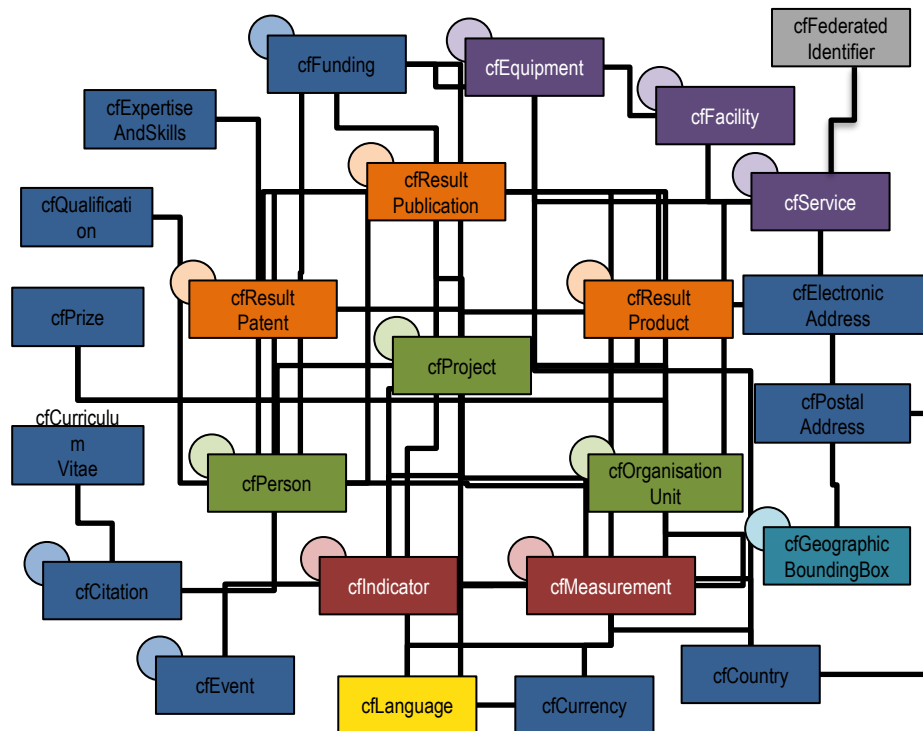


CERIF Tutorial



Jan Dvořák

May 13th, 2025

euroCRIS Membership Meeting
Leuven, Belgium

*This deck of slides is based on the CERIF Tutorial by **Brigitte Jörg**, CERIF TG Leader 2004–2012, and contains slides by **Valérie Brasse**, euroCRIS Executive for projects 2011–2015. It also contains slides co-developed by **Dragan Ivanović** for the CERIF Refactoring Project.*

Jan Dvořák

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



jan.dvorak@eurocris.org

euroCRIS

- CERIF TG Leader 2013–2023
- President since 2023

Charles University, Institute of Information Studies and Librarianship

- Researcher & Lecturer

Czech Technical University in Prague, Computing and Information Centre

- IS Analyst: V3S+EZOP – operations and further development of the in-house built institutional CRIS

InfoScience Praha (2004–2016)

- Leader of the team: Research, Development & Innovation Information System (the national CRIS for [CZ])
- In other roles for the system since 1997

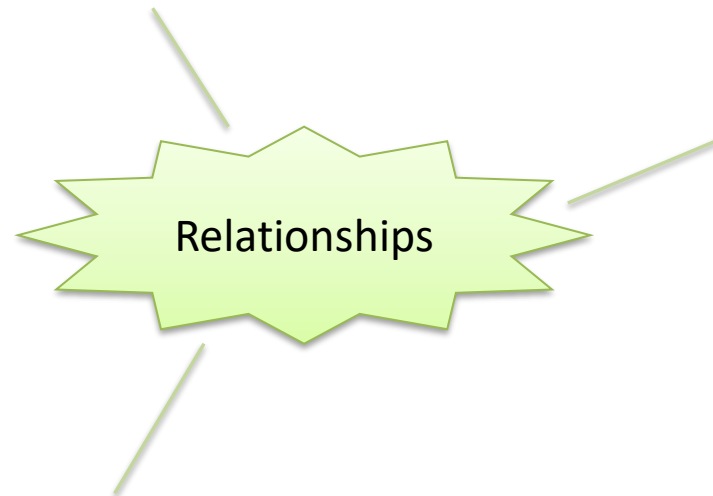
Research Information

The process of research

- Research projects
- Funding
- Research infrastructures

The research actors

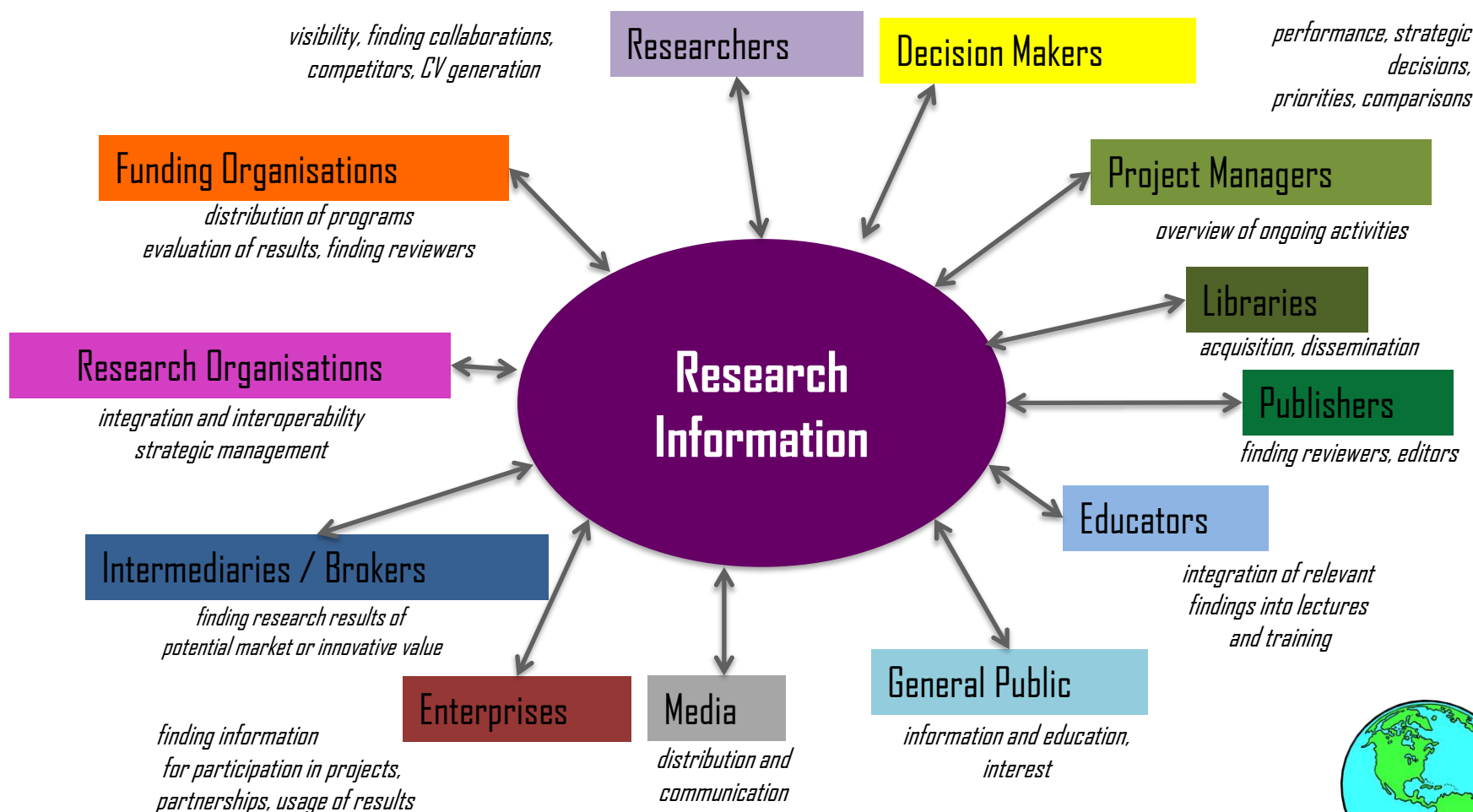
- Researchers
- Institutions
- Funders
- Publishers
- Facility operators
- Associations



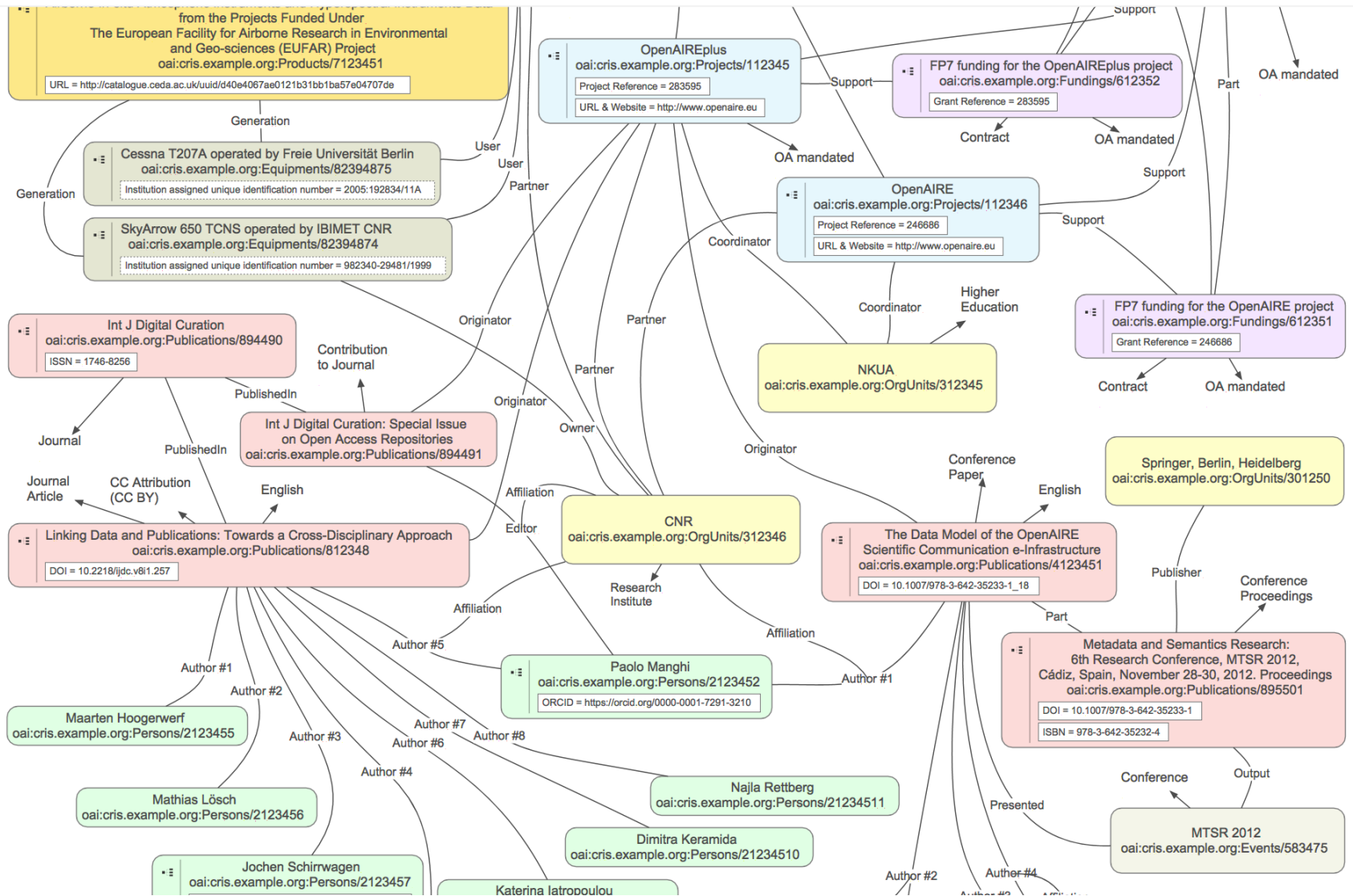
Research results

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

Who needs Research Information?



Research Information is heavily interlinked



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

What characterizes a research project?

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 research domain]

A web page (URI)

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

The screenshot shows the project details for SEQSEL. Annotations with arrows point from descriptive text to specific parts of the page:

- An acronym** points to the **SEQSEL** logo.
- A code (identifier), for ex a Grant number** points to the **Project reference: 317697**.
- A name or title** points to the **Global cooperation to develop next generation whole genome SEQUENCE SElection tools for novel traits**.
- A (planned) start date** points to **From 2012-05-01 to 2016-04-30, ongoing project**.
- A (planned) end date or duration** also points to the same date range.
- [A source of funding]** points to the **Funded under: FP7-PEOPLE**.
- A short or long description (abstract)** points to the **Objective** section.
- [A few scientific publications]** points to the **Open Access** section, specifically to the **OpenAIRE** icon and the list of publications.
- [A project coordinator]** points to the **Coordinator** section, specifically to **TEAGASC - AGRICULTURE AND FOOD DEVELOPMENT AUTHORITY**.
- [A research domain]** points to the **Subjects** section, specifically to **Life Sciences**.
- A web page (URI)** points to the source URL at the bottom of the slide.

The project details section includes:

- Total cost:** EUR 126 000
- EU contribution:** EUR 126 000
- Coordinated in:** Ireland
- Topic(s):** FP7-PEOPLE-2012-IRSES - Marie Curie Action "International Research Staff Exchange Scheme"
- Call for proposal:** FP7-PEOPLE-2012-IRSES
- Funding scheme:** MC-IRSES - International research staff exchange scheme (IRSES)

The **Objective** section states: "The objective of this project, SEQSEL, is to build upon, solidify, formalise and condense the recently initiated, but as yet informal, relationships between several world-leading research partners in the EU (Ireland, The UK, and the Netherlands) and Oceania (i.e., Australia and New Zealand) in order to conduct joint research and participate in knowledge exchange and to identify regions of the bovine genome associated with feed intake and efficiency. The study will be undertaken in dairy cattle..."

The **Open Access** section lists a publication: "Imputation of genotypes from low density (50,000 markers) to high density (700,000 markers) of cows from research herds in Europe, North America, and Australasia using 2 reference populations" by Pryce, J. E.; Johnston, J.; Hayes, B. J.; Sahana, G.; Weigel, K. A.; Mcparland, S.; Spurlock, D.; Krattenmacher, N.; Spelman, R. J.; Wall, E.; Calus, M. P. L. (2014).

The **Coordinator** section lists: TEAGASC - AGRICULTURE AND FOOD DEVELOPMENT AUTHORITY, Ireland.

