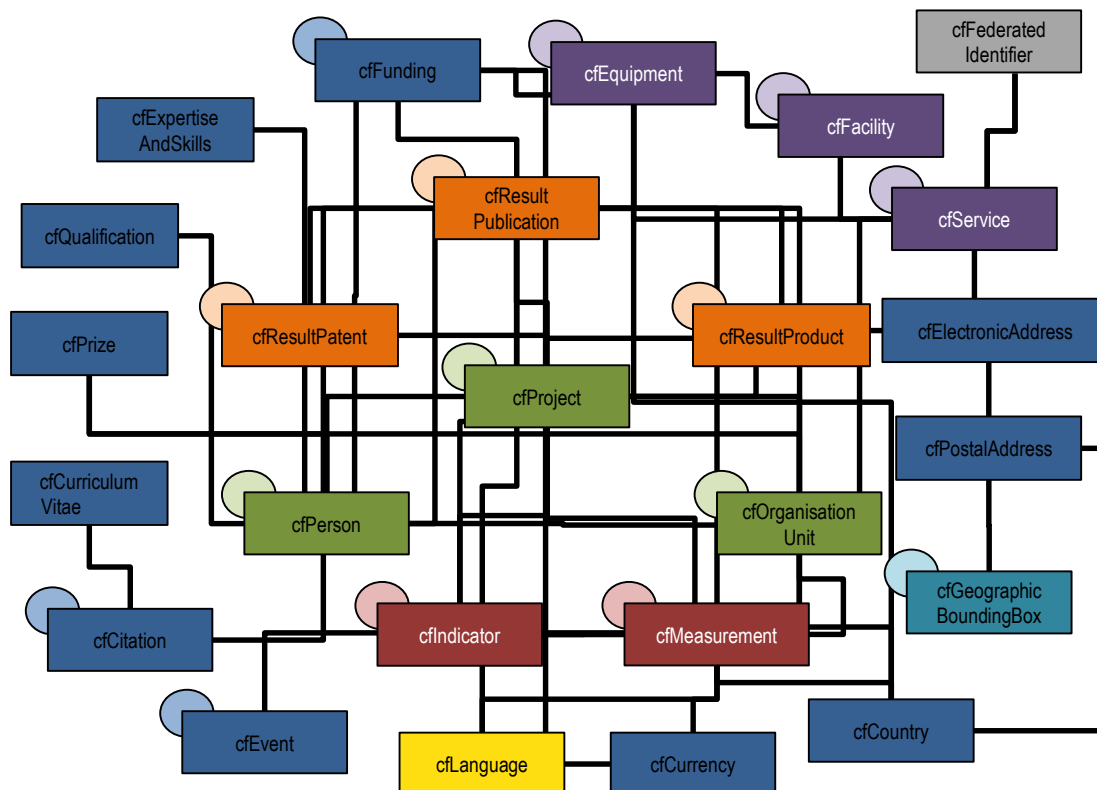


CERIF Tutorial



Jan Dvořák
June 13th, 2018
CRIS 2018 conference
Umeå, Sweden

Jan Dvořák

jan.dvorak@ff.cuni.cz, <https://orcid.org/0000-0001-8985-152X>

euroCRIS

- CERIF TG Leader since 2013
- CRIS 2012 (Prague, June 2012), Organization Committee Chair
- Membership meeting Autumn 2010, Organizer

Charles University in Prague, Faculty of Arts, Institute of Information Studies & Librarianship

- Researcher & Lecturer

Czech Technical University, Computing and Information Centre

- IS Analyst: EZOP+V3S – institutional CRIS

InfoScience Praha

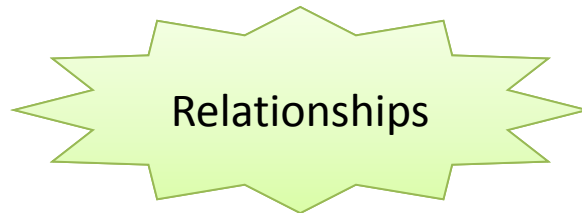
- Research, Development & Innovation Information System
(the national CRIS for [CZ] – www.isvav.cz – 2004-2016)

*This deck of slides is based on the CERIF Tutorial by **Brigitte Jörg**
CERIF TG Leader 2004-2012*

What is Research Information?

