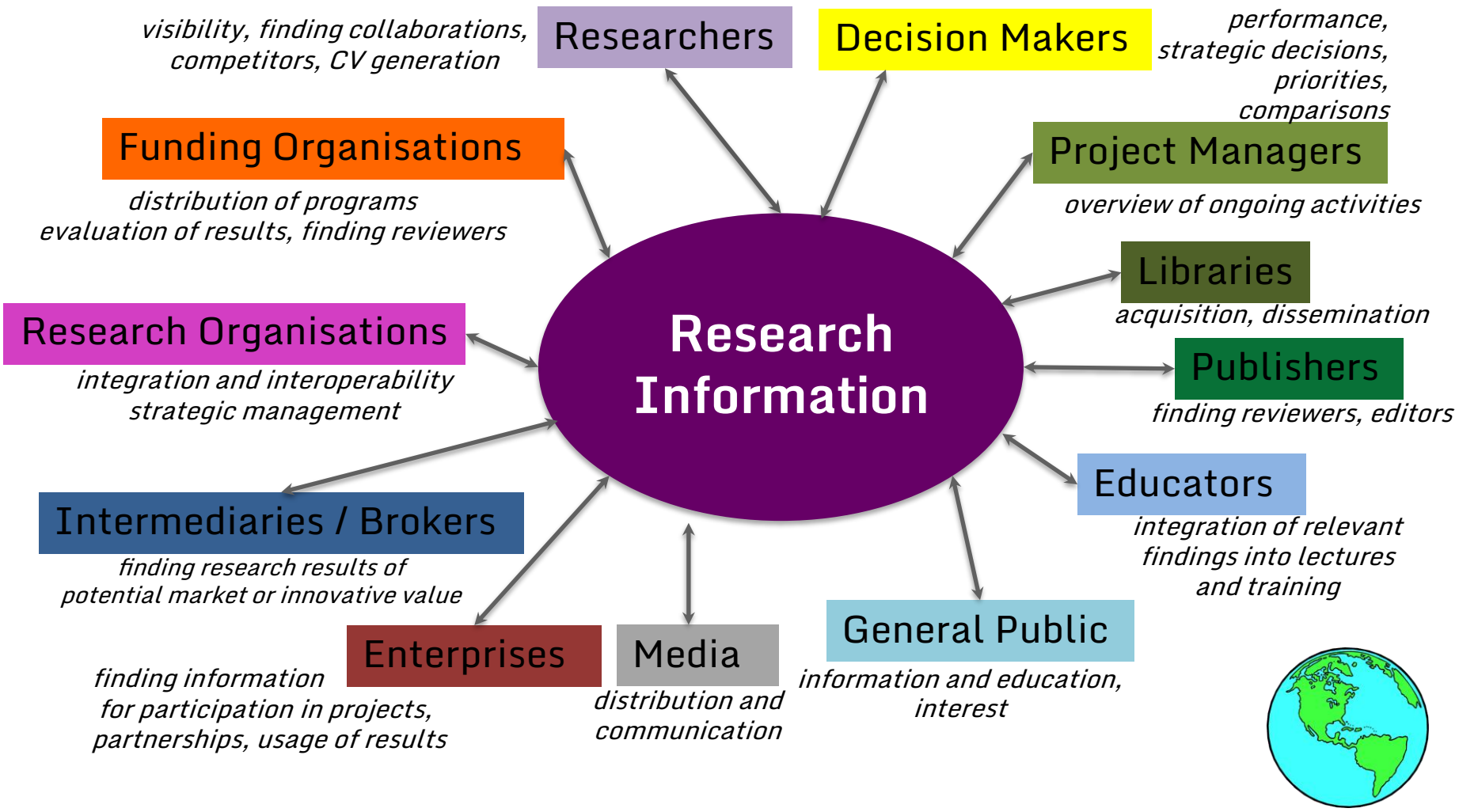
Research results

- Outputs (Publications, Research Datasets, Patents, …)
- Outcomes, Impacts

Research Information is heavily interlinked



Who needs Research Information?



Different viewpoints on Research Information

Information perspective	Useful for
Researcher, his/her activities, results, …	Researcher profile / CV system
Research Project, its consortium, team, funding, results, …	Project webpage / report
Organisational Unit, its activities, results, ...	Unit webpage / report / evaluation report
Publication	Bibliometrics / panel evaluations
Research Facility	Tracking & reporting usage

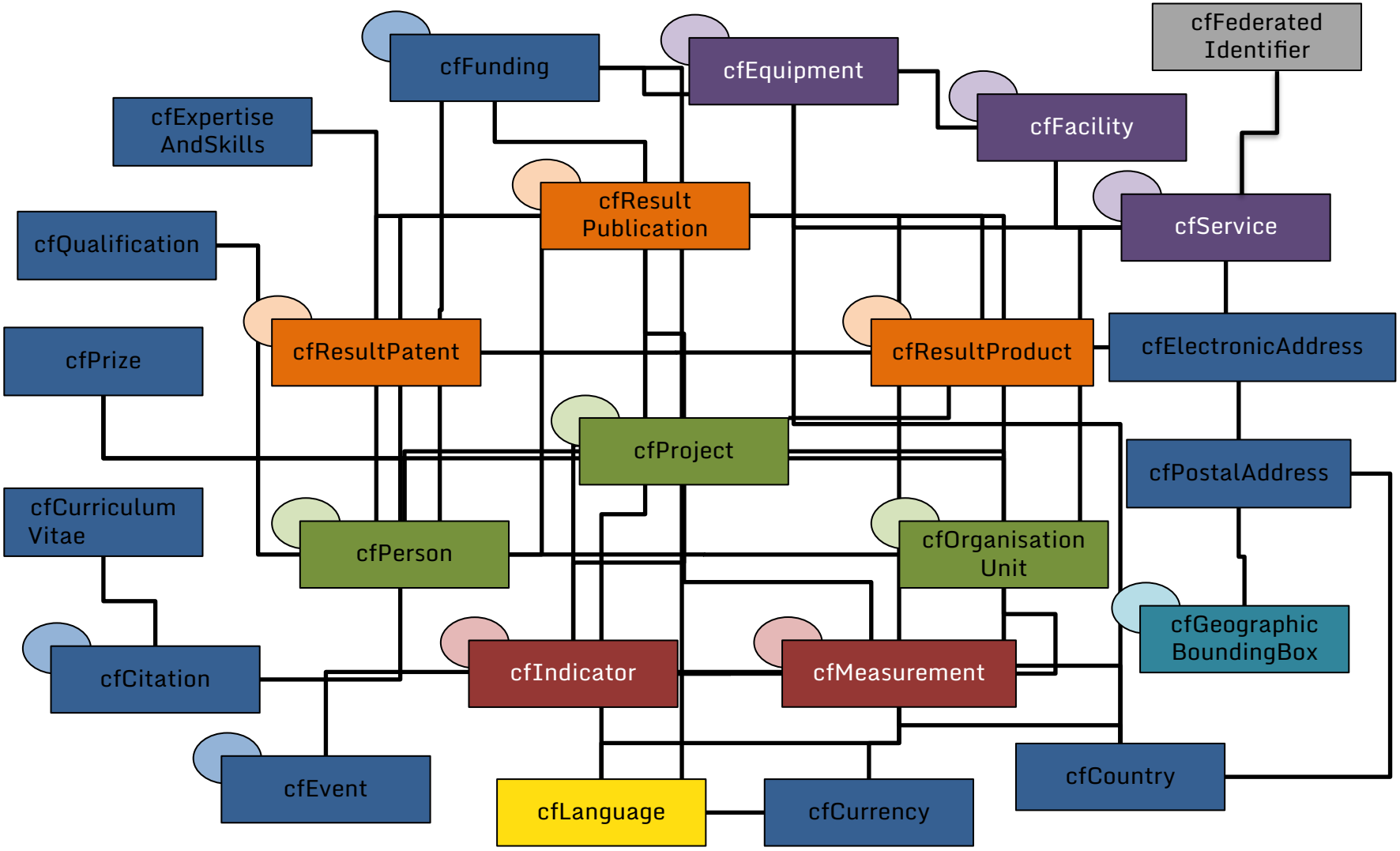
Common European Research Information Format

- CERIF is an EU Recommendation to Member States
- The European Commission (EC) has authorised euroCRIS to maintain and develop CERIF and its usage
  - By the CERIF Task Group of euroCRIS
  - GitHub repository