The **Participants** section lists: THE SCOTTISH AGRICULTURAL COLLEGE, United Kingdom; and STICHTING DIENST LANDBOUWKUNDIG ONDERZOEK, Netherlands.

The **Subjects** section lists: Life Sciences.

Home About this system Case Studies Contact us

RESEARCH COUNCILS UK Gateway to Research Innovate UK Technology Strategy Board

Social Media Analysis for Social Geography

Lead Research Organisation: [Newcastle University](#)
Department Name: Computing Sciences

[Go back](#)

Overview Organisations People Outcomes

Abstract

Many believe that social media data has the potential to give real insights into communities. However, there is a real challenge in unlocking this knowledge due to the enormous size and speed with which the data is generated, and a lack of understanding of issues such as ethics and the validity of conclusions drawn from the skewed sample of the population that social media users represent.

"Tweet My Street" is tackling these problems with the overall goal of understanding how the knowledge gained can benefit socially excluded communities. It is a multi-disciplinary collaboration involving social geographers, computer scientists and statisticians, which aims to:

- design a scalable system for social media data analysis
- create a tool that makes it easy for social scientists to ask questions and visualise results
- explore the use of modern statistical techniques to derive understanding from the data
- drive and evaluate the system through case studies, including the use of the new tools to augment existing social science methods, and to understand the way third sector organisations can make effective use of social media
- explore how the use of the tools can help the private and public sector to make key decisions.

Funded Value: £449,388
Funded Period: Feb 14 - Feb 16
Funder: ESRC
Project Status: Active
Project Category: Research Grant
Project Reference: ES/M001636/1
Principal Investigator: [Paul Watson](#)
Research Subject: Human Geography
Research Topic: Social Geography Urban Geography

A name or title

A short or long description (abstract)

A (planned) start date

A (planned) end date or duration

A code (identifier), for ex a Grant number

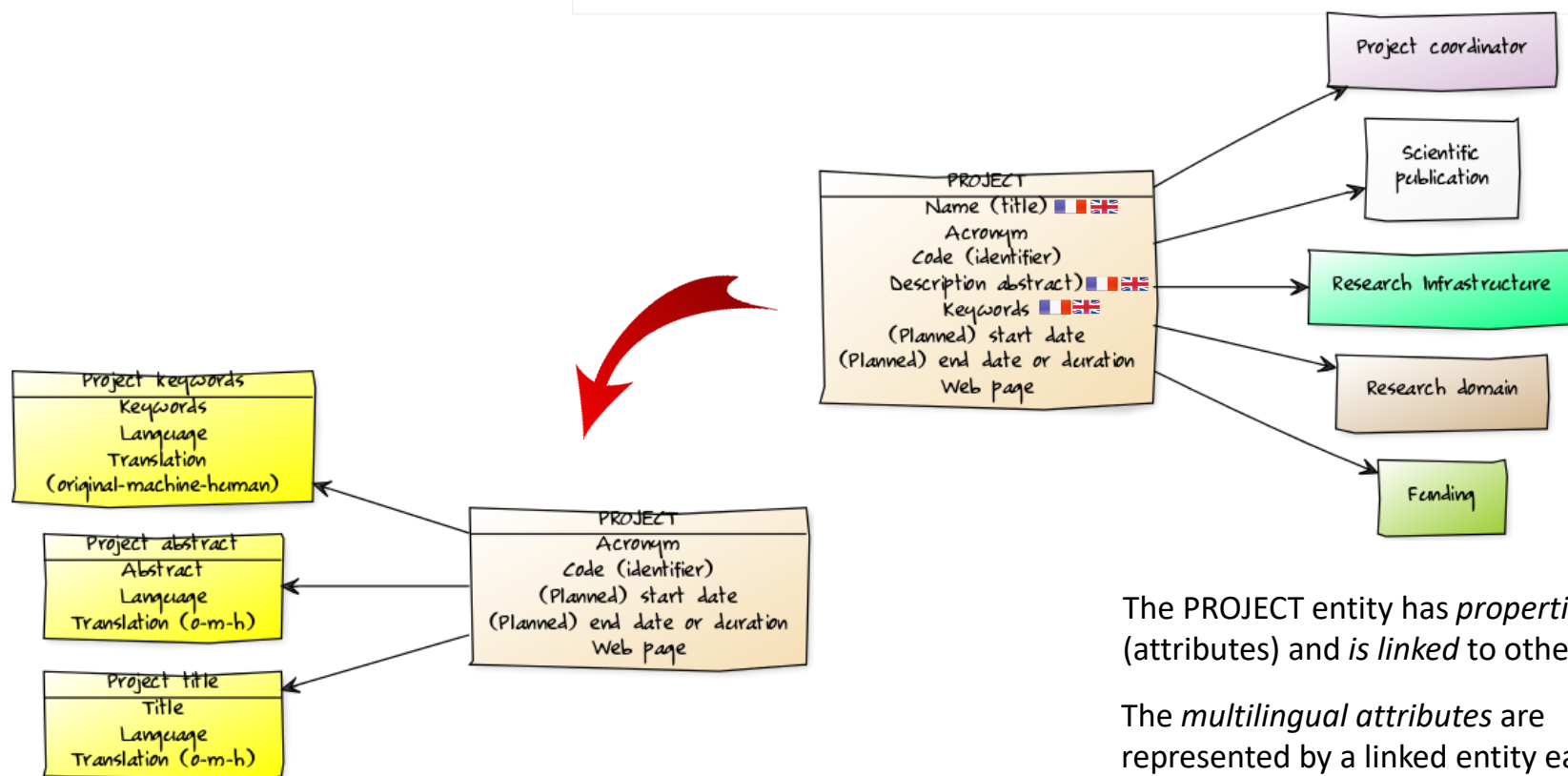
A web page (URI)

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

[A research domain]

A few keywords

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*

cfProjTitle			
cfProjId	ID	NN (PFK)	
cfLangCode	Char(5 BYTE)	NN (PFK)	
cfTrans	NChar(1)	NN (PK)	
cfTitle	Char(255 BYTE)	NN	

cfProjAbstr			
cfProjId	ID	NN (PFK)	
cfLangCode	Char(5 BYTE)	NN (PFK)	
cfTrans	NChar(1)	NN (PK)	
cfAbstr	NClob	NN	

cfProjKeyw			
cfProjId	ID	NN (PFK)	
cfLangCode	Char(5 BYTE)	NN (PFK)	
cfTrans	NChar(1)	NN (PK)	
cfKeyw	Char(255 BYTE)	NN	

Project keywords
Keywords
Language
Translation (original-machine-human)

Project abstract
Abstract
Language
Translation (o-m-h)

Project title
Title
Language
Translation (o-m-h)

PROJECT
Acronym
Code (identifier)
(Planned) start date
(Planned) end date or duration
Web page

cfPrj			
cfProjId	ID	NN (PK)	
cfStartDate	Date		
cfEndDate	Date		
cfAcro	Char(16 BYTE)		
cfURI	Char(128 BYTE)		

Example in DB

cfProj

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

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

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

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.

INTERMEDIATE SUMMARY

- CERIF is:
 - A conceptual model
 - A storage format in relational database
 - A set of exchange formats (XML, Linked Data)
- CERIF supports multilingual data, storing the original value of a literal attribute, and for any other language, a value translated by a machine and/or a human
- So far, we have seen the CERIF Entity "PROJECT" (cfProj)



Common
European
Research
Information
Format

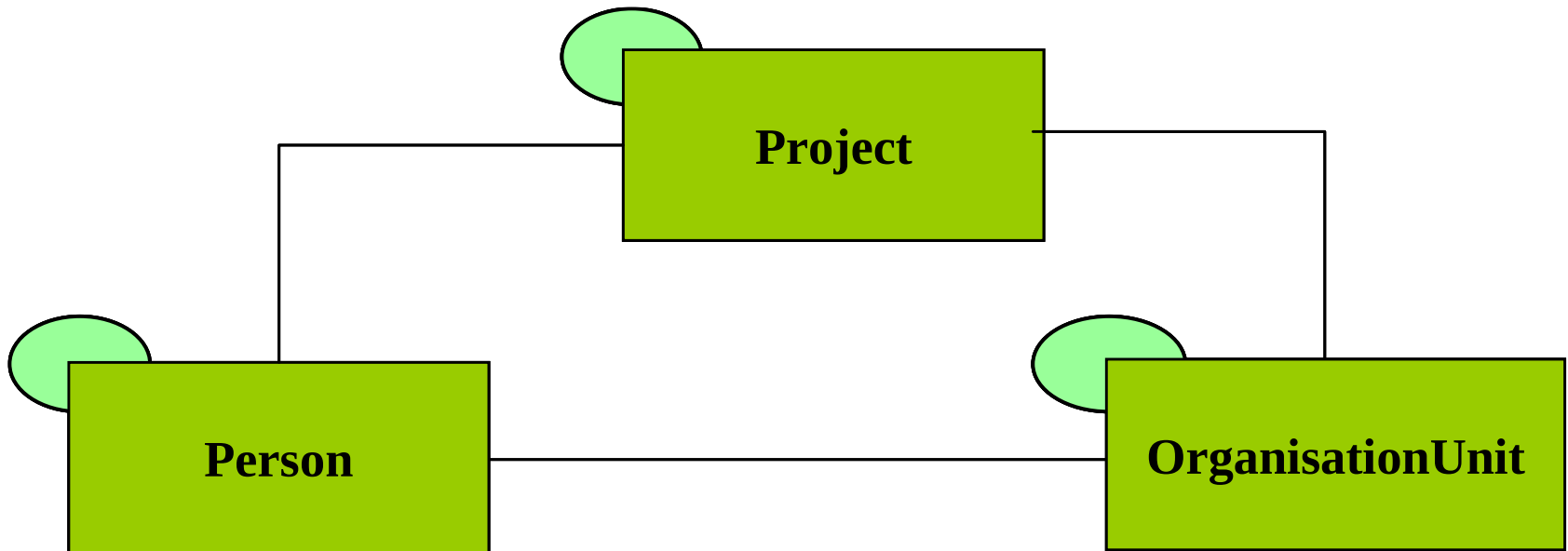
We have seen how to represent, store or exchange metadata about research projects.

Similarly:

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



CERIF Base Entities



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.

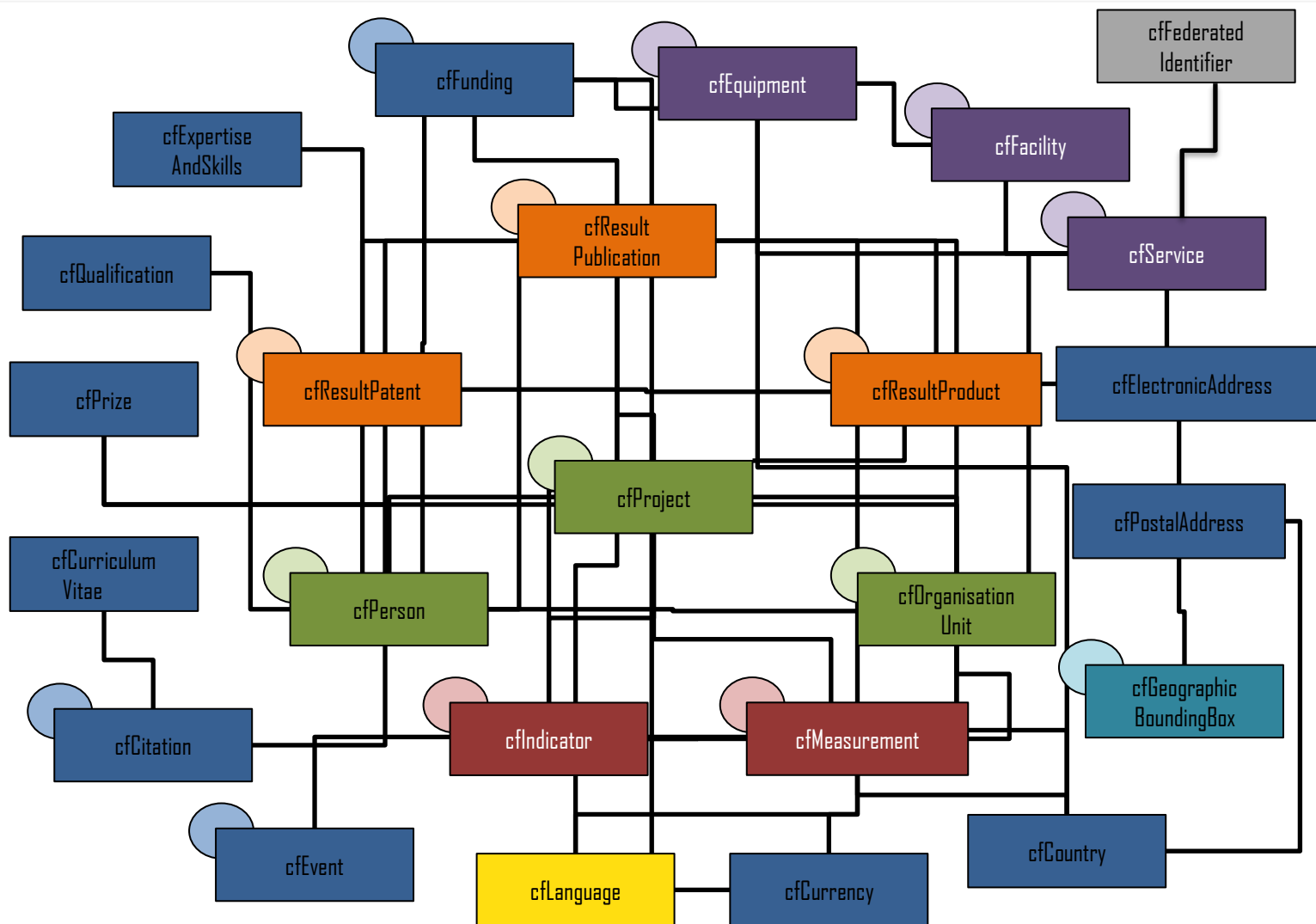
OrgUnit definition

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.