The process of research

- Research projects
- Funding
- Research infrastructures



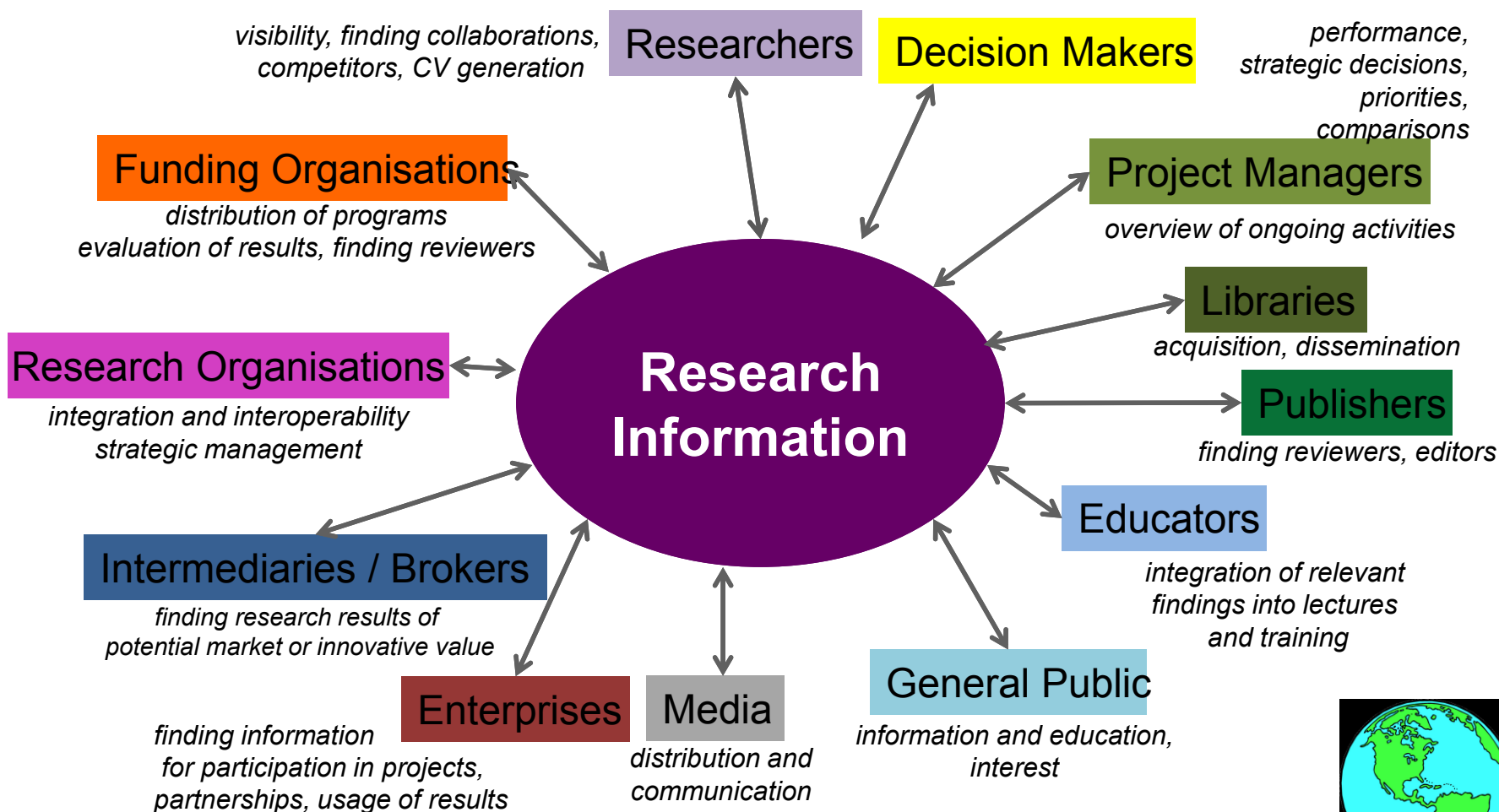
The research actors

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

Research results

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

Who needs Research Information?