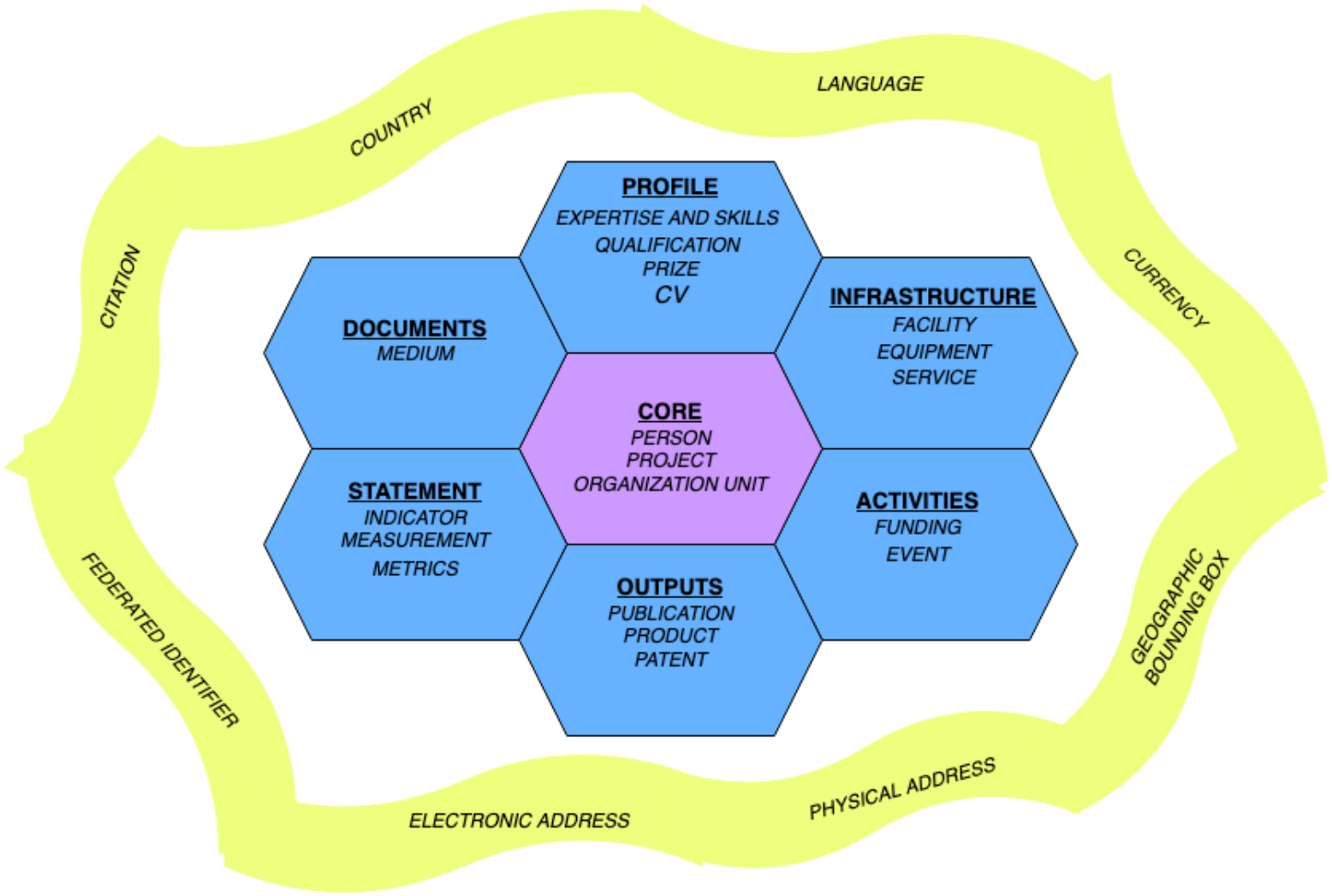
Current Research Information System

- Ongoing
- Past, of current interest
- Currently planned or decided upon

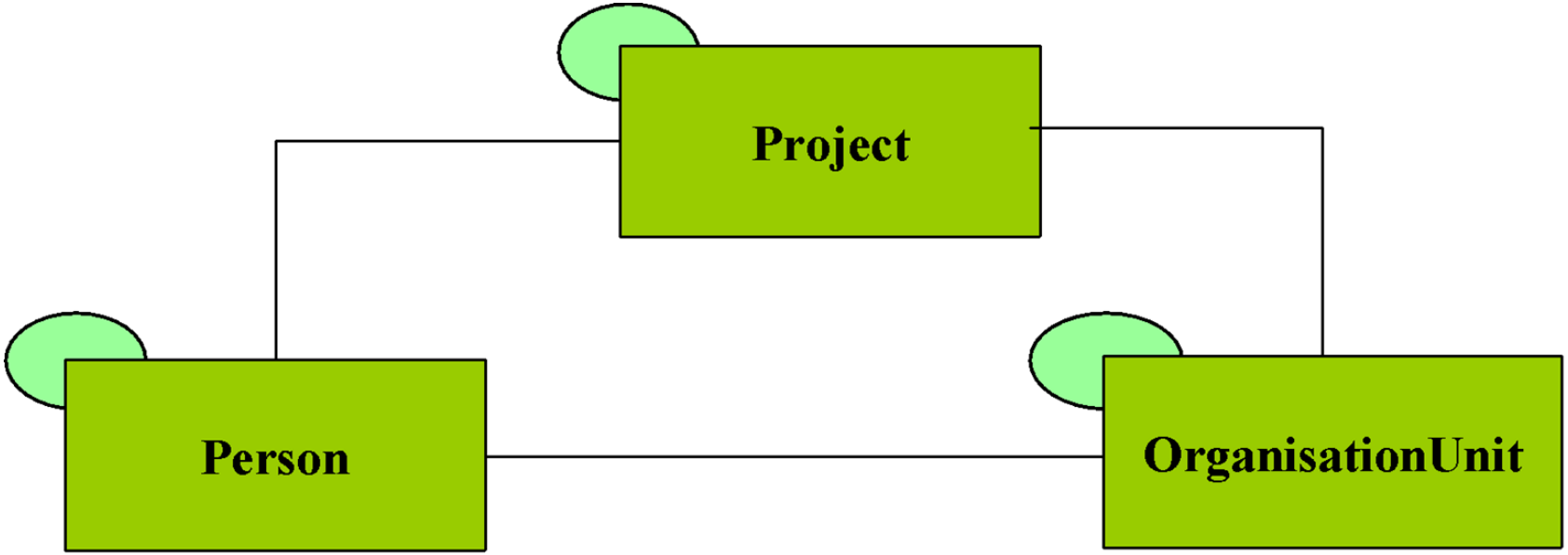
Common European Research Information Format



Common European Research Information Format



CERIF Base Entities



Project definition

A temporary endeavor undertaken to create a unique product, service or result.
Source: the Project Management Institute, https://www.pmi.org/about/learn-about-pmi/what-is-project-management

In the research information domain, one typically tracks:
(1) **research** projects, where the result is an addition to the body of knowledge of the mankind,
(2) technology **development** projects, where the result is a particular technology or product,
(3) **innovation** projects, where the result is an improvement of a product or process, and
(4) projects that **create or enhance infrastructure** for research, technology development or innovation.

Depending on the scope one can also track finer levels of granularity: stages, work packages, sometimes even down to individual tasks. All such activities are also modelled using the Project entity and linked using the recursive link relationship.

The Project entity only captures details of the project scope and plan. Information about the resources needed to execute the project such as the funding (i.e., the grants received), the people and organisations involved, the supporting infrastructures, the outputs produced, etc. is contained in separate entities (the Funding entity, the Person entity, the OrgUnit entity, the infrastructure entities, the result entities respectively) and is linked to the Project.

Person definition

A human being as an individual.
Source: https://en.oxforddictionaries.com/definition/person

The kind of involvement of a Person in the research ecosystem is specified in the links with the organisations, the services, etc. This typically includes:
(1) **researchers** (Persons performing research in an Organisation Unit as employees or students);
(2) **authors** and **contributors** (Persons signing a publication, creators of data sets, software developers, etc.);
(3) **investigators** and project **participants** (Persons involved in a Project as principal investigators, co investigators, project managers, consultants, etc.);
(4) **management** (directors, rectors, deans, department heads, etc.);
(5) **support** staffs (technicians, responsible for Equipment, librarians and digital asset curators, administrative staff, etc.).
One Person typically has many of these relationships.

OrgUnit definition

Organisation Unit: an organisation, a unit therein, a committee or any other group of people that has a collective goal. Organisation Units are not necessarily formalized as legal entities.

In the research information domain Organisation Units typically represents:
(1) organisations that **perform research** (universities, research institutes, corporations) and their subdivisions (faculties, schools, departments, research groups) and other associated bodies (boards, advisory bodies);
(2) organisations that **fund research** (funders, their divisions and evaluation panels);
(3) scientific **associations and networks**;
(4) **publishers**, facility **operators** and other **service providers** in the research space;
(5) **authorities**, such as patent offices and standardization or supervision bodies; and
(6) other bodies: editorial **boards**, evaluation **panels**, or **committees** of all kinds.

Project webpage

An acronym

A code (identifier), for ex a Grant number

A name or title

A (planned) start date

A (planned) end date or duration

[A source of funding]

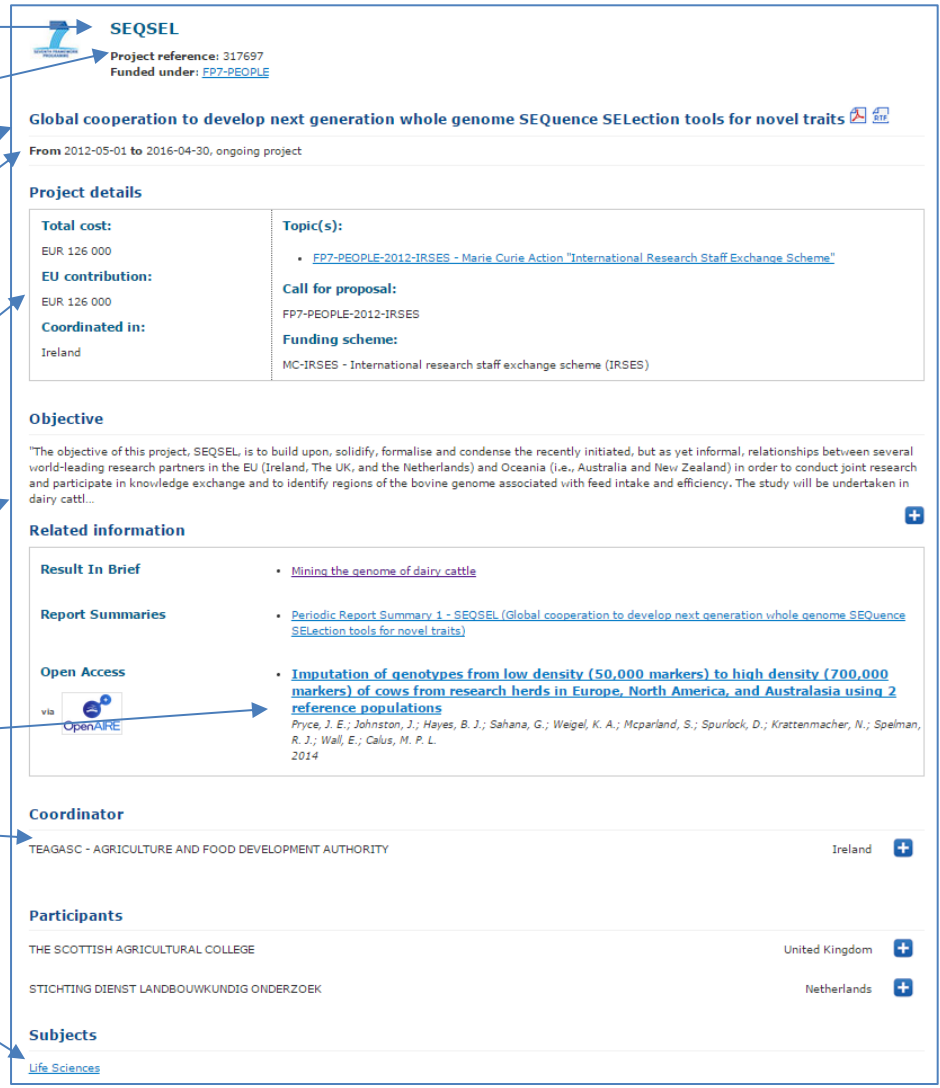
A short or long description (abstract)