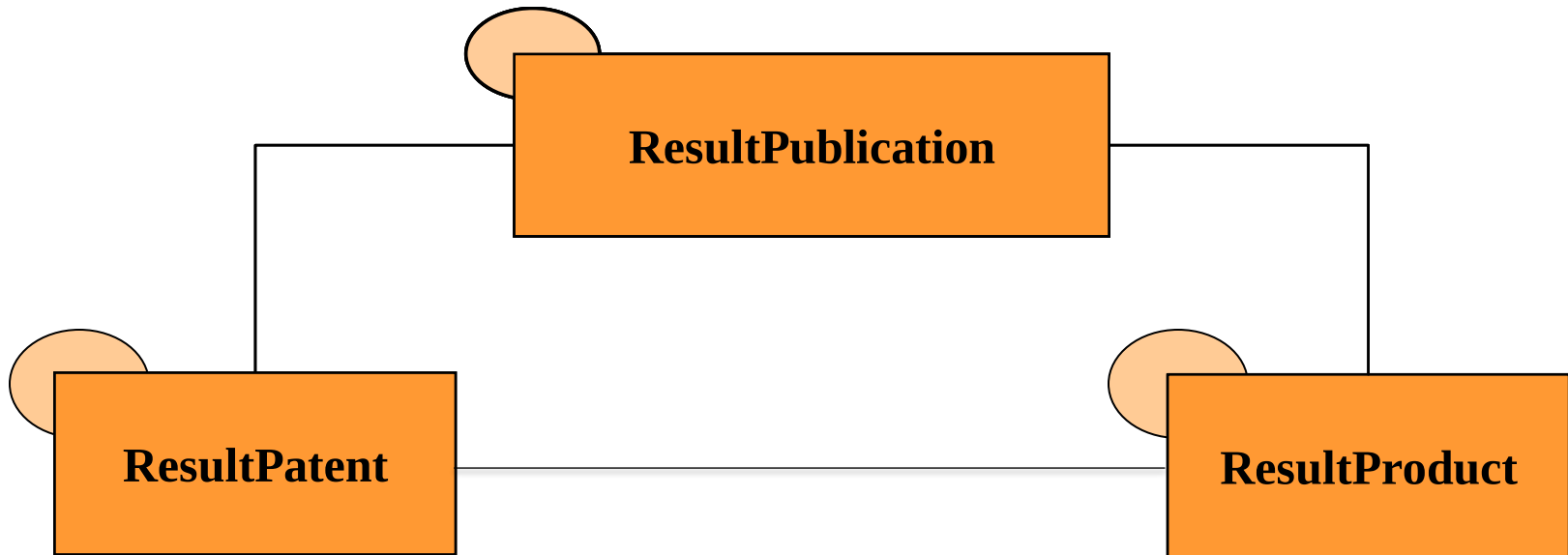
In the research information domain Organisation Units typically represents:

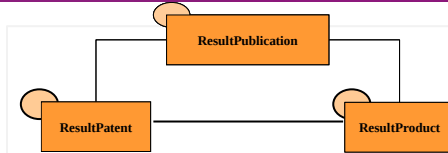
- (1) organisations that **perform research** (universities, research institutes, corporations) and their subdivisions (faculties, schools, departments, research groups) and other associated bodies (boards, advisory bodies);
- (2) organisations that **fund research** (funders, their divisions and evaluation 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.

Common European Research Information Format



CERIF Result Entities





CERIF Result Entities

ResultPublication

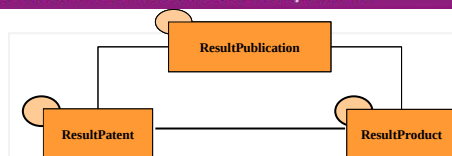
ID
 URI
 Title
 Subtitle
 Abstract
 Bibl. Note
 PublicationDate
 TotalPages
 StartPage
 EndPage
 Keywords
 Type
 Status
 Subject

ResultPatent

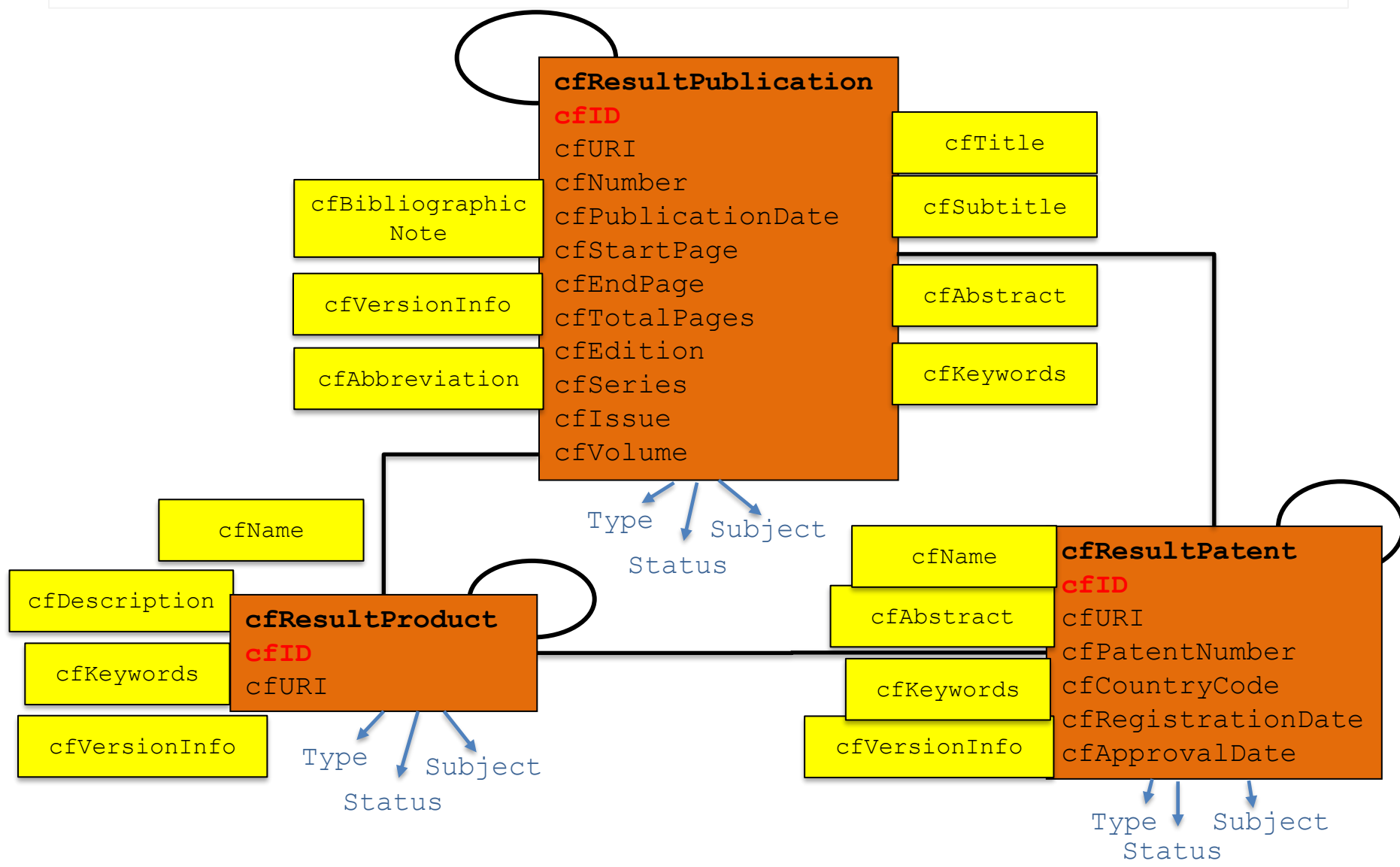
ID
 URI
 PatentNumber
 Title
 CountryCode
 RegistrationDate
 ApprovalDate
 Description
 Keywords
 Type
 Status
 Subject

ResultProduct

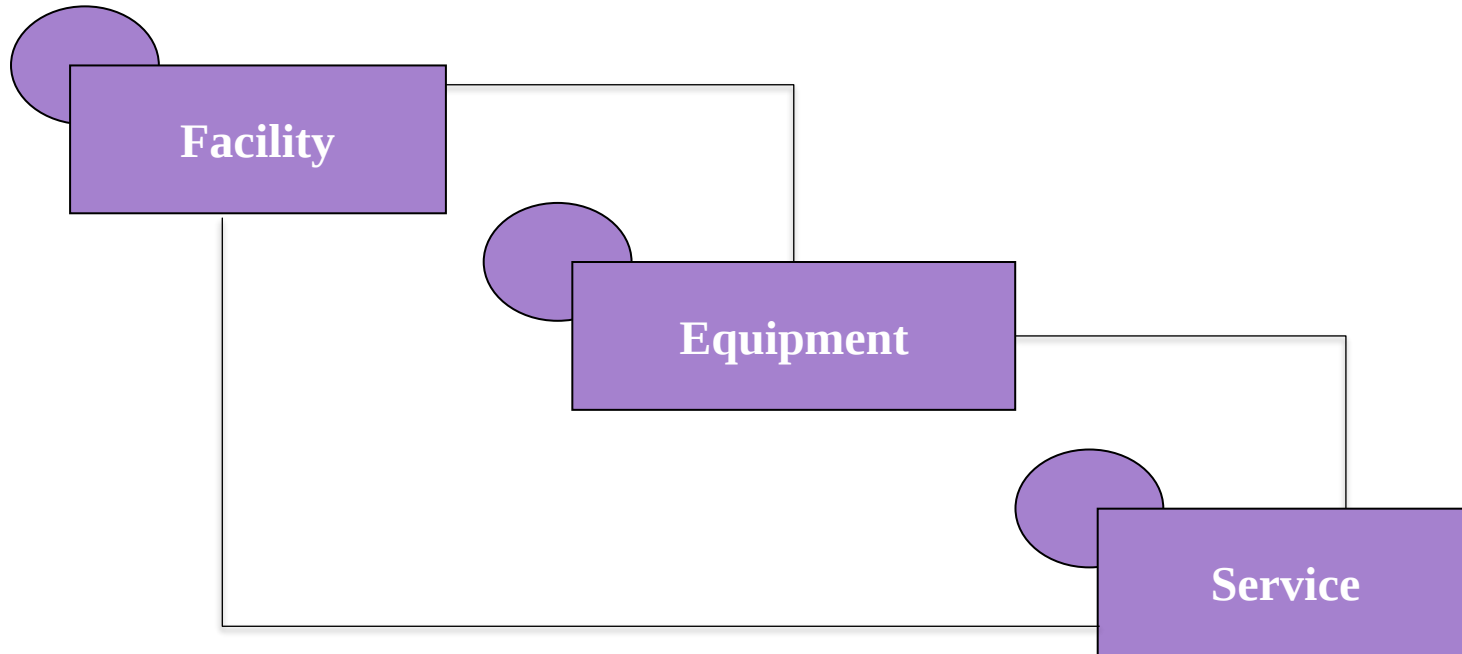
ID
 URI
 Type
 Status
 Subject



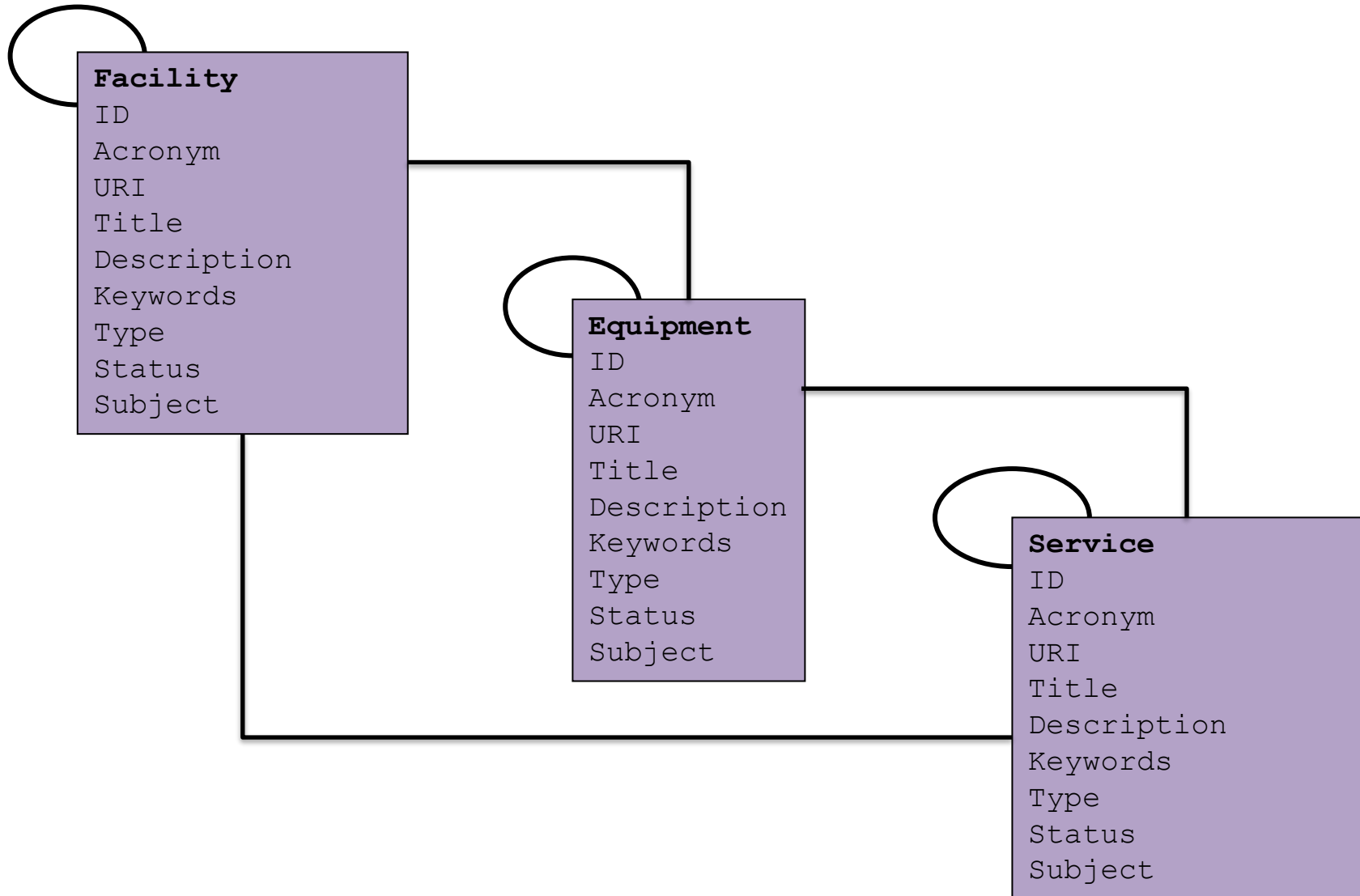
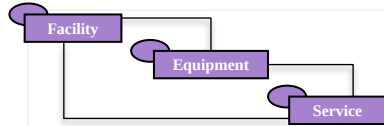
CERIF Result Entities