Common European Research Information Format

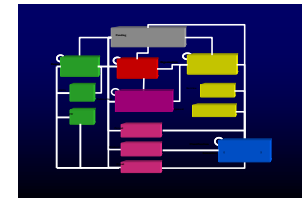
- CERIF is an EU Recommendation to Member States
- The European Commission (EC) has authorised euroCRIS to maintain and develop CERIF and its usage

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

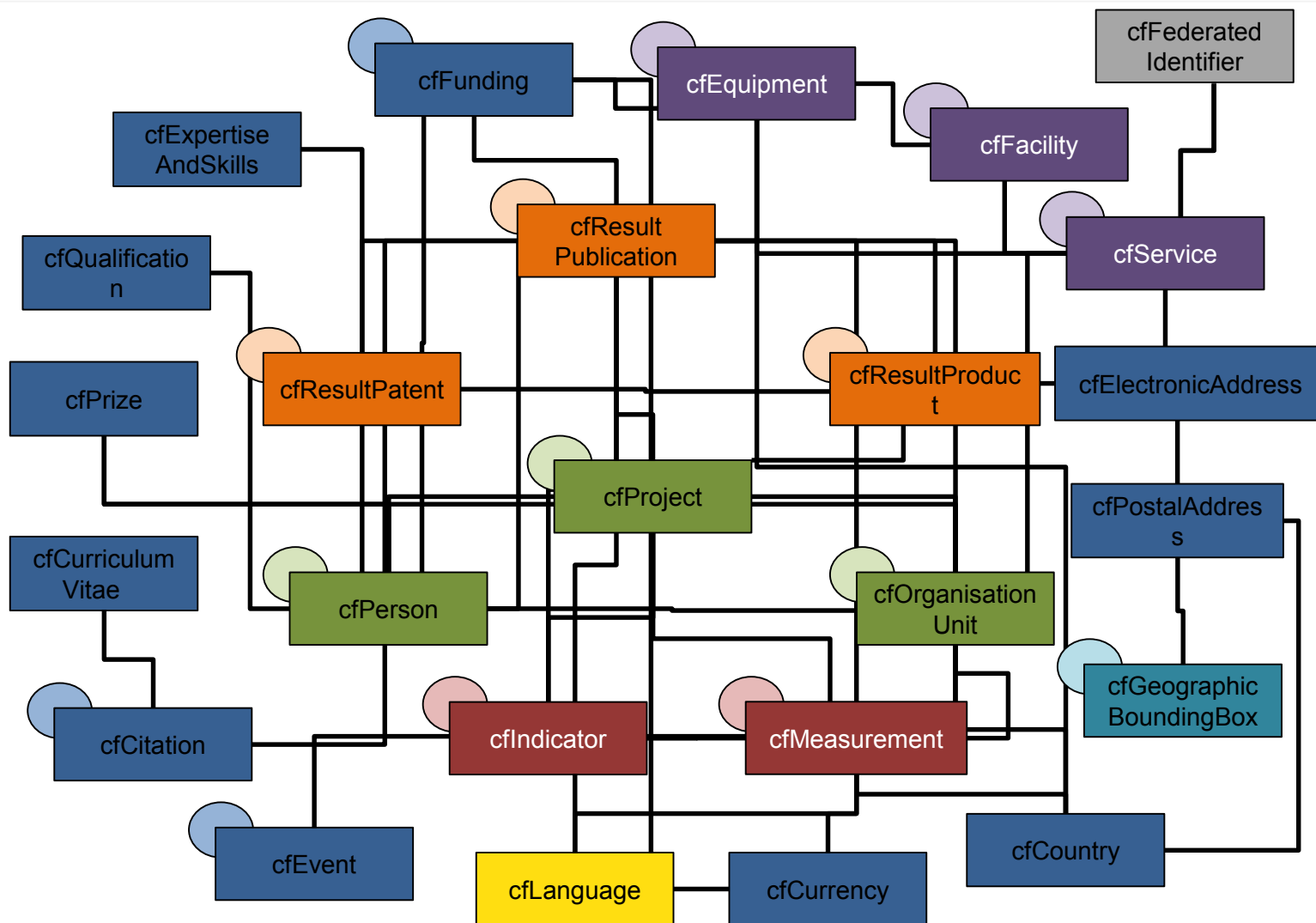
Model Levels



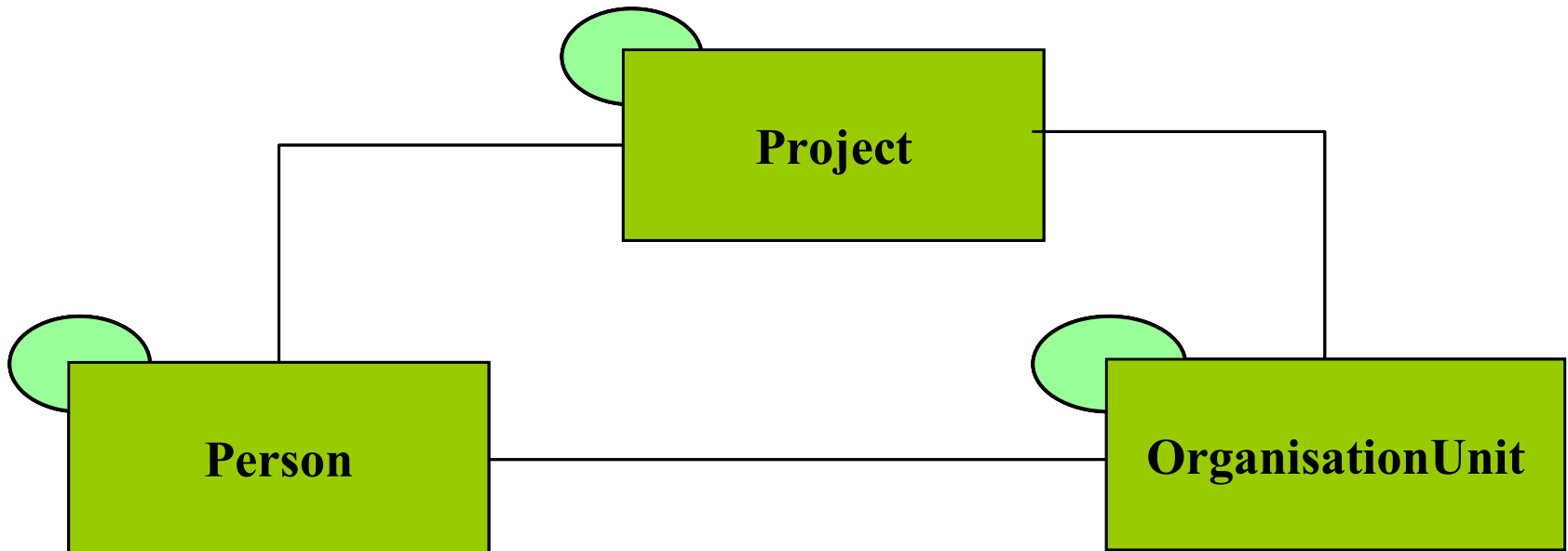
- **Conceptual Level** (Specification)
Concepts relevant for the research domain and their relationships
- **Logical Level** (ER Model)
Entities and their relationships
- **Semantic Layer** (Declared Semantics)
A formalized controlled vocabulary describing a general contextual semantics of the research domain inline with the conceptual, logical and machine description

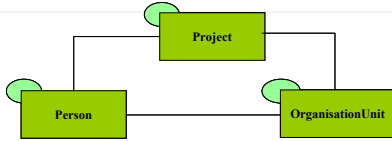


Common European Research Information Format



CERIF Base Entities





CERIF Base Entities

Project

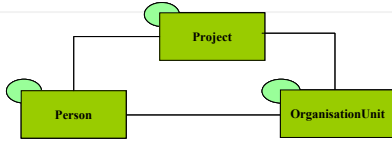
ID
URI
Acronym
StartDate
EndDate
Title
Abstract
Keywords