[A few scientific publications]

[A project coordinator]

A web page (URI)

[A research domain]



Source: http://cordis.europa.eu/project/rcn/106635_en.html

A name or title

A short or long description (abstract)

[A project coordinator]

[A source of funding]

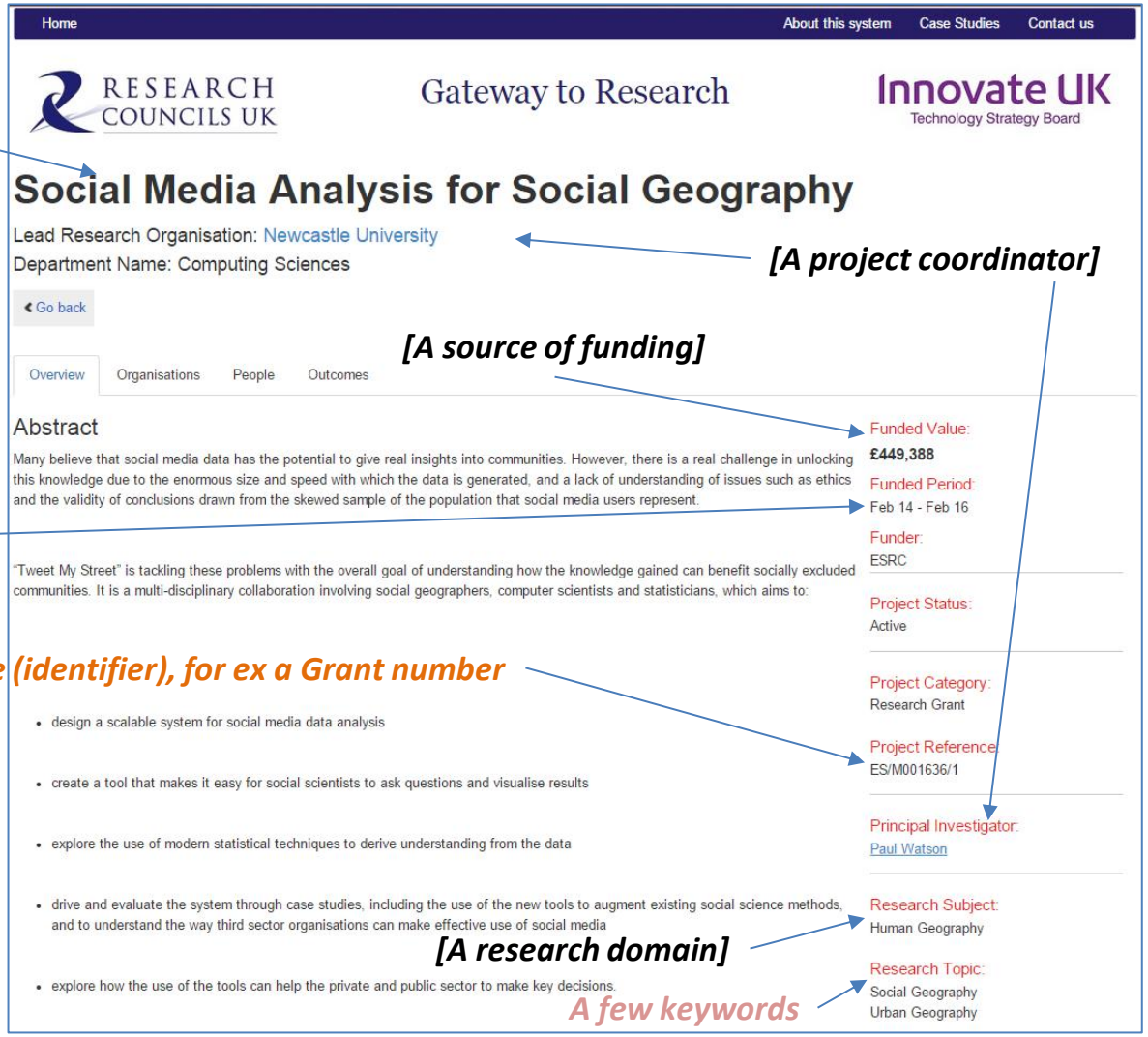
A (planned) start date
A (planned) end date or duration

A code (identifier), for ex a Grant number

A web page (URI)

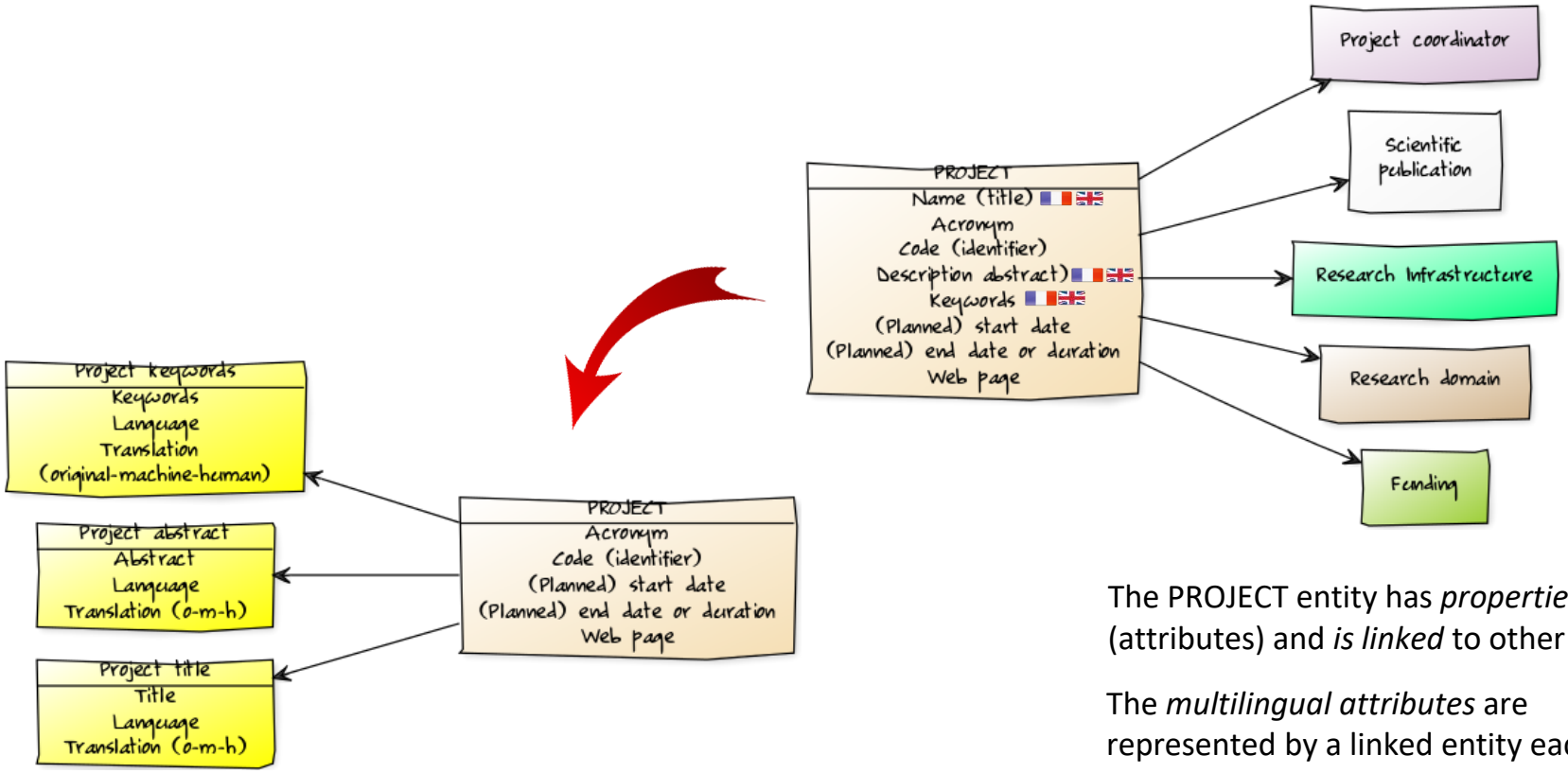
[A research domain]

A few keywords



Source: http://gtr.rcuk.ac.uk/project/A49CA721-687A-4D55-8FDF-9B60375B6EA8

Metadata for a Research Project



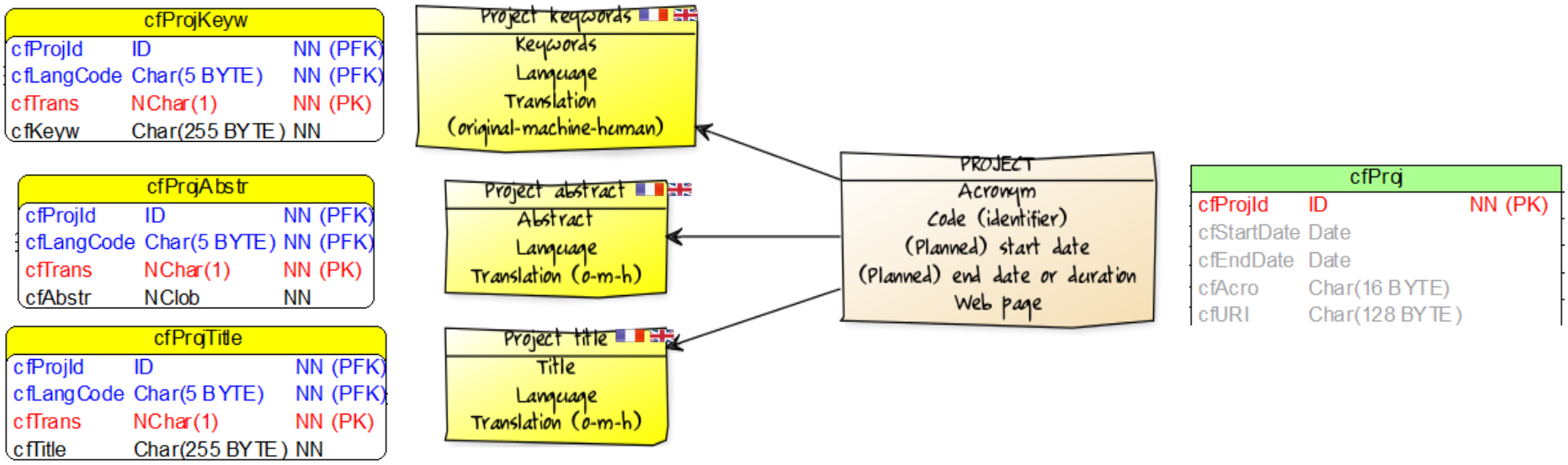
The PROJECT entity has *properties* (attributes) and *is linked* to other entities.