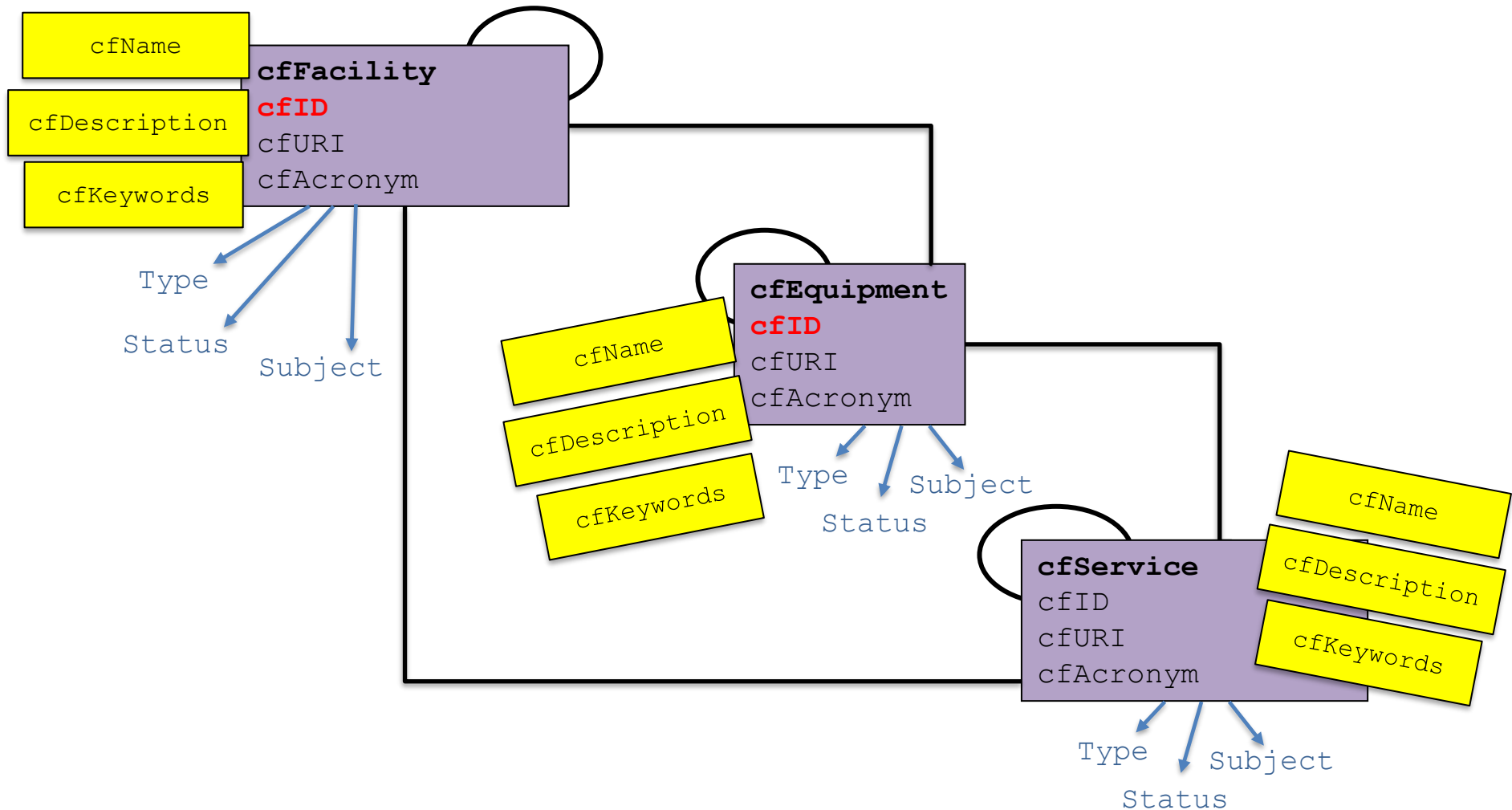
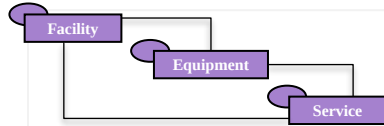
CERIF Infrastructure Entities



CERIF Infrastructure Entities



CERIF Infrastructure Entities

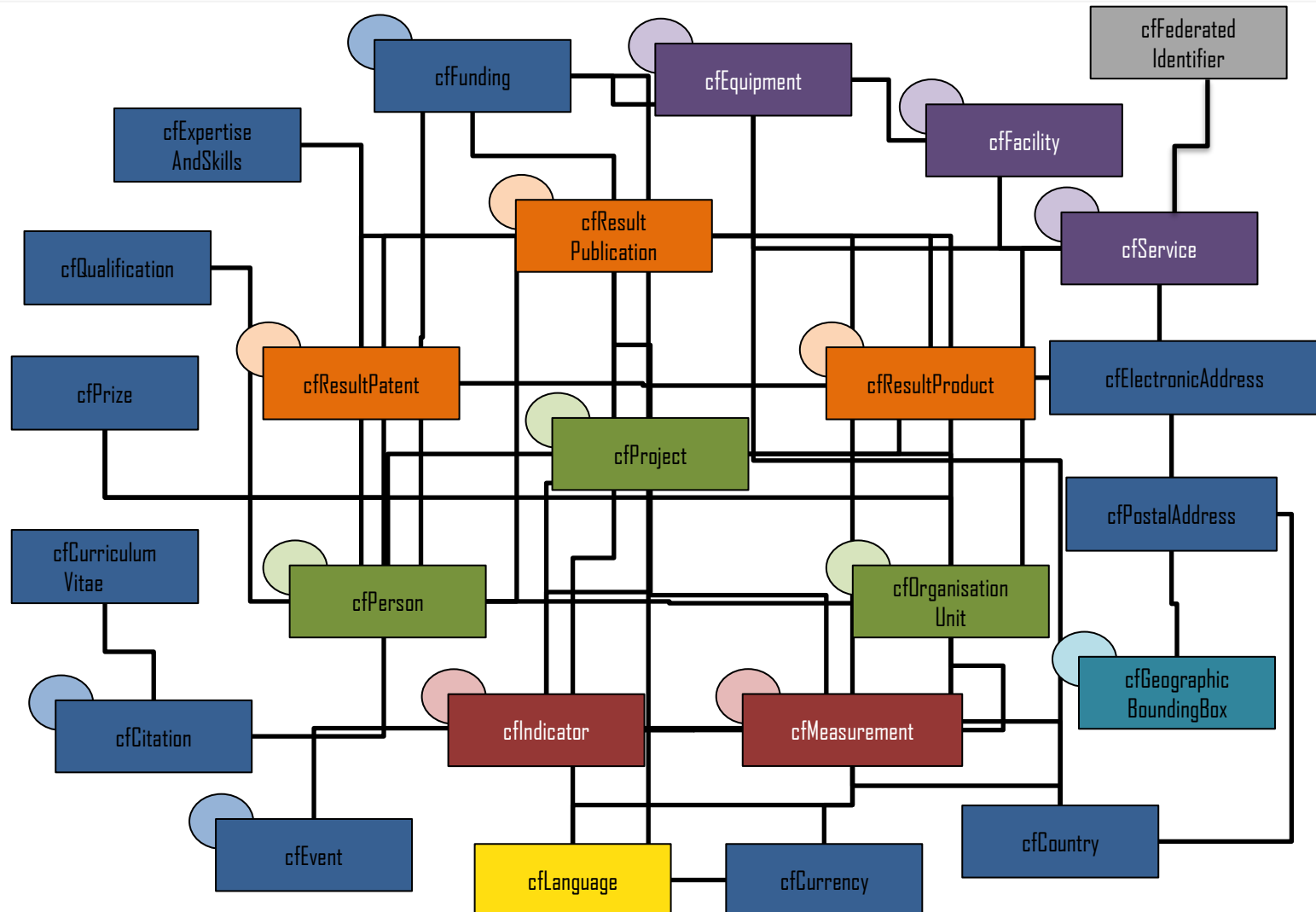


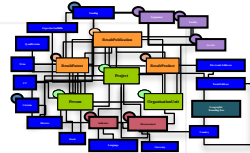
CERIF General Pattern

A typical CERIF entity has:

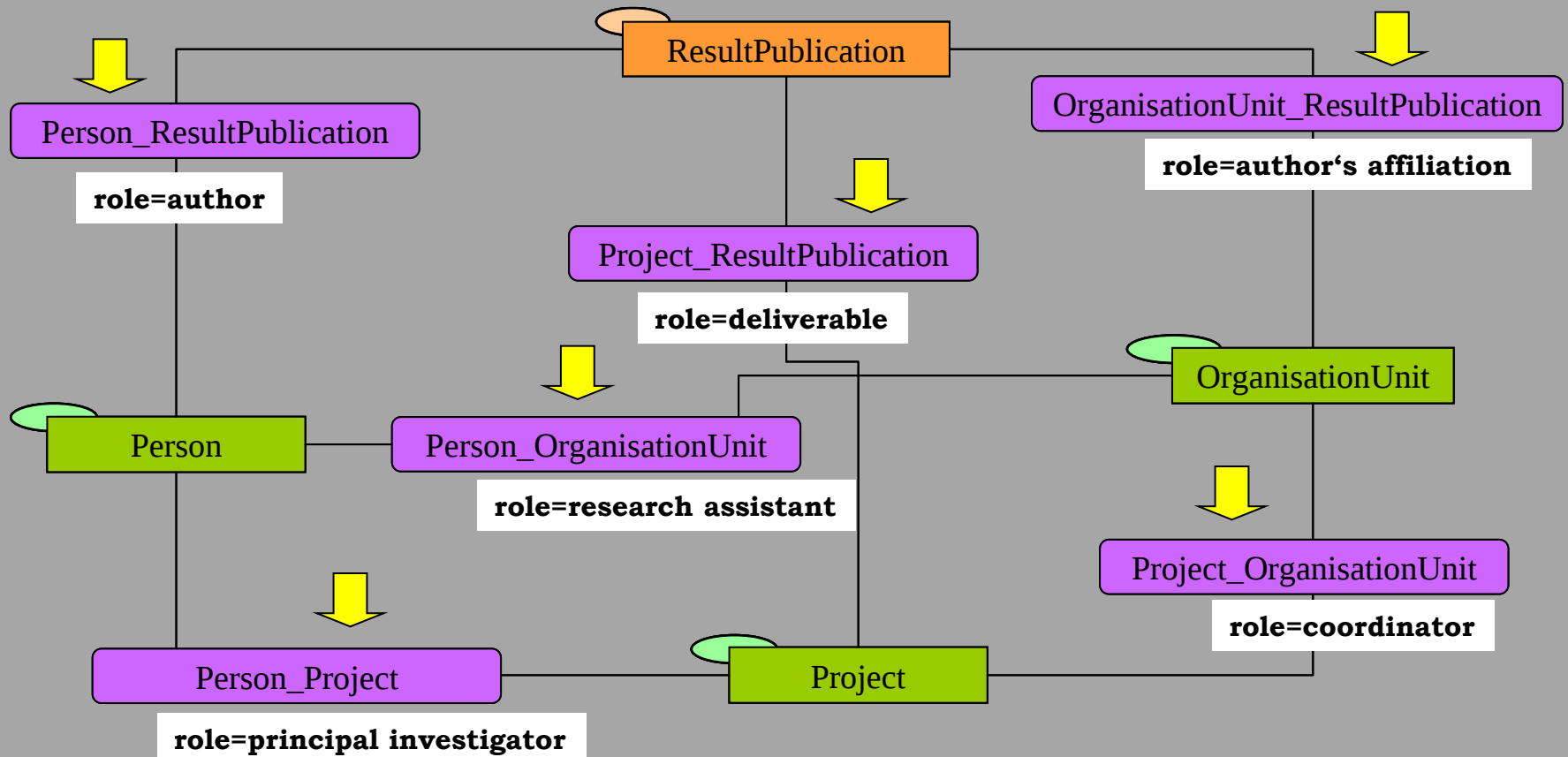
- Identifier (internal)
- Attributes
 - the basic ones
 - the multi-lingual ones
- External Identifiers
- Classifications
 - Type
 - Status
 - Subject area
 - + other
- Links
 - to other entities
 - recursive, a.k.a. self-referential