Person

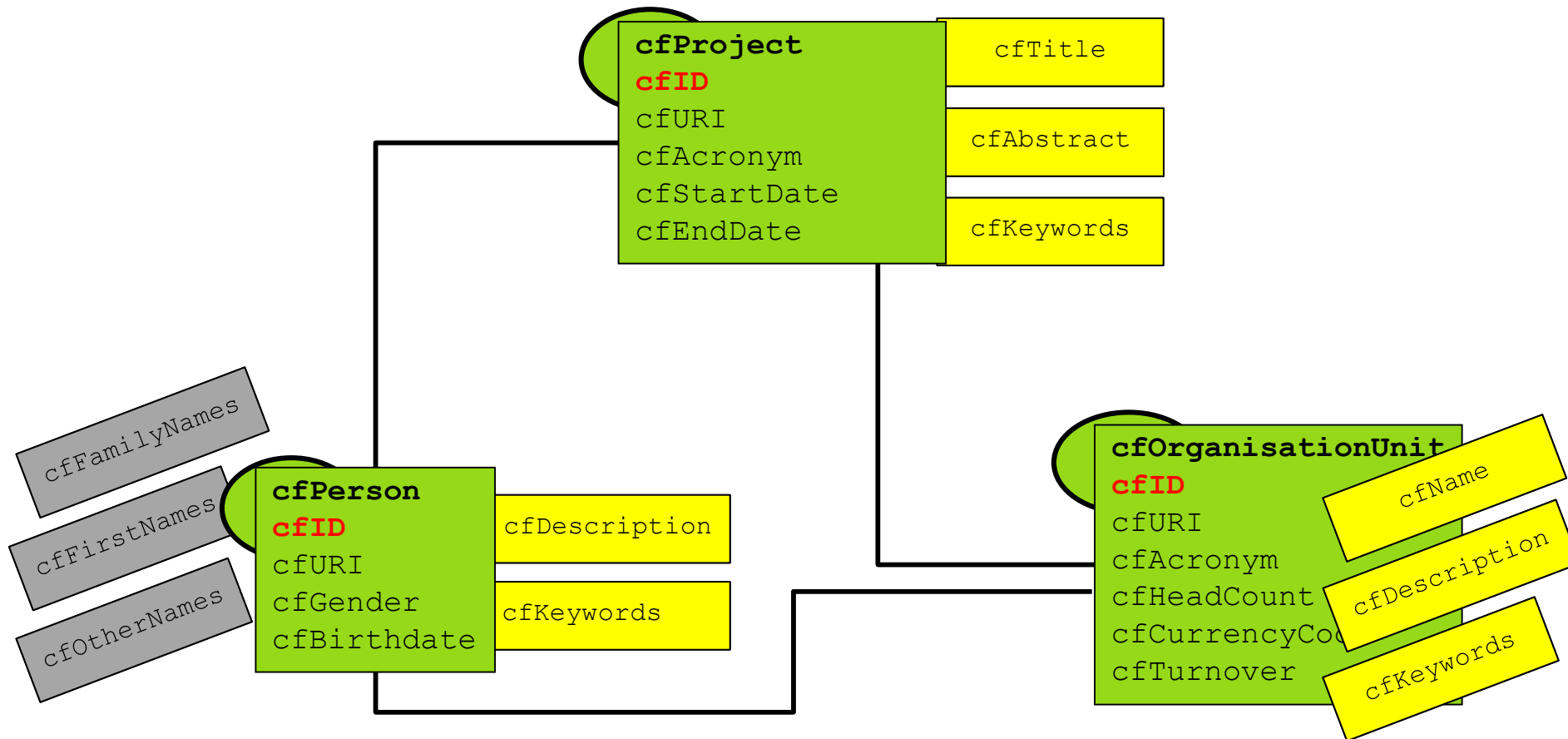
ID
URI
Gender
FirstNames
OtherNames
FamilyNames
NameVariants
ResearchInterest
Keywords