The *multilingual attributes* are represented by a linked entity each.

* "start date" and "end date" are deprecated in v1.6

Metadata for a Research Project

CERIF naming rule: in English, abbreviated, starting with *cf*
Example: Project title = cfProjTitle



Example in DB

CERIF Project example database record			
Data	**Attribute**	**Table**	**Type**
project-ist-world *	**cfProjId**	**cfProj**	**base**
IST World	**cfAcro**	**cfProj**	**base**
http://www.ist-world.org/	**cfURI**	**cfProj**	**base**
2005-04-01	**cfStartDate**	**cfProj**	**base**
2007-11-30	**cfEndDate**	**cfProj**	**base**
Knowledge Base for RTD Competencies in IST	**cfTitle**	**cfProjTitle**	**lang[en,o]**
Wissensbasis für RTD Kompetenzen im Bereich IST	**cfTitle**	**cfProjTitle**	**lang[de,h]**
IST, Research Information, NMS, Portal,	**cfKeyw**	**cfProjKeyw**	**lang**
The objective of the project is to set up and populate an information portal with innovative functionalities …	**cfAbstr**	**cfProjAbstr**	**lang**

cfProj

cfProjId (PK)	cfAcro	cfURI	cfStartDate	cfEndDate
project-ist-world	IST World	http://...	2005-04-01	2007-11-30

cfProjTitle, PK = cfProjID + cfLangCode + cfTrans

cfProjID (FK)	cfLangCode	cfTrans	cfTitle
project-ist-world	EN	O	Knowledge Base for RTD Competencies in IST
project-ist-world	DE	H	Wissensbasis für RTD Kompetenzen im Bereich IST

cfProjKeyw, PK = cfProjID + cfLangCode + cfTrans

cfProjID (FK)	cfLangCode	cfTrans	cfKeyw
project-ist-world	EN	O	IST, Research Information, NMS, Portal

cfProjAbstr, PK = cfProjID + cfLangCode + cfTrans

cfProjID (FK)	cfLangCode	cfTrans	cfAbstr
project-ist-world	EN	O	The objective of the project is to set...

Source: http://www.eurocris.org/Uploads/Web%20pages/CERIF-1.3/Specifications/CERIF1.3_FDM.pdf

We have seen how to represent, store or exchange metadata about research projects.

Similarly:

- What characterises a person (researcher, Ph.D.,…)?
- What characterises an organisation (research laboratory, institute,…)?

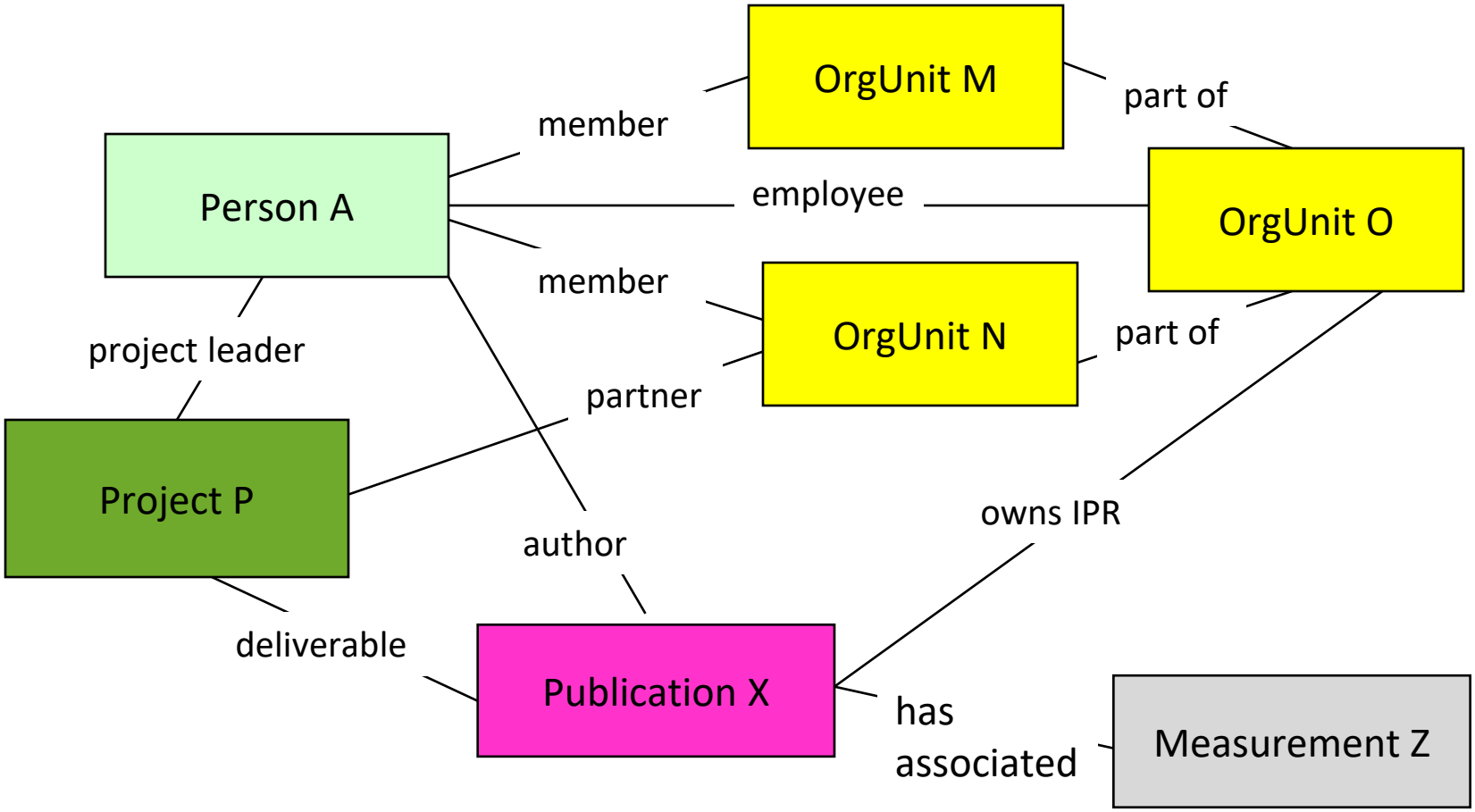
Source: https://pixabay.com/en/chemistry-teacher-science-1027781/

CERIF General Pattern

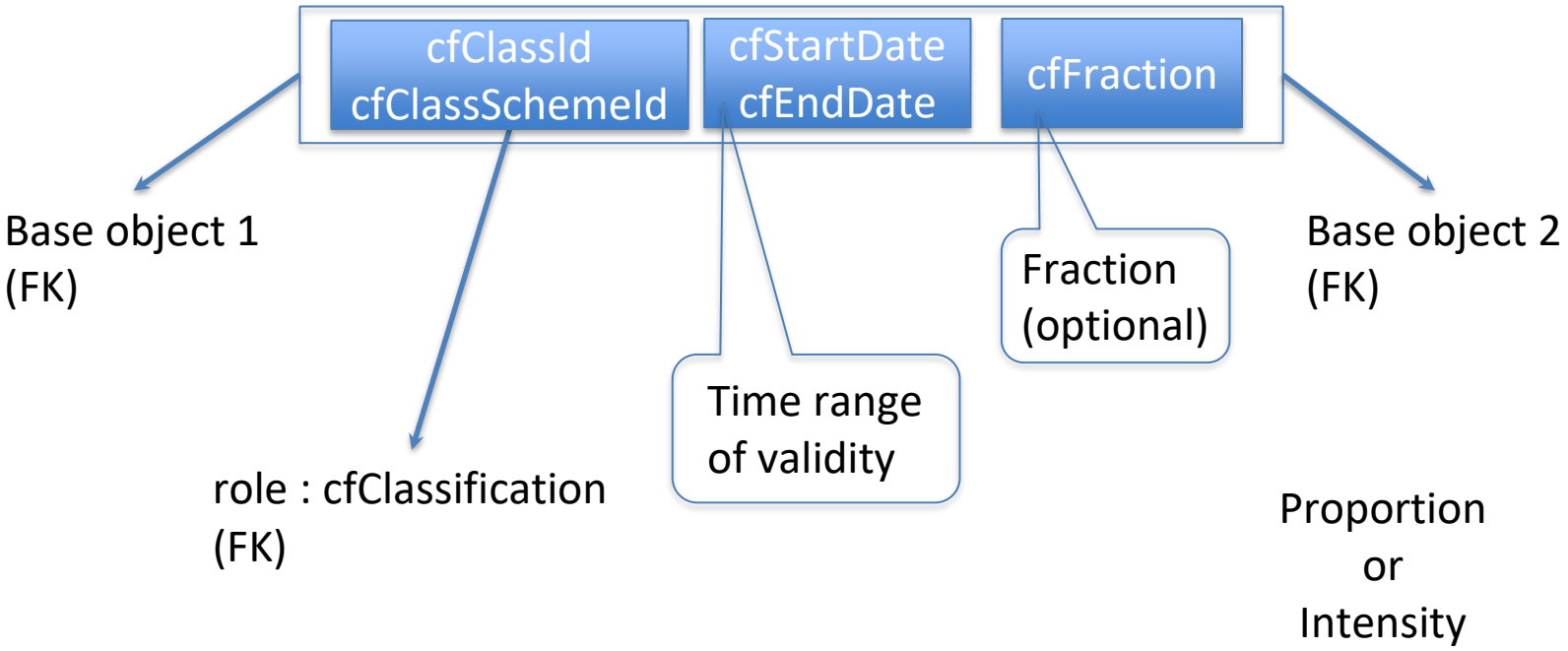
A typical CERIF entity has:

- Identifier (internal)
- Attributes
  - the basic ones
  - the multi-lingual ones
- Links
  - to other entities
  - recursive, a.k.a. self-referential
- External Identifiers
- Classifications (semantic layer)
  - Type
  - Status
  - Subject area
  - + other

CERIF link entities
(slide by Keith Jeffery)

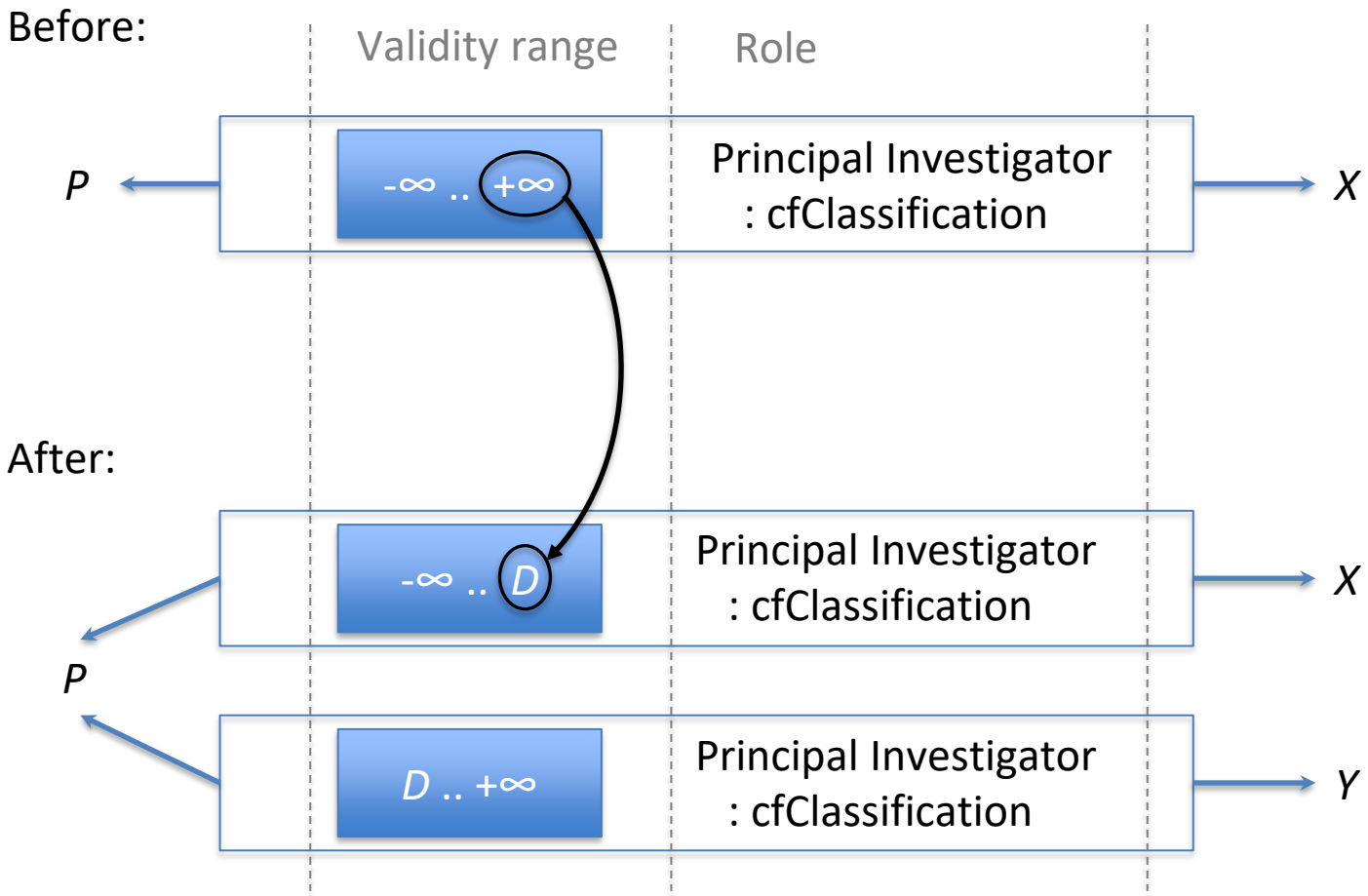


Generic Linking Entity Structure



Recording Change in CERIF

Example: The Principal Investigator of project *P* changes: *X* is replaced by *Y* effective date *D*.

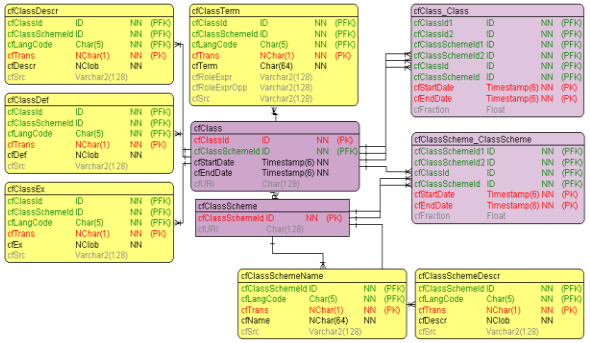


CERIF Semantic Layer

Central place to store declared semantic classifications

Allows to capture any Schema or Structure

- Flat Lists
- Thesauri
- Classification Systems (e.g. SKOS, …)
- Taxonomies
- Ontologies



Open / Extensible in all directions

- New Schemas
- New Concepts / Terms
- New Relationships

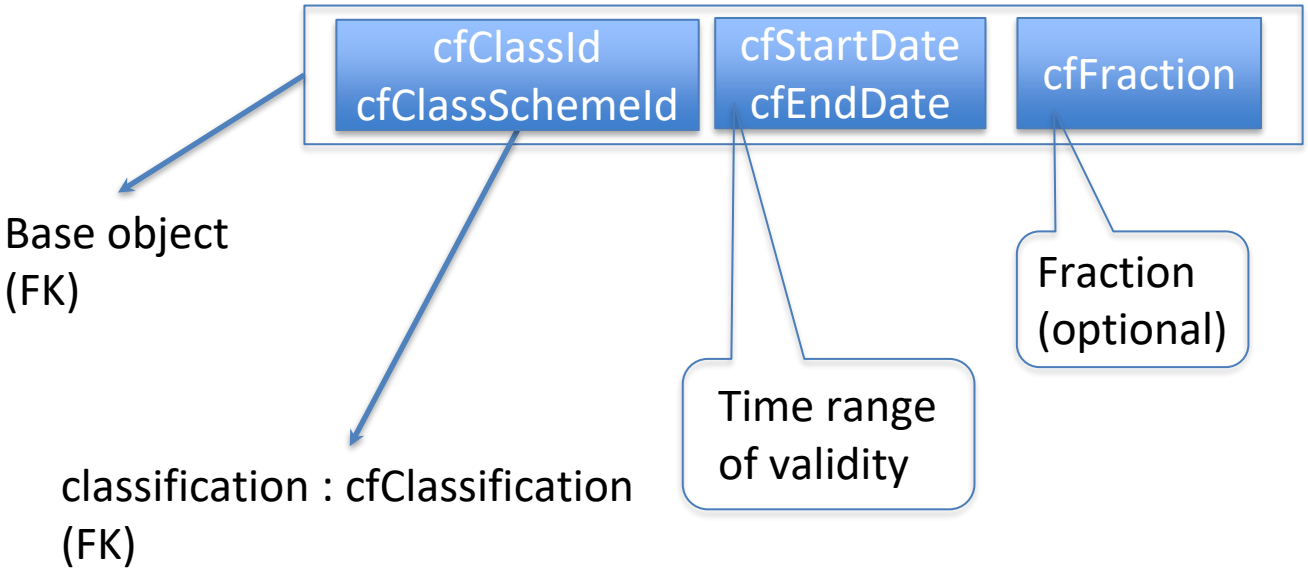
Enables to manage

- Roles, Types, Statuses, … Semantics
- Subject Headings
- Versioning and archiving (start+end dates)

Allows for Mappings between Schemes

- skos:closeMatch, skos:exactMatch or any other mapping you need

Generic Classification Link Structure



Any aspect that can be expressed by a controlled vocabulary:

- Type
- Status
- Subject area
- License
- …

CERIF Federated Identifiers

- Records the “tag” by which an object is known elsewhere
- For any CERIF research entity
- “Identifier Types” classification scheme
- (optionally) Connects to a Service representing the issuer of the identifier
    - Usually an information system

CERIF Federated Identifiers: Examples

Publication

- DOI
- PubMed Central ID
- WoS UT, Scopus EID
- ISBN, ISSN

Person

- ORCID iD
- ISNI
- ResearcherID, Scopus Author ID
- IdRef, DAI, Número Lattes
- Social Security Number or another administrative identifier
- Staff Id in an HR system

Project/Grant

- Funder's reference number
- Organisation's reference number

Organisation

- ROR ID
- FundRefID
- Wikidata reference
- ISNI
- RingGold number
- Org Enhanced (WoS)
- VAT Identification Number or another administrative identifier

Classifications: External Codes

CERIF PROFILES

Interoperability

1. Agreed Semantics
2. Agreed Format
3. Agreed Protocol

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

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

CERIF Profiles

Useful subsets of CERIF
for specific research information
interchange scenarios

- Entities & attributes:
  - Profile ⊂ CERIF
- Semantic vocabularies:
  - Profile – specific choices
  - Sources: CERIF & beyond
- Integrity constraints:
  - Profile ⊃ CERIF

Profile data is CERIF

CERIF Profiles

Producers know what to include

Consumers know what to expect

OpenAIRE Guidelines for CRIS Managers 1.2

- Dvořák, Jan, Andreas Czerniak, & Dragan Ivanović. (2023). OpenAIRE Guidelines for CRIS Managers 1.2 (Version 1.2.0). Zenodo. DOI 10.5281/zenodo.8050936

- Document on the web https://openaire-guidelines-for-cris-managers.readthedocs.io/en/latest/index.html