CERIF 1.6

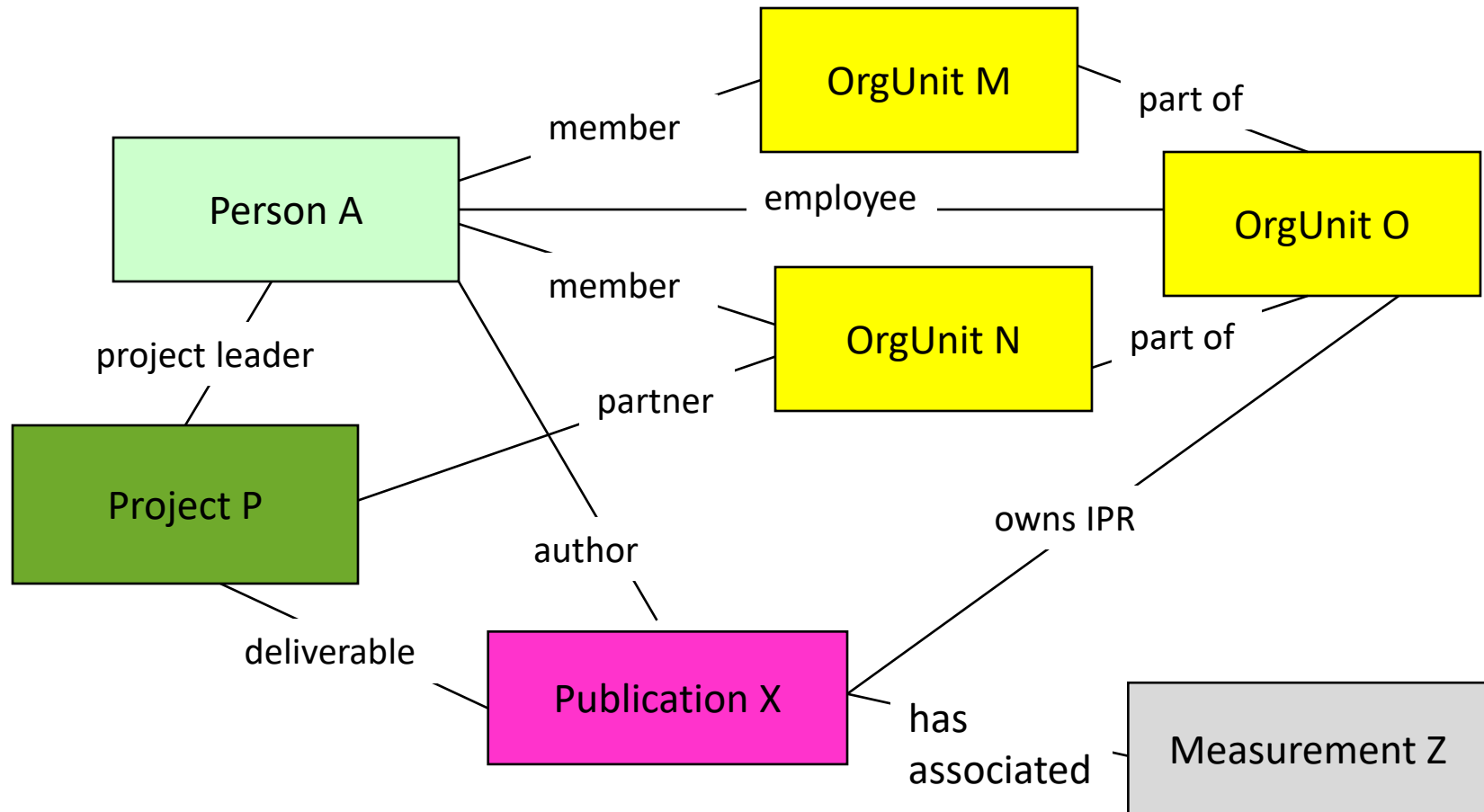




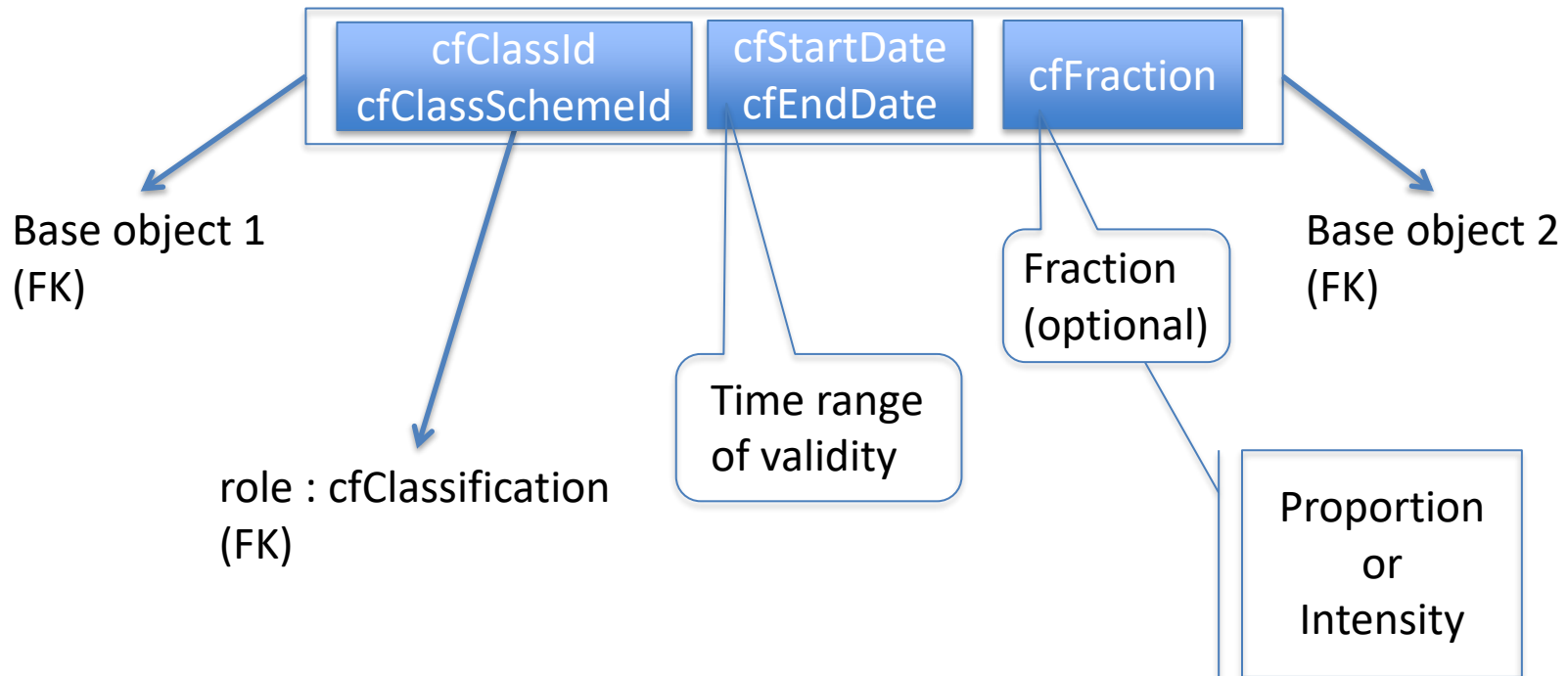
Some CERIF Link Entities



Another example
(slide by Keith Jeffery)



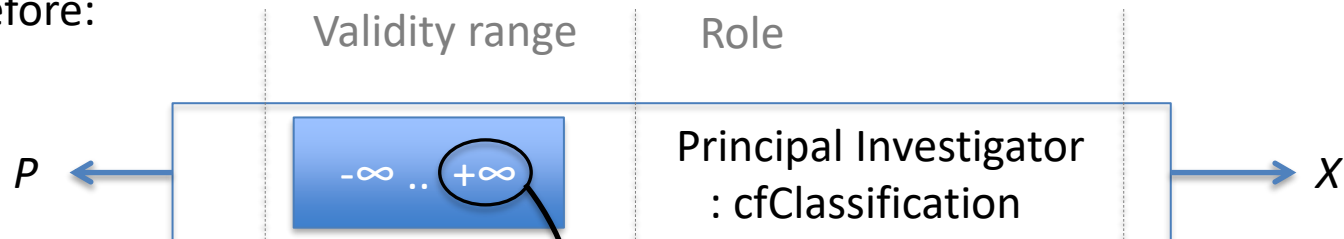
Generic Linking Entity Structure



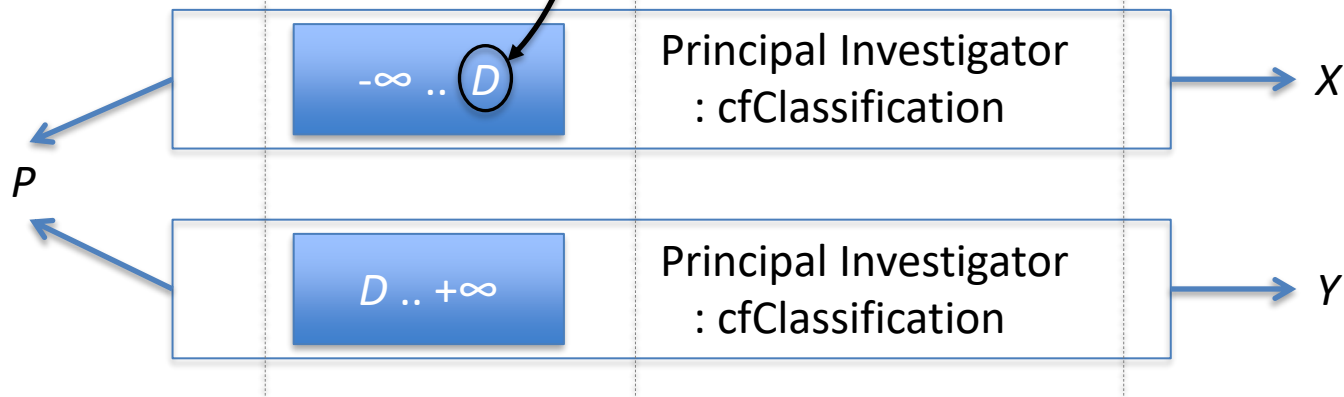
Recording Change in CERIF

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

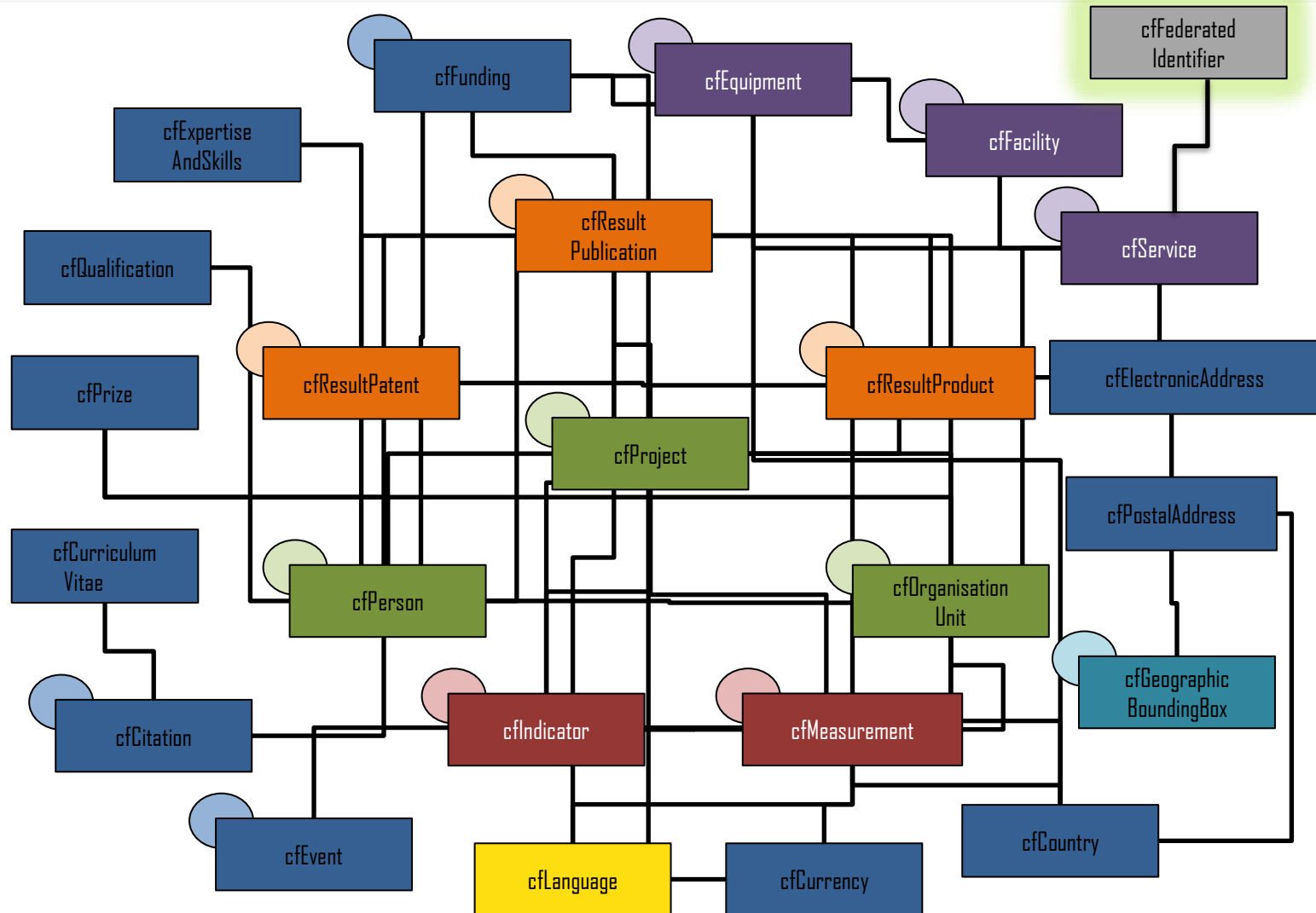
Before:



After:



CERIF 1.6



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, Handle, ARK
- 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
- GrantDOI
- RAiD IDs

Organisation

- ROR ID
- FundRefID
- Wikidata reference
- ISNI
- RingGold number
- VAT Identification Number or another administrative identifier

Classifications: External Codes

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

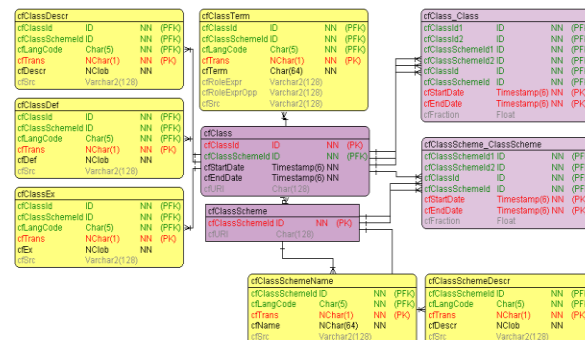
- New Schemas
- New Concepts / Terms
- New Relationships

Enables to manage

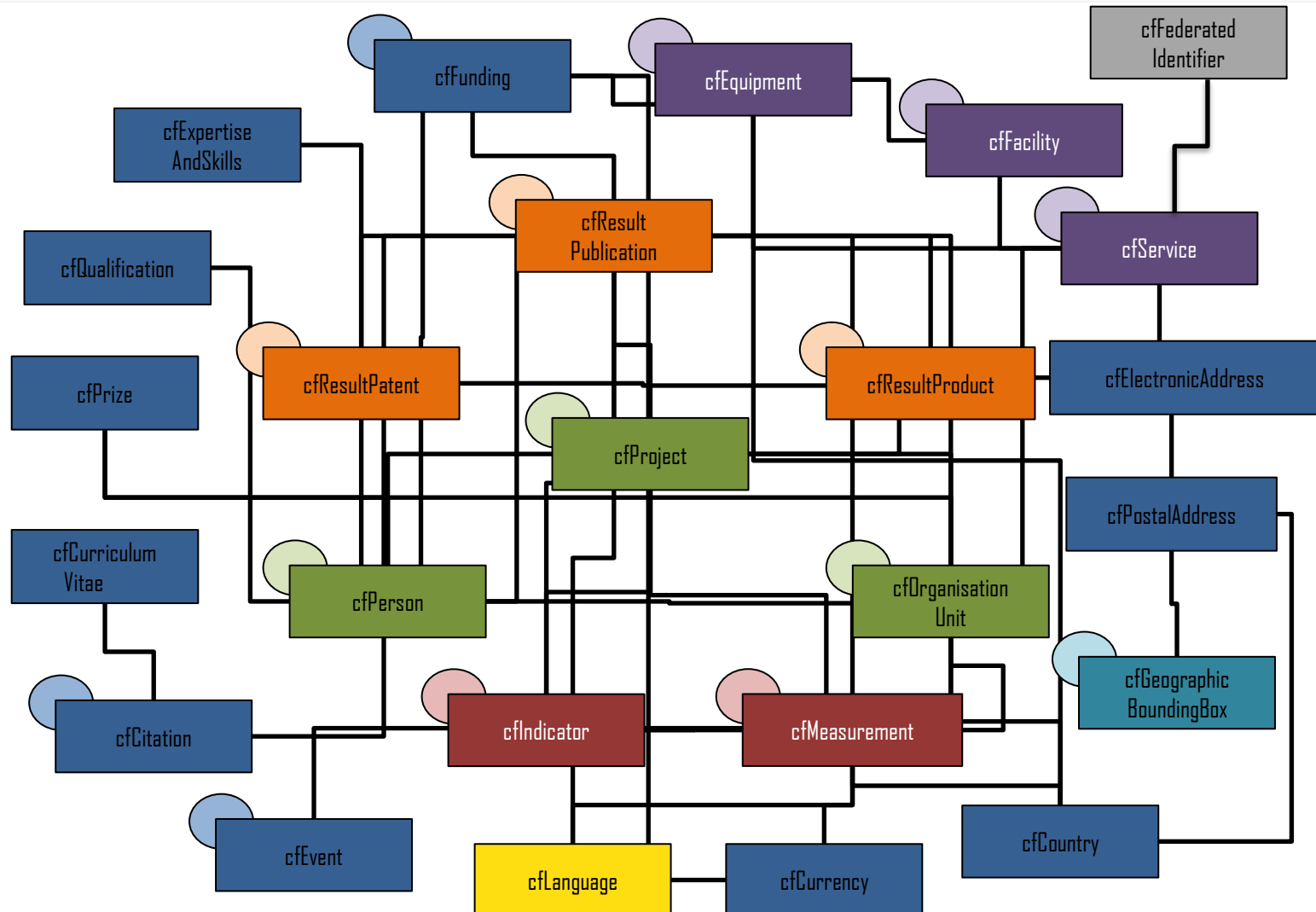
- Roles, Types, Statuses, ... Semantics
- Subject Headings
- Versioning and archiving (start+end dates)

Allows for Mappings between Schemes

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



CERIF 1.6



Metadata Layers

Discovery metadata

DC, VIVO, MODS, METS, eGMS, DCAT, ...