OrganisationUnit

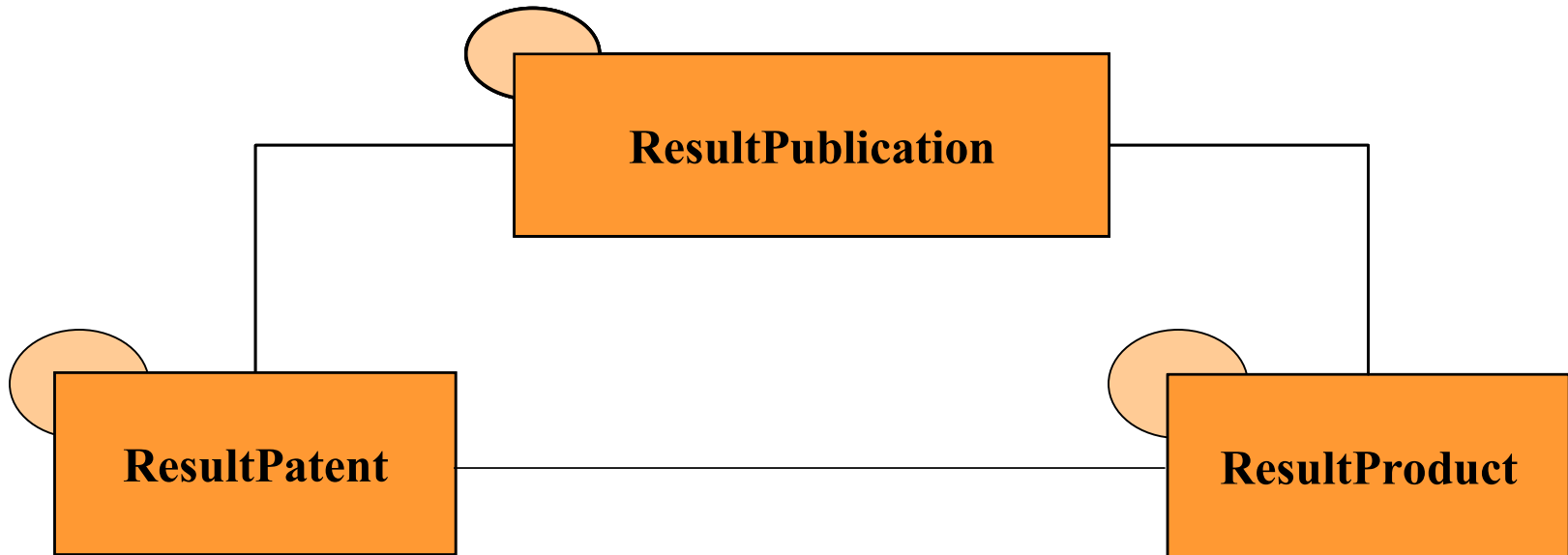
ID
URI
Acronym
Name
HeadCount
CurrencyCode
Turnover
ResearchActivity
Keywords