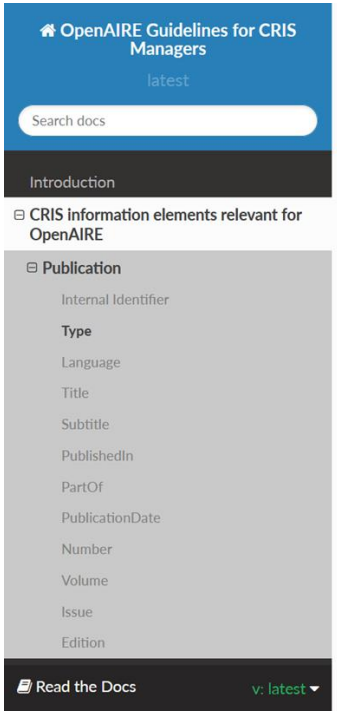
  - Introduction

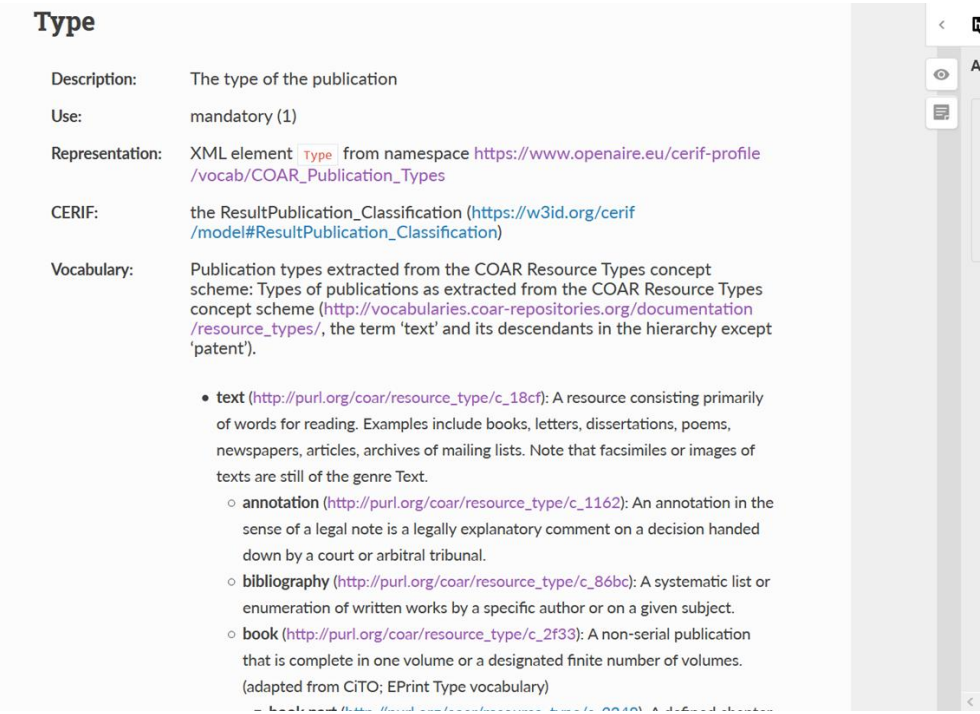
  - CRIS Information elements relevant for OpenAIRE
    - CERIF, CERIF XML

  - Technical implementation guidelines
    - CERIF XML as payload of OAI-PMH 2.0

Version 1.2: Aligned vocabularies

- COAR Resource Types (v. 3.1)
- COAR Access Rights
- ISSN Media List





Type

Description: The type of the publication

Use: mandatory (1)

Representation: XML element `Type` from namespace https://www.openaire.eu/cerif-profile/vocab/COAR_Publication_Types

CERIF: the ResultPublication_Classification (https://w3id.org/cerif/model#ResultPublication_Classification)

Vocabulary: Publication types extracted from the COAR Resource Types concept scheme: Types of publications as extracted from the COAR Resource Types concept scheme (http://vocabularies.coar-repositories.org/documentation/resource_types/, the term 'text' and its descendants in the hierarchy except 'patent').

- **text** (http://purl.org/coar/resource_type/c_18cf): A resource consisting primarily of words for reading. Examples include books, letters, dissertations, poems, newspapers, articles, archives of mailing lists. Note that facsimiles or images of texts are still of the genre Text.
  - **annotation** (http://purl.org/coar/resource_type/c_1162): An annotation in the sense of a legal note is a legally explanatory comment on a decision handed down by a court or arbitral tribunal.
  - **bibliography** (http://purl.org/coar/resource_type/c_86bc): A systematic list or enumeration of written works by a specific author or on a given subject.
  - **book** (http://purl.org/coar/resource_type/c_2f33): A non-serial publication that is complete in one volume or a designated finite number of volumes. (adapted from CiTO; EPrint Type vocabulary)

Information scope of the OpenAIRE Guidelines for CRIS Managers 1.2

The Core:

- The CRIS as a Service
- Organisations
- Researchers

Activities:

- Projects
- Funding
- Events

Outputs:

- Publications
- Datasets, Software & other Products
- Patents

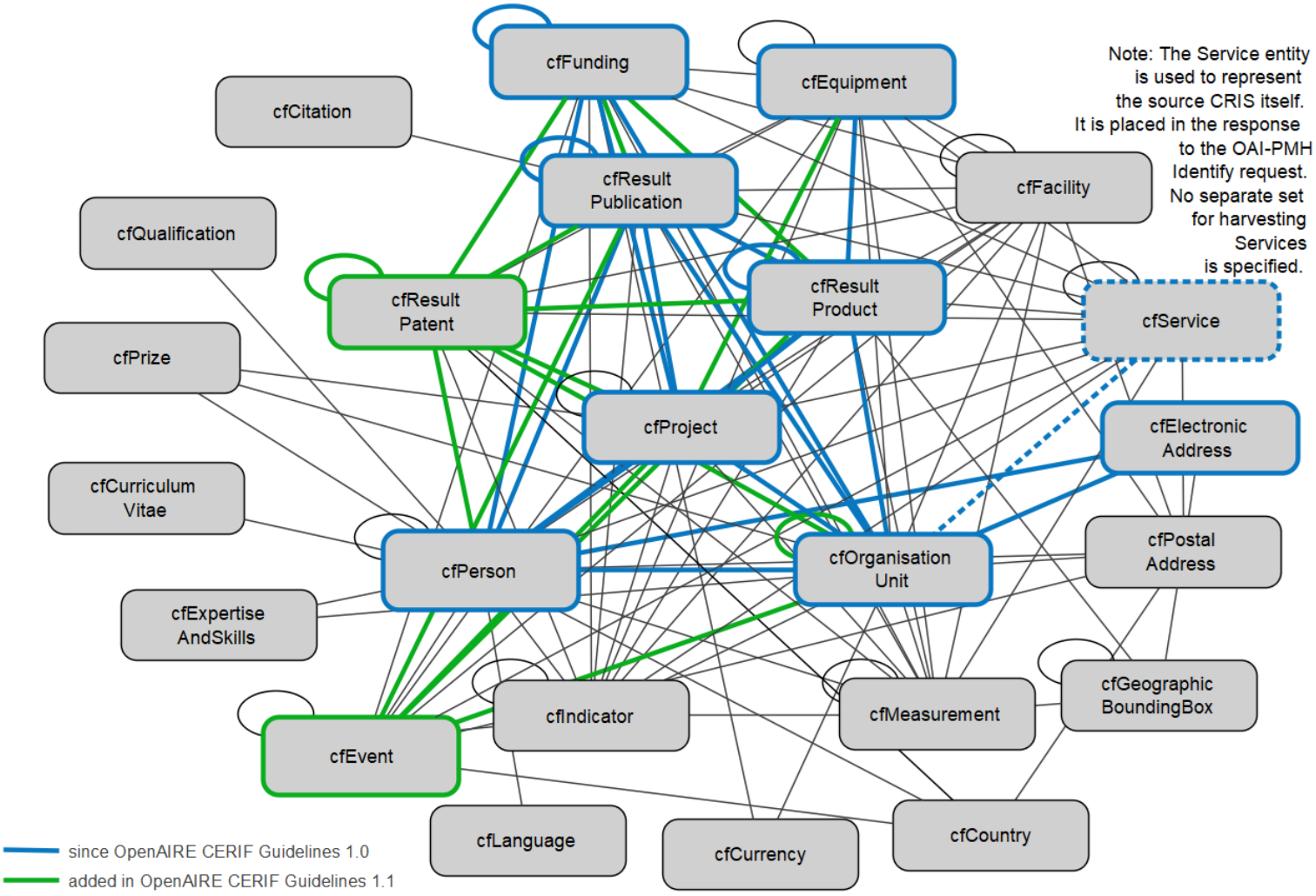
Infrastructure:

- Equipment

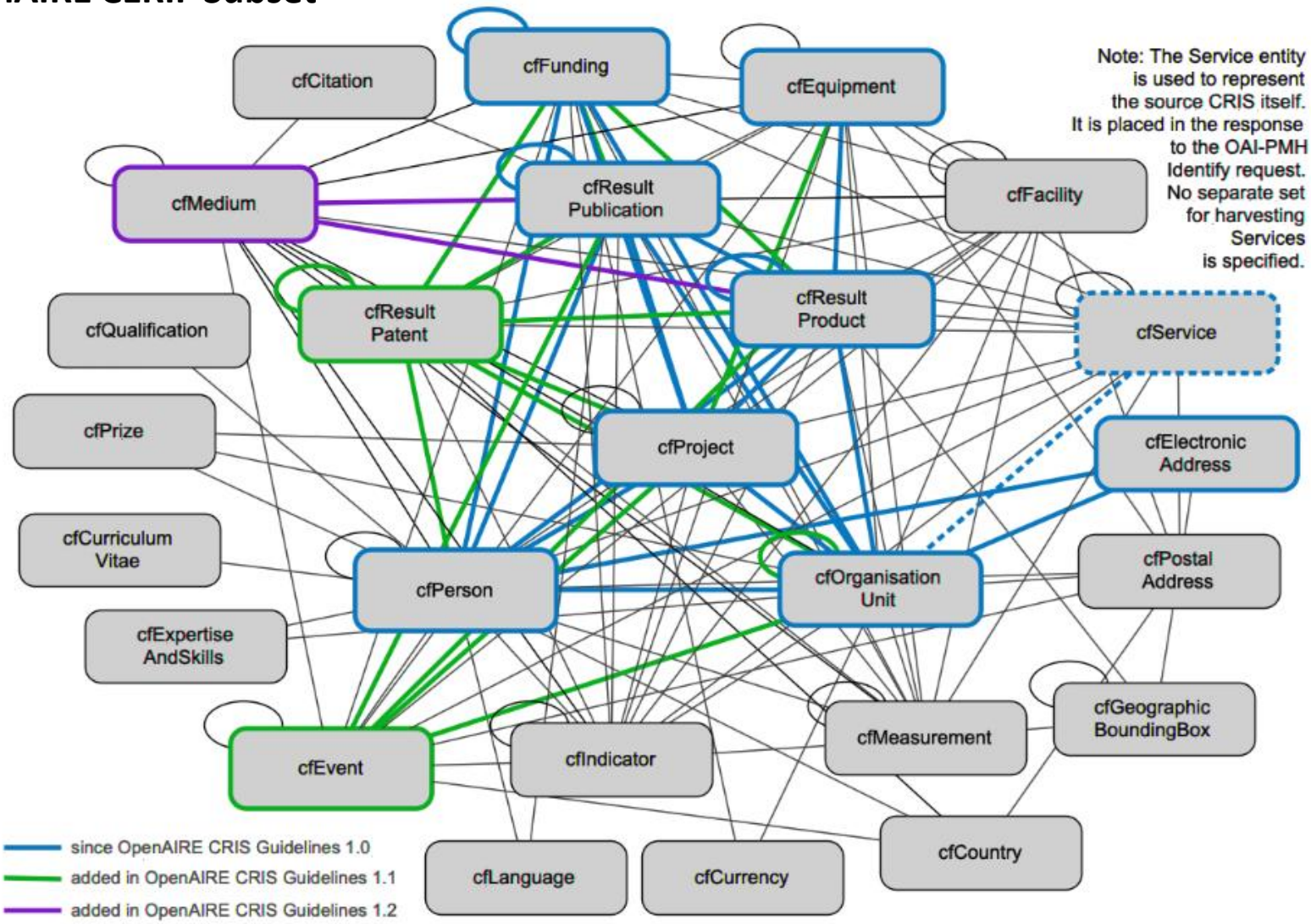
Two non top entities:

- Medium
- Electronic Address

OpenAIRE CERIF Subset



OpenAIRE CERIF Subset



Note: The Medium and the ElectronicAddress entities are not supported as top-level entities, no OAI-PMH sets are defined for them. They also do not expose their internal identiifiers .

Identifiers

- Publication:
  - DOI
  - Handle
  - PMCID
  - ISI Number (WoS Accession Number)
  - SCP Number
  - ISSN
  - ISBN
  - URL
  - URN
  - ZDB-ID
- Persons:
  - ORCID iD
  - ISNI
  - ResearcherID
  - Scopus Author ID
- Funding:
  - GrantDOI
- Products:
  - DOI
  - ARK
  - Handle
  - URL
  - URN

Plus a generic option

Supporting artifacts

- XML Schema
  - Namespace https://www.openaire.eu/cerif-profile/1.2/
  - A few Schematron integrity rules embedded
- A comprehensive set of examples
- Prototype validator

```
<Publication xmlns="https://www.openaire.eu/cerif-profile/1.2/" id="Publications/852734">
<!-- Linking Data and Publications: Towards a Cross-Disciplinary Approach →
    <Type xmlns="https://www.openaire.eu/cerif-profile/vocab/COAR_Publication_Types">
        http://purl.org/coar/resource_type/c_6501<!-- journal article --></Type>
    <Title xml:lang="en">Strong selection against hybrids maintains a narrow contact zone between morphologically cryptic
        lineages in a rainforest lizard</Title>
        <PublishedIn>
            <Publication id="Publications/893204"><!-- Evolution →
            <Type xmlns="https://www.openaire.eu/cerif-profile/vocab/COAR_Publication_Types">
                http://purl.org/coar/resource_type/c_0640<!-- journal --></Type>
                <Title xml:lang="en">Evolution</Title>
                <ISSN>1558-5646</ISSN>
                <Publishers>
                    <!-- [ … ] -->
                </Publishers>
            </Publication>
        </PublishedIn>
        <Volume>66</Volume>
        <Issue>5</Issue>
        <StartPage>1474</StartPage>
        <EndPage>1489</EndPage>
        <DOI>10.1111/J.1558-5646.2011.01539.X</DOI>
        <!-- [ … ] -->
</Publication>
```

```
<Publication xmlns="https://www.openaire.eu/cerif-profile/1.1/" id="852734">
        <!-- [ … ] -->
        <Authors>
                <Author>
                        <Person>
                                <PersonName>
                                        <FamilyNames>Singhal</FamilyNames>
                                        <FirstNames>Sonal</FirstNames>
                                </PersonName>
                        </Person>
                        <Affiliation>
                                <OrgUnit id="301248">
                                        <Name xml:lang="en">Museum of Vertebrate Zoology</Name>
                                        <PartOf>
                                                <OrgUnit id="329384">
                                                        <Name xml:lang="en">University of
California, Berkeley</Name>
                                                </OrgUnit>
                                        </PartOf>
                                </OrgUnit>
                        </Affiliation>  <!-- [ … ] -->
                </Author>  <!-- [ … ] -->
        </Authors>
        <!-- [ … ] -->
</Publication>
```

<Publication xmlns="*https://www.openaire.eu/cerif-profile/1.1/*" id="***852734***">

<!-- […] -->

<Keyword xml:lang="***en***">**hybrid zones**</Keyword>

<Keyword xml:lang="***en***">**phylogeography**</Keyword>

<Abstract xml:lang="***en***">**Phenotypically cryptic lineages comprise an important yet understudied part of biodiversity;** <!-- […] --></Abstract>

<References>

<Product id="***729487***">

<Type xmlns="*https://www.openaire.eu/cerif-profile/vocab/COAR_Product_Types*">**http://purl.org/coar/resource_type/c_ddb1**<!-- dataset --></Type>

<Name xml:lang="***en***">**Data from: Strong selection against hybrids maintains a narrow contact zone between morphologically cryptic lineages in a rainforest lizard**</Name>

<VersionInfo xml:lang="***en***">**1**</VersionInfo>

<DOI>**10.5061/DRYAD.4GH6HF5G**</DOI>

<!-- […] -->

</Product>

</References>

</Publication>

CERIF REFACTORING (2.X)

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

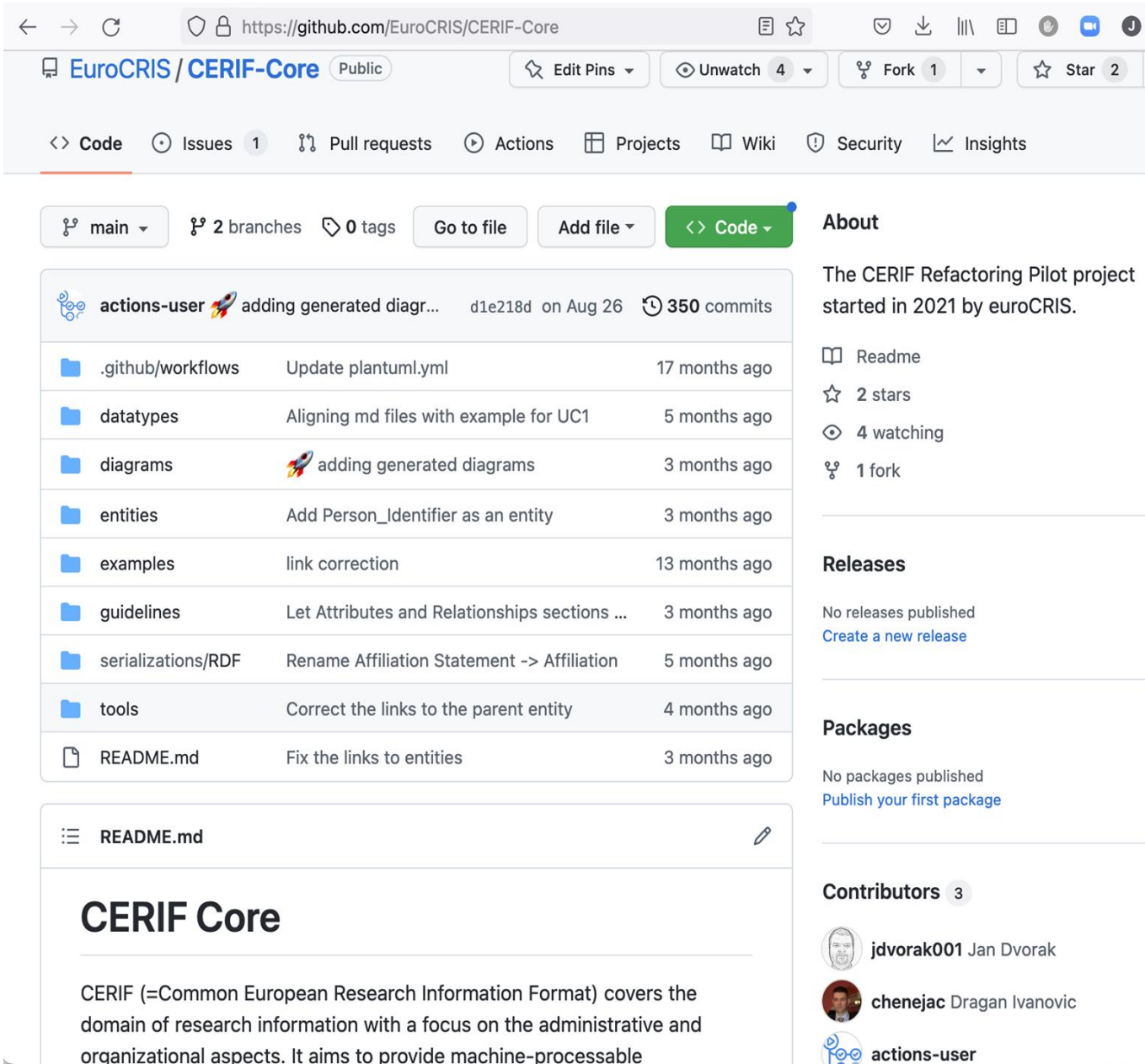
Contains:

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

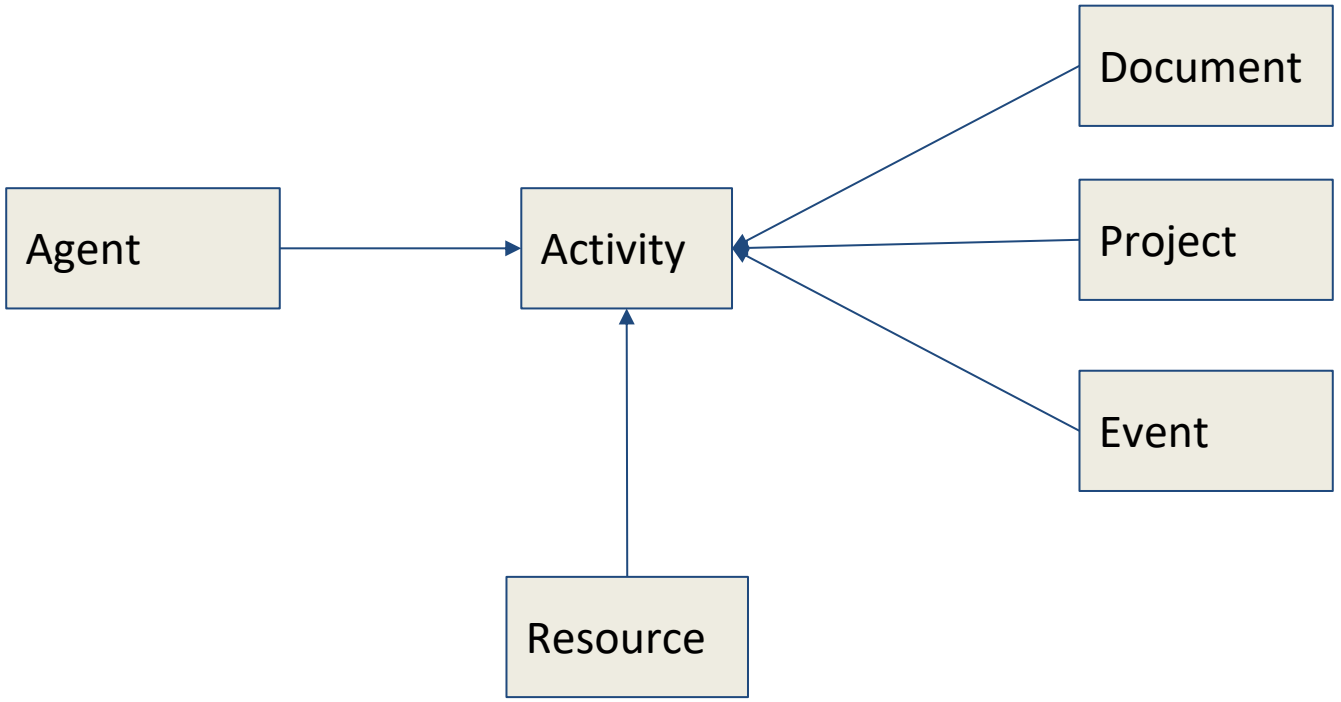
The basis for:

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

A prototype CERIF Core repository at https://github.com/EuroCRIS/CERIF-Core



CERIF Core "View from the Top"



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 agency
- Research software metadata
- Bibliometric indicators
- Research Infrastructure description

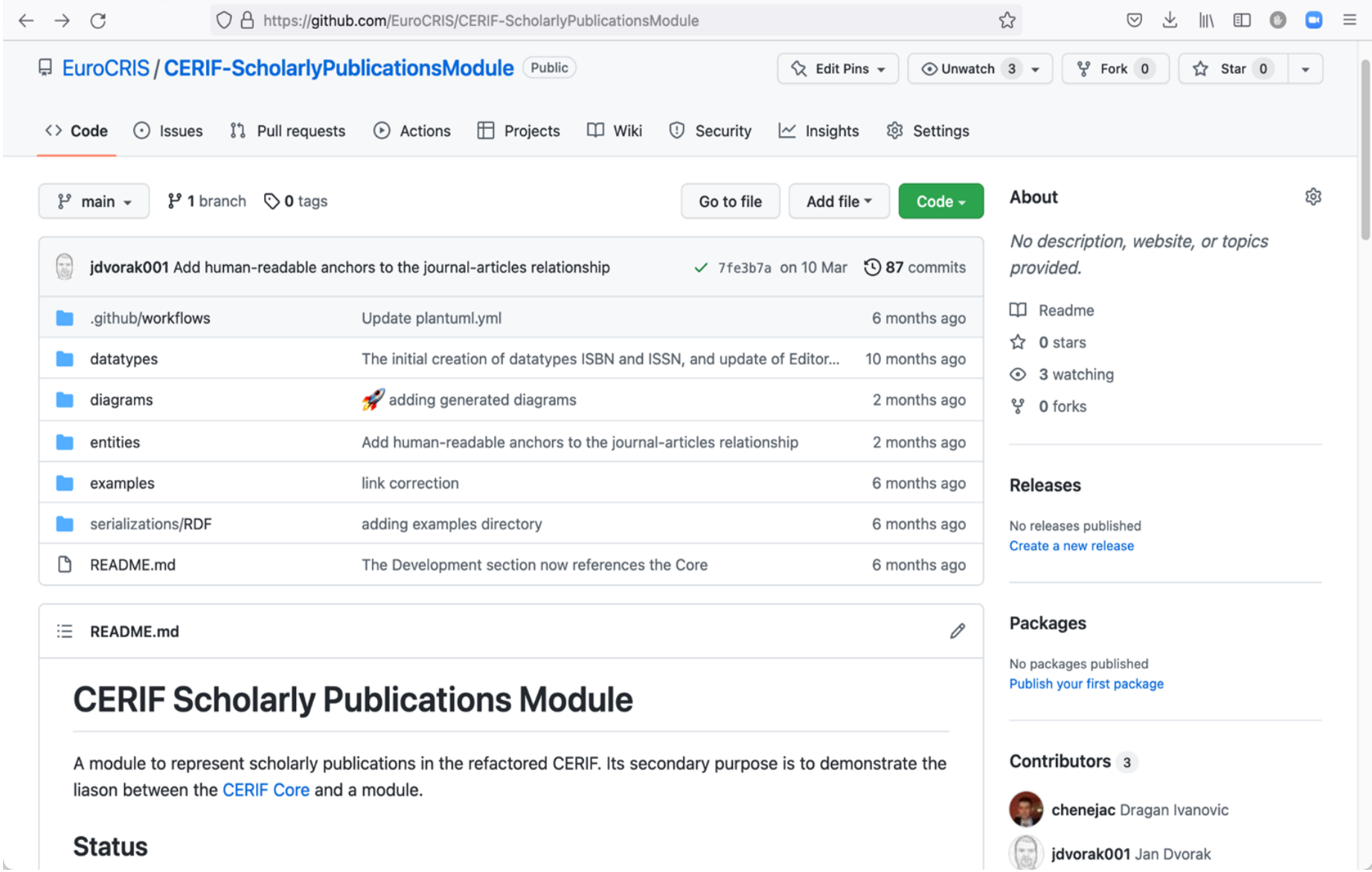
CERIF Scholarly Publication Module

1st module started along with the Core

Scope:

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

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



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

EuroCRIS / CERIF-ScholarlyPublicationsModule Public

Edit Pins | Unwatch 3 | Fork 0 | Star 0

Code | Issues | Pull requests | Actions | Projects | Wiki | Security | Insights | Settings

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

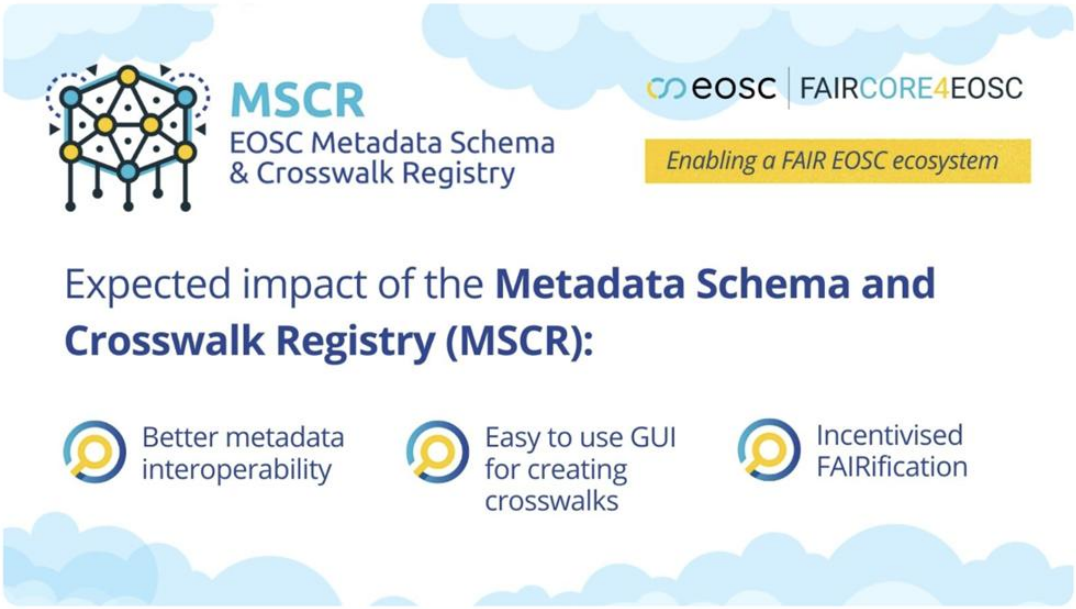
chenejac Dragan Ivanovic
jdvorak001 Jan Dvorak

Developed in the FAIRCORE4EOSC project

- CERIF as an intermediate model for Metadata Schema and Crosswalk Registry

https://faircore4eosc.eu/eosc-core-components/metadata-schema-and-crosswalk-registry-mscr

Metadata Schema and Crosswalk Registry (MSCR)



MSCR
EOSC Metadata Schema
& Crosswalk Registry

Related Case Studies:

- EOSC Service Providers & RDM Communities
- European Integration of National-level Services
- Mathematics
- Climate Change
- Social Sciences & Humanities

Overview

The MSCR allows registered users and communities to create, register and version schemas and crosswalks with PIDs. The published content can be searched, browsed and downloaded without restrictions. The MSCR also provides an API to facilitate the transformation of data from one schema to another via registered crosswalks.

Partners:

GRNET

Thank you!