Contextual metadata

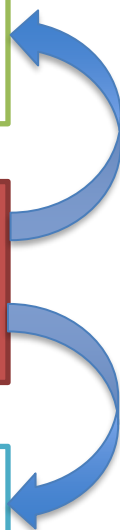
CERIF

Detailed metadata

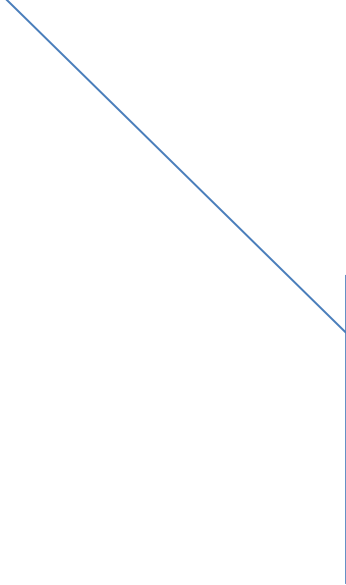
Domain-specific standards

Generate

Reference



Current Research Information System



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

CERIF development

By the CERIF Task Group of euroCRIS

Adopting open-source software projects
tools & best practices:

- <https://github.com/EuroCRIS/CERIF-DataModel>
- <https://github.com/EuroCRIS/CERIF-Vocabularies>
- CC BY license

Two branches:

- master: latest official release (1.6.1)
- develop: on-going development

Model Info

Model Info

ER Diagram

Entities

Attributes

Keys

Relationships

Domains

Basic Information

Project	CERIF: the Common European Research Information Format
Model	CERIF Data Model
Version	CERIF current development DM
Company	euroCRIS
Author	CERIF Task Group
Created	30.6.2014
Last Modified	22.12.2017

Statistic Information

Entities	295			
Attributes	1829			
Domains	1			
User Data Types	0			
Relationships	671	Identifying	638	Non-identifying 33
Indexes	0			
Keys	298	Primary Keys	295	Alternate Keys 3
Views	0			
View Relationships	0			
Procedures	0			
Defaults	0			
Schemas	0			
Users	0			
User Groups	0			
Notes	0			

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

<http://cordis.europa.eu/cerif/>

INTEROPERABILITY

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

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 1.0.

DOI [10.5281/zenodo.17065](https://doi.org/10.5281/zenodo.17065)

Ex.: OpenAIRE Guidelines for CRIS Managers 1.2.

DOI [10.5281/zenodo.8050936](https://doi.org/10.5281/zenodo.8050936)

OpenAIRE Guidelines for CRIS Managers 1.2

- Dvořák, Jan, Czerniak, Andreas, Ivanović, Dragan. (2023). OpenAIRE Guidelines for CRIS Managers 1.2 (Version 1.2.0). Zenodo. DOI [10.5281/zenodo.8050936](https://doi.org/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 DAI-PMH 2.0
- Development in <https://github.com/openaire/guidelines-cris-managers> with Andreas Czerniak and Dragan Ivanović

Version 1.1: Aligned vocabularies

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

OpenAIRE Guidelines for CRIS Managers

latest

Search docs

Introduction

CRIS information elements relevant for OpenAIRE

Publication

- Internal Identifier
- Type
- Language
- Title
- Subtitle
- PublishedIn
- PartOf
- PublicationDate
- Number
- Volume
- Issue
- Edition

Read the Docs v: latest

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 1.1

The Core:

- The CRIS as a Service
- Organisations
- Researchers

Activities:

- Projects
- Funding
- Events

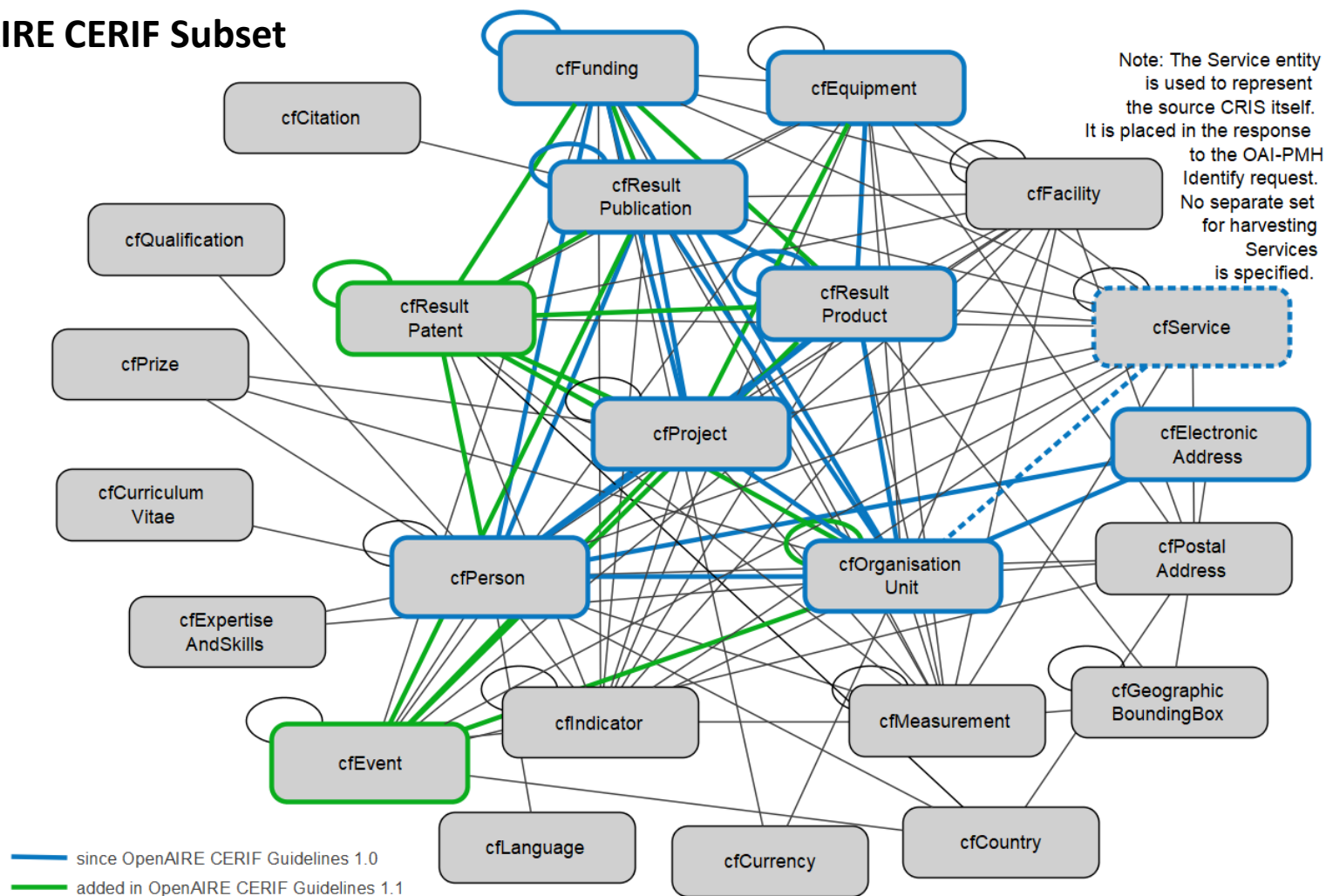
Outputs:

- Publications
- Datasets, Software & other Products
- Patents

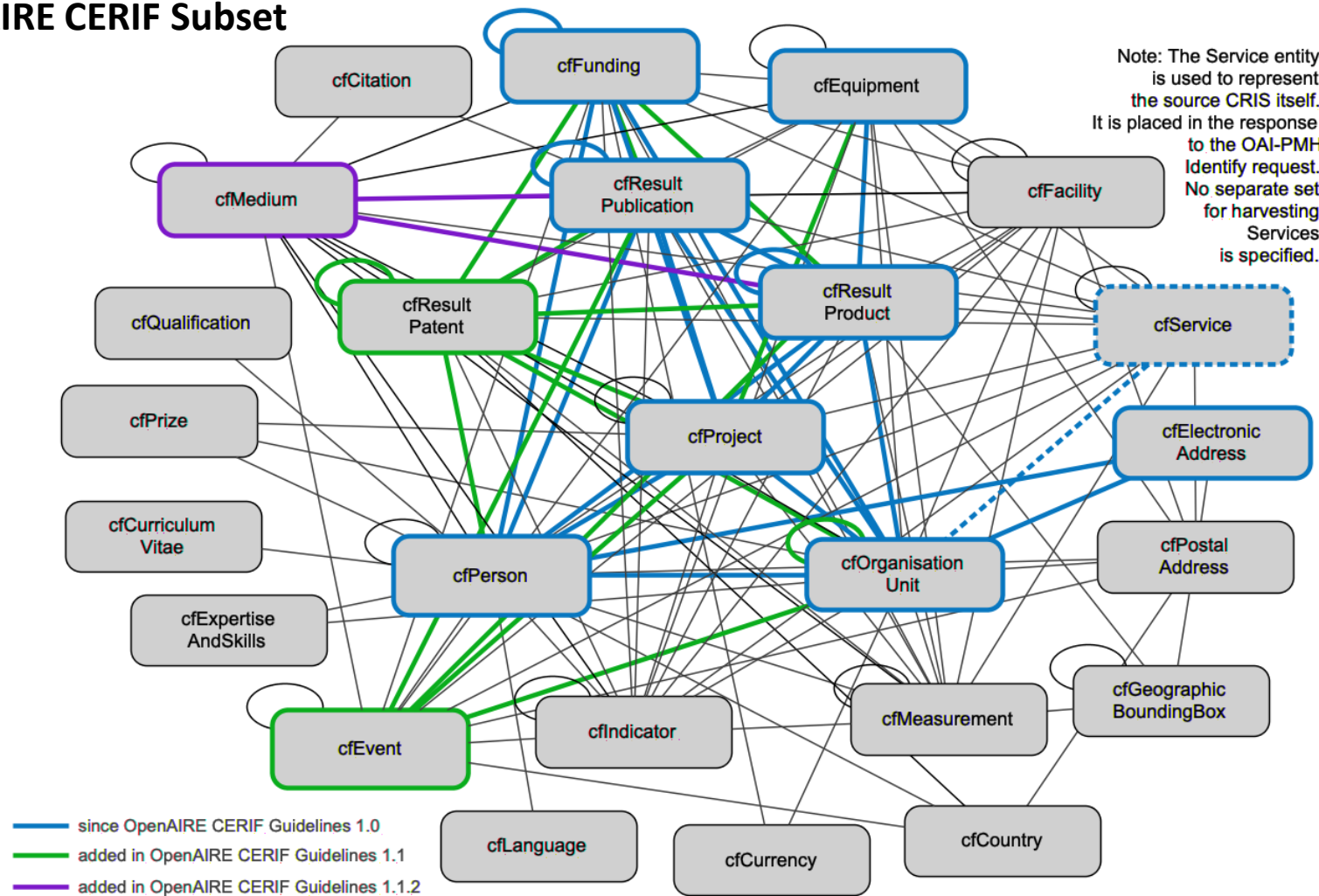
Infrastructure:

- Equipment

OpenAIRE CERIF Subset



OpenAIRE CERIF Subset



Identifiers

- Publication:
 - DOI
 - Handle
 - PMCID
 - ISI Number (WoS Accession Number)
 - SCP Number
 - ISSN
 - ISBN
 - URL
 - URN
 - ZBD-ID
- Persons:
 - ORCID iD
 - ISNI
 - ResearcherID
 - Scopus Author ID
 - DAI
- Organizations:
 - ROR ID
 - GRID ID
 - ISNI
 - FundRef ID
- Products:
 - DOI
 - ARK
 - Handle
 - URL
 - URN
- Funding:
 - Grant DOI

Plus a generic option

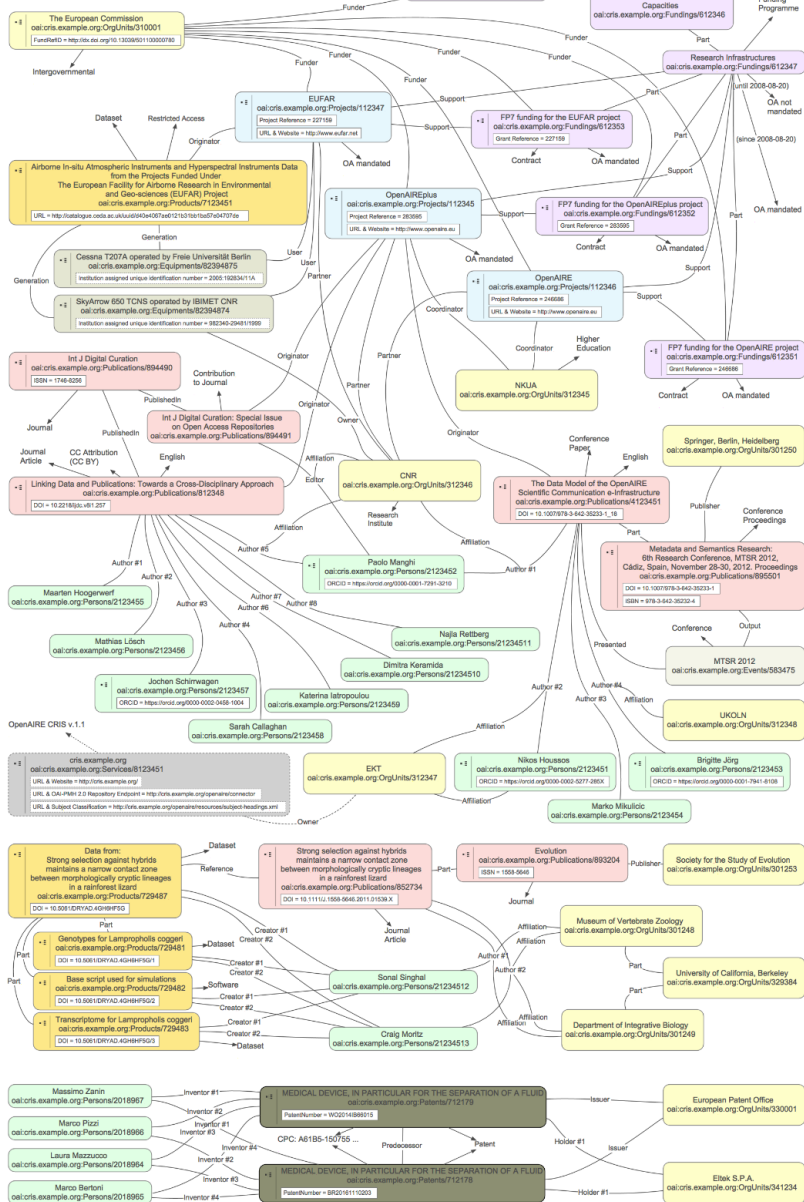
Supporting artifacts

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

The examples for the OpenAIRE Guidelines for CRIS Managers v.1.1: An overview

The examples depict some real-world objects in their (limited) contexts.