CERIF Base Entities

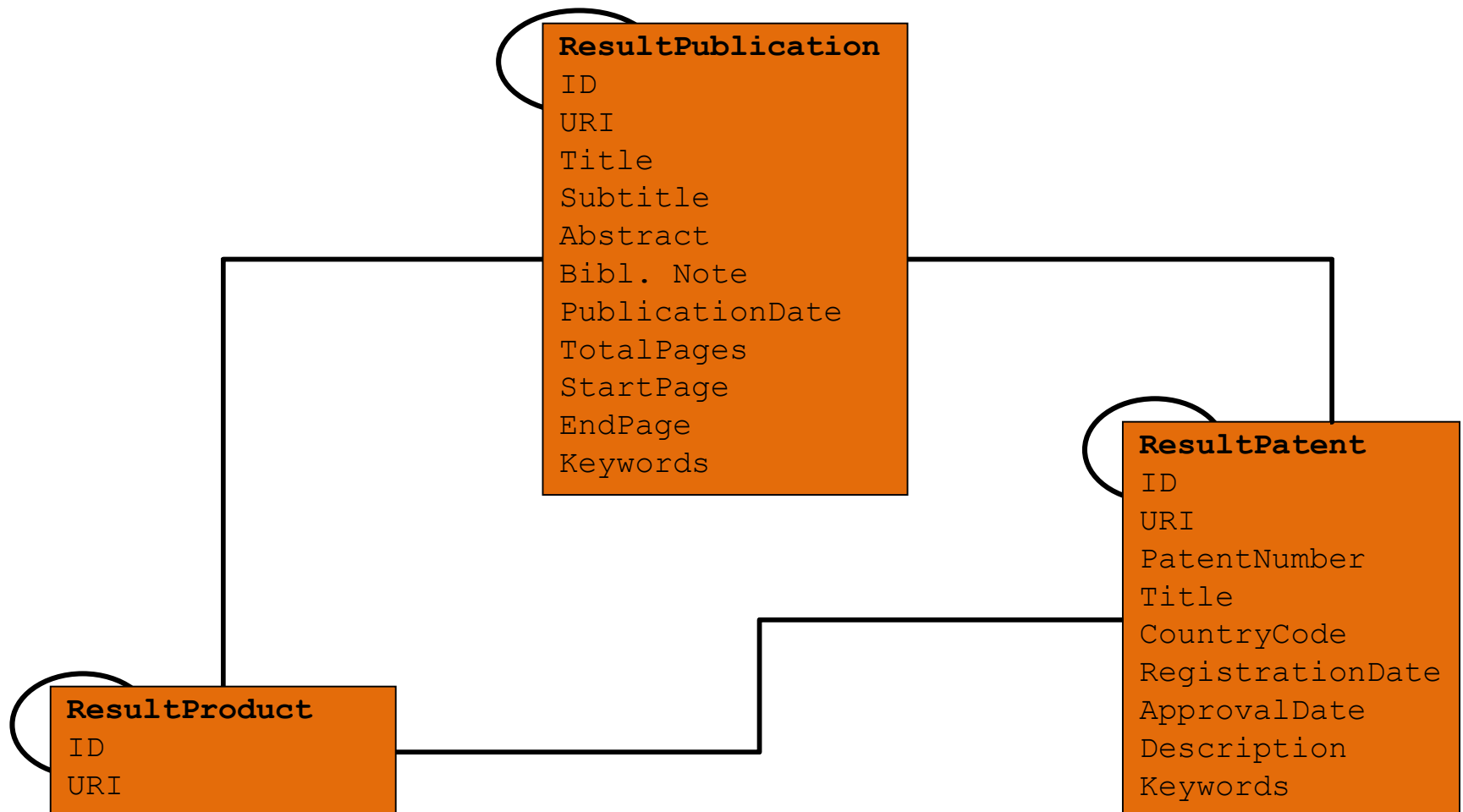


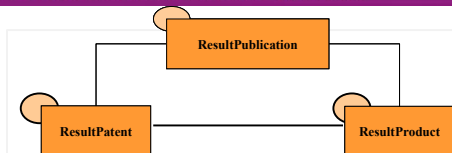
CERIF Result Entities



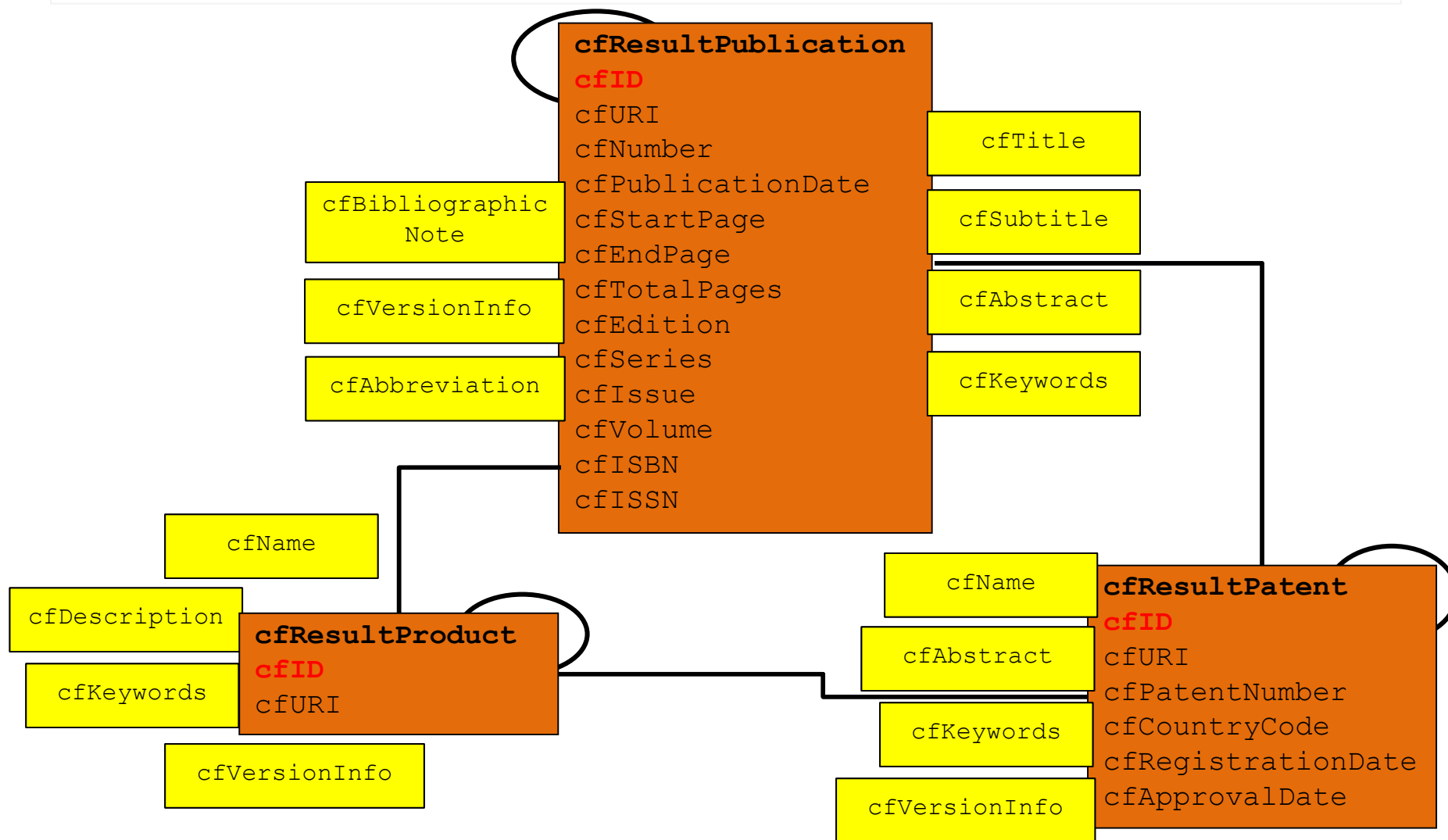