Dashed lines and edges denote illustrative claims that are there for the sake of the example.



A Research Graph: The example set from the OpenAIRE Guidelines for CRIS Managers 1.1

The diagram illustrates a complex semantic network of relationships between various entities. The entities are represented by colored boxes, and the relationships are indicated by lines with labels.

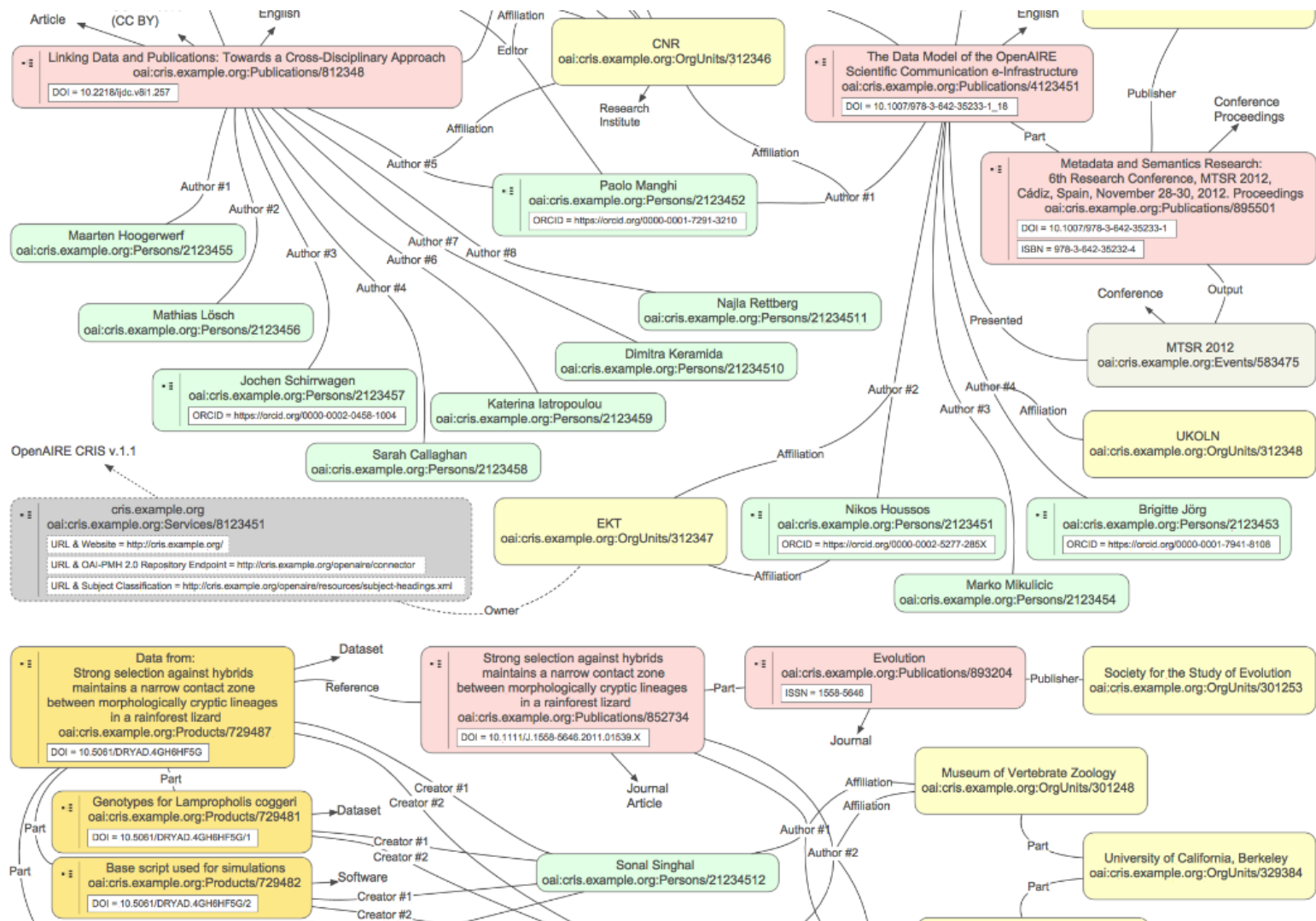
Entities and their details:

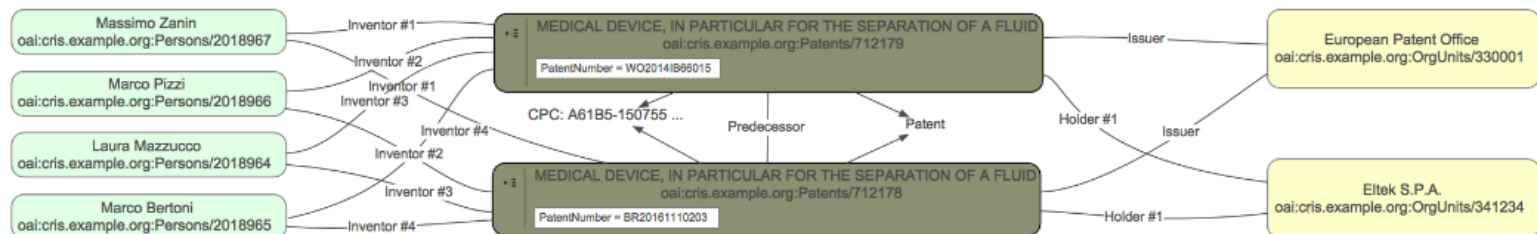
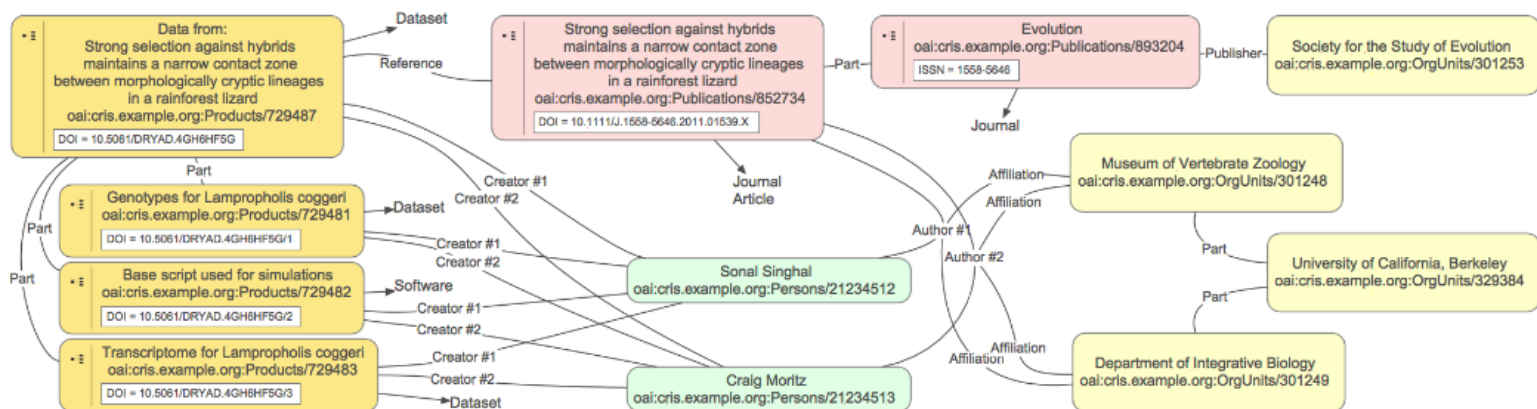
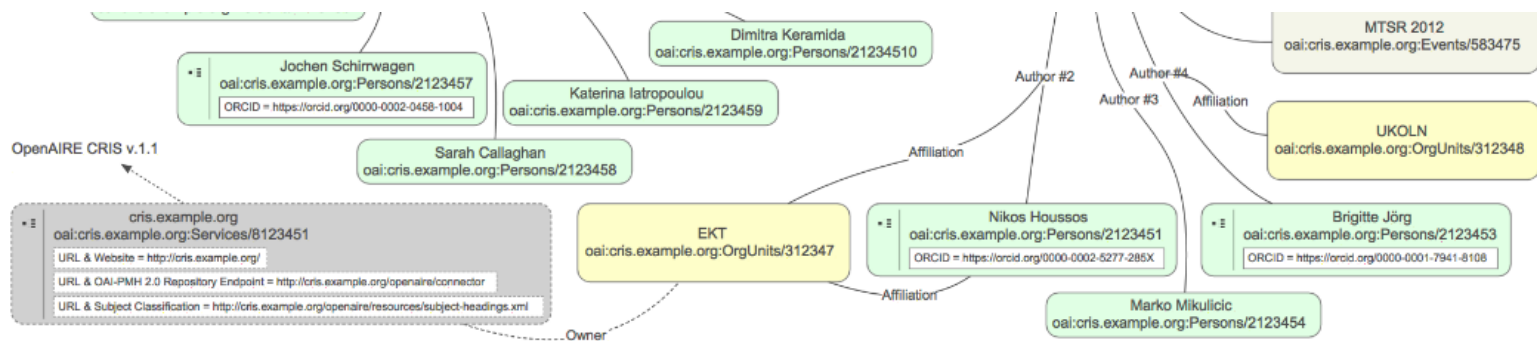
- FP7** (purple box): oai:cris.example.org:Fundings/612345
- Capacities** (purple box): oai:cris.example.org:Fundings/612346
- Research Infrastructures** (purple box): oai:cris.example.org:Fundings/612347
- FP7 funding for the EUFAR project** (purple box): oai:cris.example.org:Fundings/612353, Grant Reference = 227159
- FP7 funding for the OpenAIREplus project** (purple box): oai:cris.example.org:Fundings/612352, Grant Reference = 283595
- OpenAIREplus** (blue box): oai:cris.example.org:Projects/112345, Grant Reference = 283595, URL: http://www.opensai.eu
- OpenAIRE** (blue box): oai:cris.example.org:Projects/112346, Project Reference = 246686, URL & Website = http://www.opensai.eu
- NKUA** (yellow box): oai:cris.example.org:OrgUnits/312345
- Springer, Berlin, Heidelberg** (yellow box): oai:cris.example.org:OrgUnits/301250
- CNR** (yellow box): oai:cris.example.org:OrgUnits/312346
- The Data Model of the OpenAIRE Scientific Communication e-Infrastructure** (pink box): oai:cris.example.org:Publications/4123451

Relationships:

- Funder:** FP7 to Capacities, FP7 to Research Infrastructures, FP7 to EUFAR project, FP7 to OpenAIREplus project, OpenAIREplus to OpenAIRE.
- Part:** Capacities to Research Infrastructures, Research Infrastructures to EUFAR project, Research Infrastructures to OpenAIREplus project, Research Infrastructures to OpenAIRE project.
- Support:** Research Infrastructures to EUFAR project, Research Infrastructures to OpenAIREplus project, Research Infrastructures to OpenAIRE project, OpenAIREplus to OpenAIRE project.
- Contract:** EUFAR project to OpenAIREplus project, OpenAIREplus project to OpenAIRE project, OpenAIREplus project to OpenAIRE project.
- OA mandated:** EUFAR project, OpenAIREplus project, OpenAIRE project, OpenAIRE project.
- Coordinator:** OpenAIREplus to OpenAIRE project, OpenAIRE project to NKUA.
- Originator:** OpenAIREplus to OpenAIRE project, OpenAIRE project to NKUA.
- Conference Paper:** The Data Model of the OpenAIRE Scientific Communication e-Infrastructure.
- English:** The Data Model of the OpenAIRE Scientific Communication e-Infrastructure.
- Publisher:** The Data Model of the OpenAIRE Scientific Communication e-Infrastructure.

Dashed lines and edges denote illustrative claims that are there for the sake of the example.





```
<Publication xmlns="https://www.openaire.eu/cerif-profile/1.2/" id="852734">
  <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="893204">
      <Type scheme="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 medium="http://issn.org/vocabularies/Medium#Online">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/1.2/" 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/1.2/" 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 Profiles

Producers know
what to include



Consumers know
what to expect

CERIF Profiles

Useful subsets of CERIF
for specific research information
interchange scenarios

- Entities & attributes:
 - Profile $\subset$ CERIF
- Semantic vocabularies:
 - Profile – specific choices
 - Sources: CERIF & beyond
- Integrity constraints:
 - Profile $\supset$ CERIF

Profile data is
CERIF

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 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



Product ▾ Solutions ▾ Resources ▾ Open Source ▾ Enterprise ▾ Pricing



EuroCRIS / CERIF-Core

Public



<> Code Issues 16 Pull requests Actions Projects Security Insights

main ▾

5 Branches 0 Tags

Go to file

<> Code ▾



jdvorak001

Only consider relationship links from the Relationships sections



4fa4e91 · 3 days ago



607 Commits



.github/workflows

Also regenerate the OWL representation of the Core

4 months ago



datatypes

Add the various forms of documents

3 days ago



diagrams

adding generated files

3 days ago



entities

Add the Repository entity

3 days ago



examples

link correction

4 years ago



guidelines

Fix missing links between markdown files

2 years ago



serializations/RDF

adding generated files

3 days ago



tools

Only consider relationship links from the Relationships...

3 days ago



README.md

Add the Repository entity

3 days ago



README



CERIF Core

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

CERIF Agent

Def: Anything that has the ability to perform actions.

Generalizes the former cfPerson + cfOrgUnit

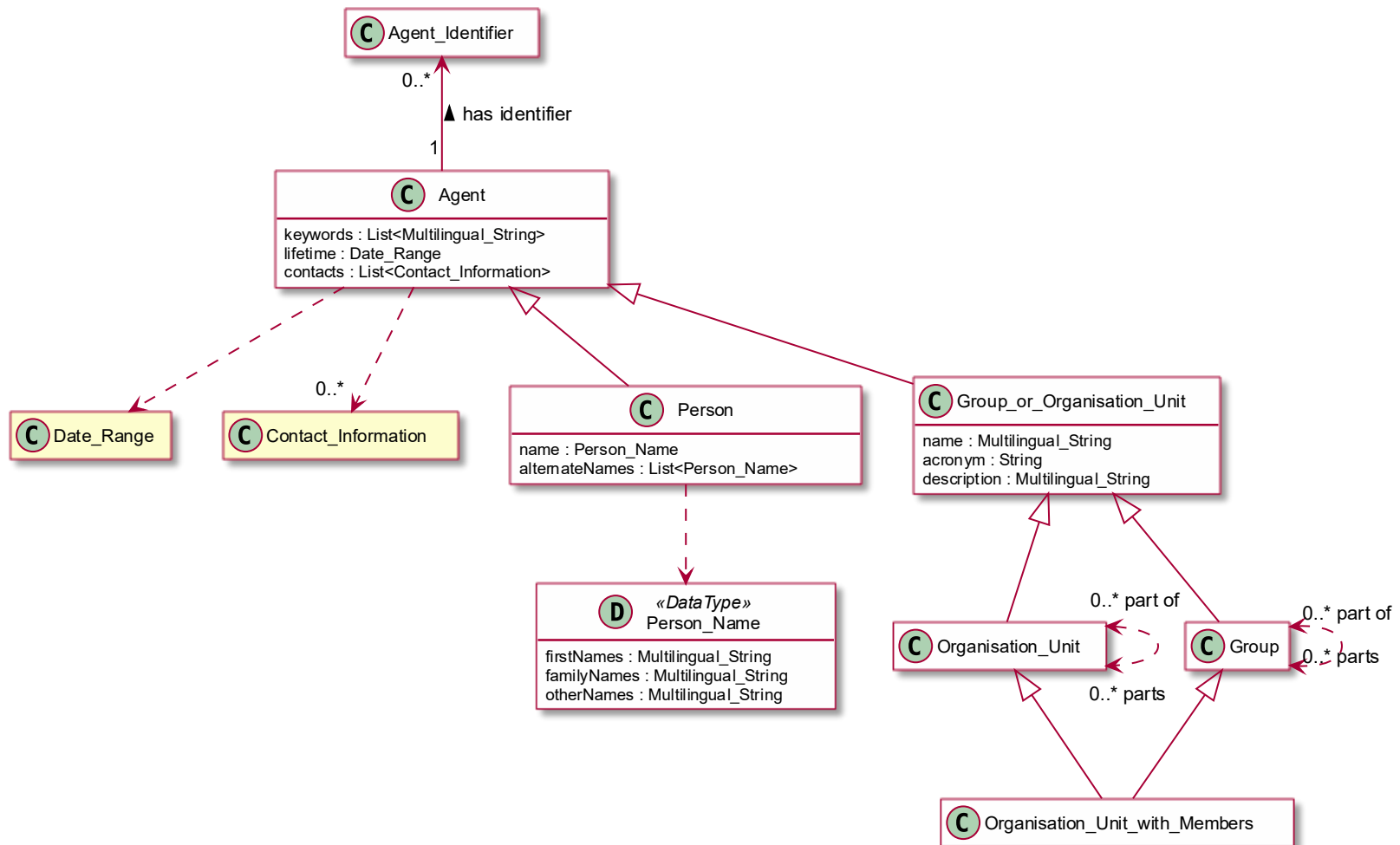
Subclassed as:

- Person
 - Group_or_Organisation_Unit
 - Organisation_Unit (formalized)
 - Group (has members)
- 
- ```
graph LR; A[Organisation_Unit_with_Members] --> B[Organisation_Unit (formalized)]; A --> C[Group (has members)];
```

(following FOAF)

- AI Agent?

# The Agent Hierarchy



## CERIF Activity

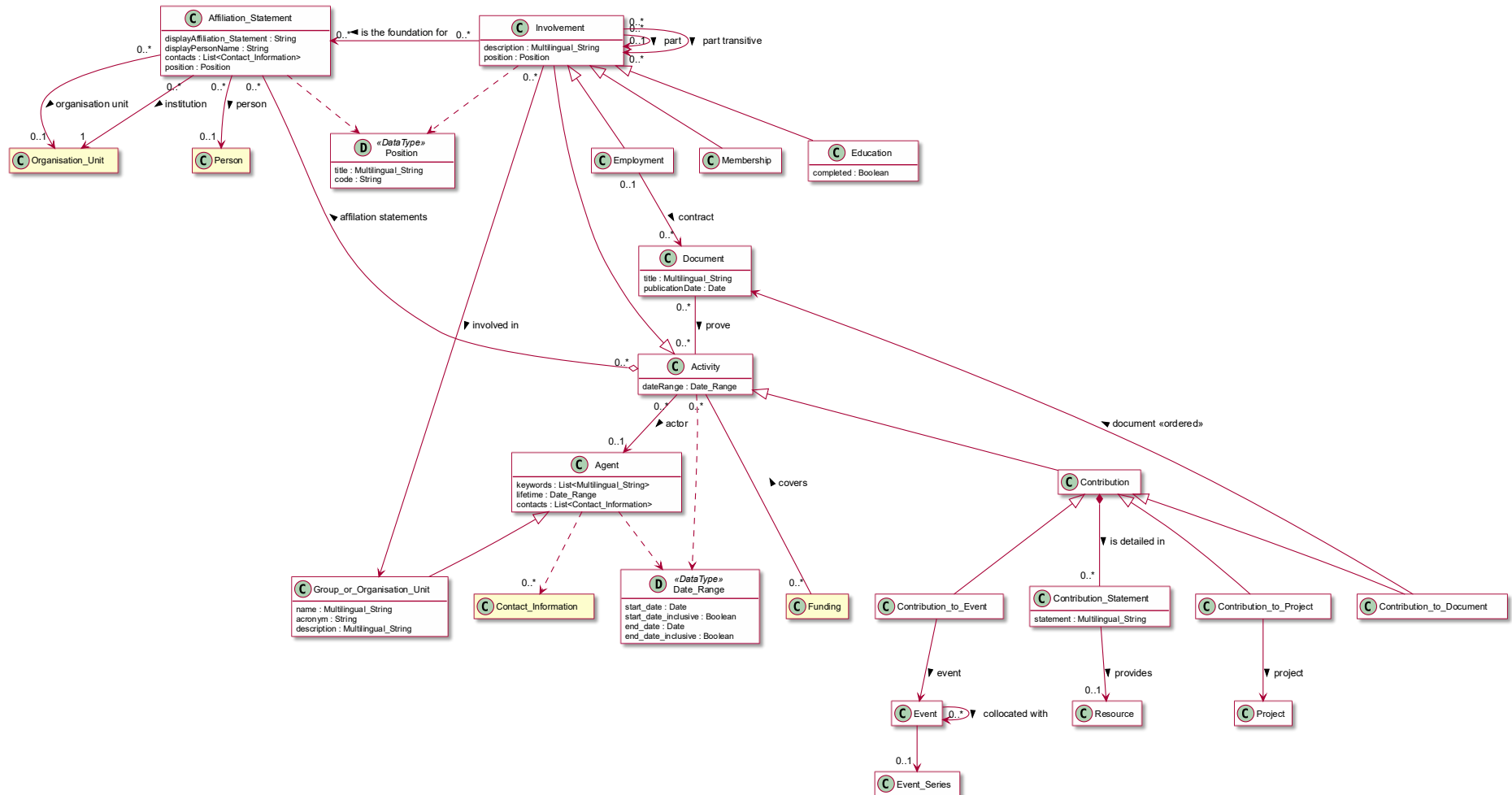
Def: Anything an Agent does.

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



# CERIF AffiliationStatement

To give context to the Agent in his/her/its Activities.

Person

OrgUnit



Person is  
member

Person with  
affiliation to OrgUnit  
is member

OrgUnit is member,  
represented by  
Person

OrgUnit is  
member,  
represented  
by Position

OrgUnit  
is  
member

Where does the  
person come from?  
Why are they there?

Who represents  
the orgunit? Who  
should we talk to?

## CERIF Project & Funding

Project: A temporary endeavor undertaken to create a unique product, service or result.

(Source: the Project Management Institute)

Funding: Money provided by an organization or government for a particular purpose.

(Source: the Oxford Learner's Dictionary of Academic English)

Common confusion: funders, institutions and researchers alike may say “Project” when they mean “Project with Funding” or “Funding”.

# Funding: a specific type of Resource

It's not just Funding that enables research. There is also:

- The expertise, time & effort of Researchers
- Prior knowledge (from Documents)
- Housing the research, the administrative backing by Institutions
- Services offered by Facilities:
  - Laboratories
  - Supercomputing centres
  - Data archives
  - Libraries
  - ...
- Other Services
  - Publishing

Most of these can be had in exchange for some Funding

# Resources exist in different modes

Planned

Sought for

Constructed / Secured

Offered / Invitation to use

Considered

Applied for / Negotiated

Approved / Agreed upon

Contracted

Used

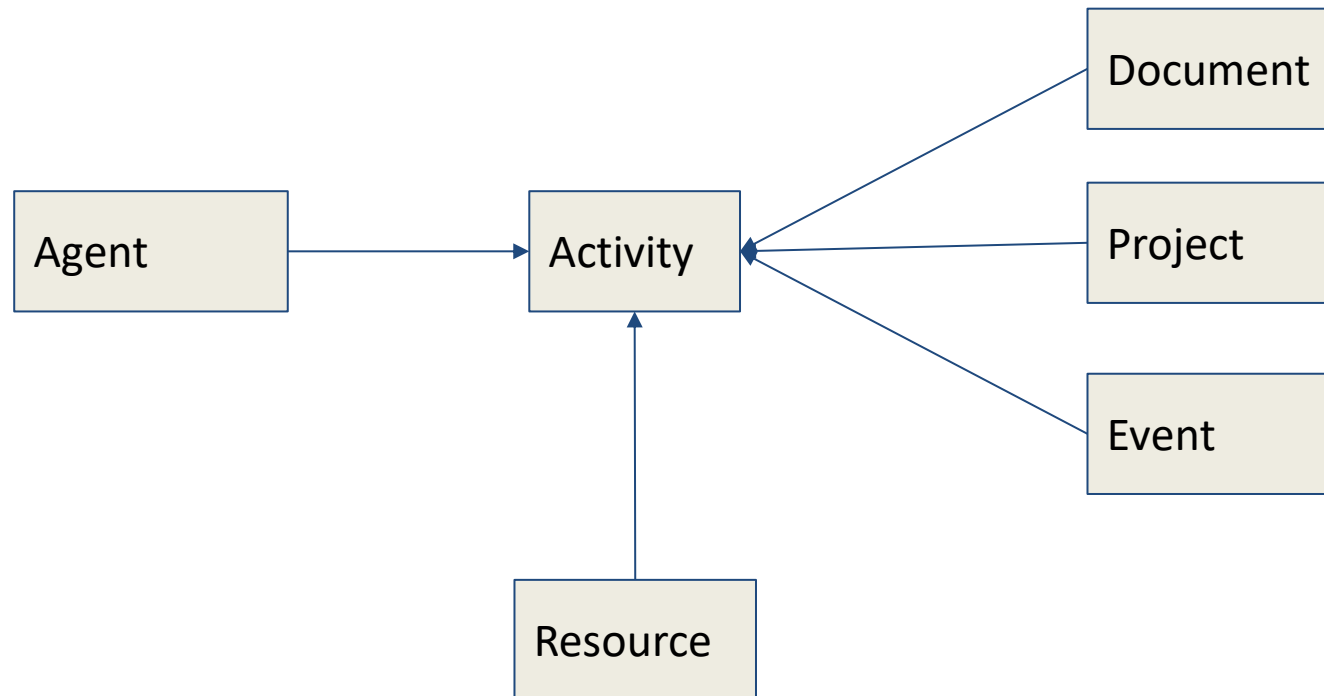
## CERIF Document

Def: A document is a bounded representation of a body of information designed with the capacity (and usually intent) to communicate. A document may manifest symbolic, diagrammatic or sensory-representational information.

(Source: The Bibo Ontology,  
<http://purl.org/ontology/bibo/Document>)

Accommodates current cfResultPublication, cfResultPatent and some of cfResultProduct

# 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
- Research software metadata
- Bibliometric indicators
- Project reporting
- Research Infrastructure description

# **CERIF Scholarly Publication Module**

1st module started along with the Core

Scope:

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



Product ▾ Solutions ▾ Resources ▾ Open Source ▾ Enterprise ▾ Pricing

Search or jump to

EuroCRIS / CERIF-ScholarlyPublicationsModule Public

<> Code 2 Issues Pull requests Actions Projects Security Insights

main 1 Branch 0 Tags

Go to file

<> Code ▾



actions-user adding generated diagrams

33c2726 · 6 months ago 100 Commits

|                    |                                                              |              |
|--------------------|--------------------------------------------------------------|--------------|
| .github/workflows  | Not list the diagrams subdirectory                           | 2 years ago  |
| datatypes          | The initial creation of datatypes ISBN and ISSN, and upda... | 3 years ago  |
| diagrams           | adding generated diagrams                                    | 6 months ago |
| entities           | Add publication series md file (#26)                         | 6 months ago |
| examples           | link correction                                              | 3 years ago  |
| serializations/RDF | Make it consistent with current Core: Contributorship -> ... | 2 years ago  |
| README.md          | Convert to the new template (follow the Core)                | 2 years ago  |

README



## 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.

# Developed in the FAIRCORE4EOSC project

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

Home / EOSC Components /

## Metadata Schema and Crosswalk Registry (MSCR)



**MSCR**

EOSC Metadata Schema  
& Crosswalk Registry

eosc | FAIRCORE4EOSC

Enabling a FAIR EOSC ecosystem

### Expected impact of the Metadata Schema and Crosswalk Registry (MSCR):



Better metadata  
interoperability



Easy to use GUI  
for creating  
crosswalks



Incentivised  
FAIRification



**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!**