CERIF Result Entities

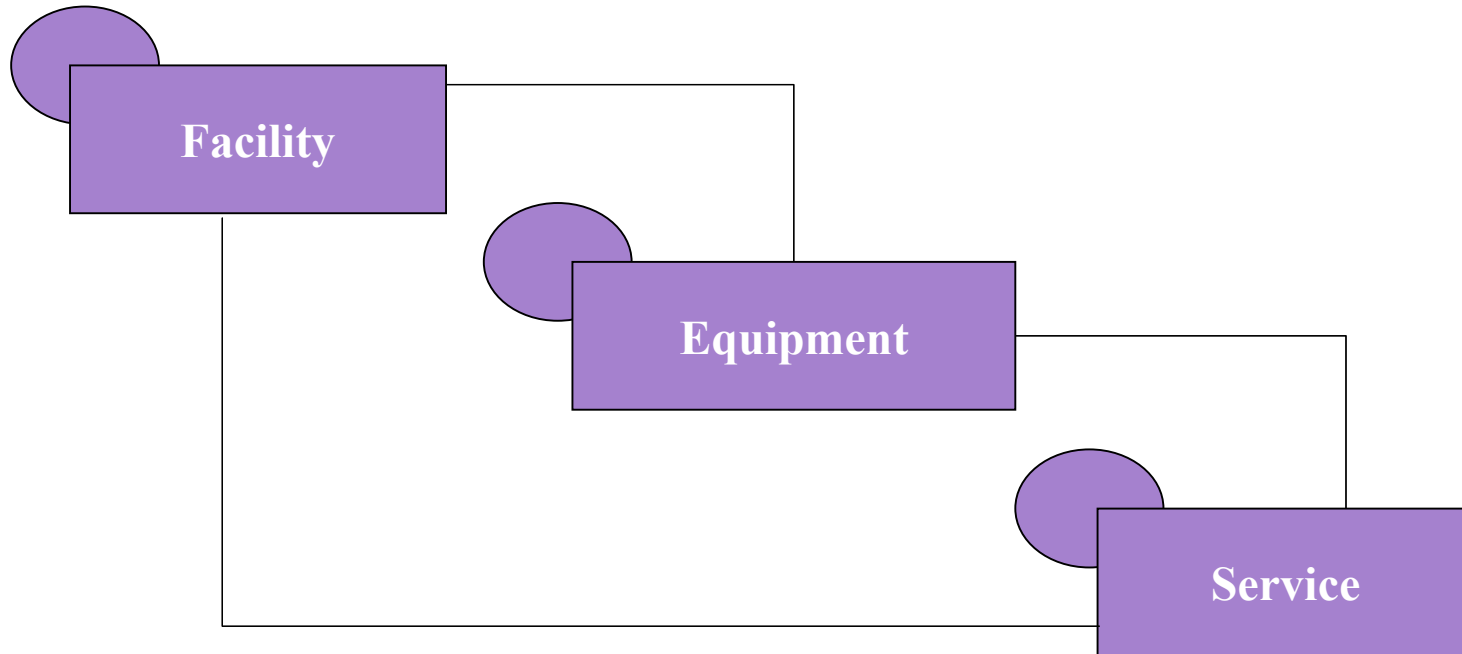


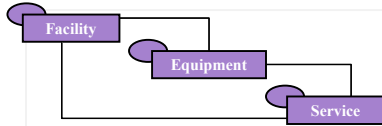


CERIF Result Entities

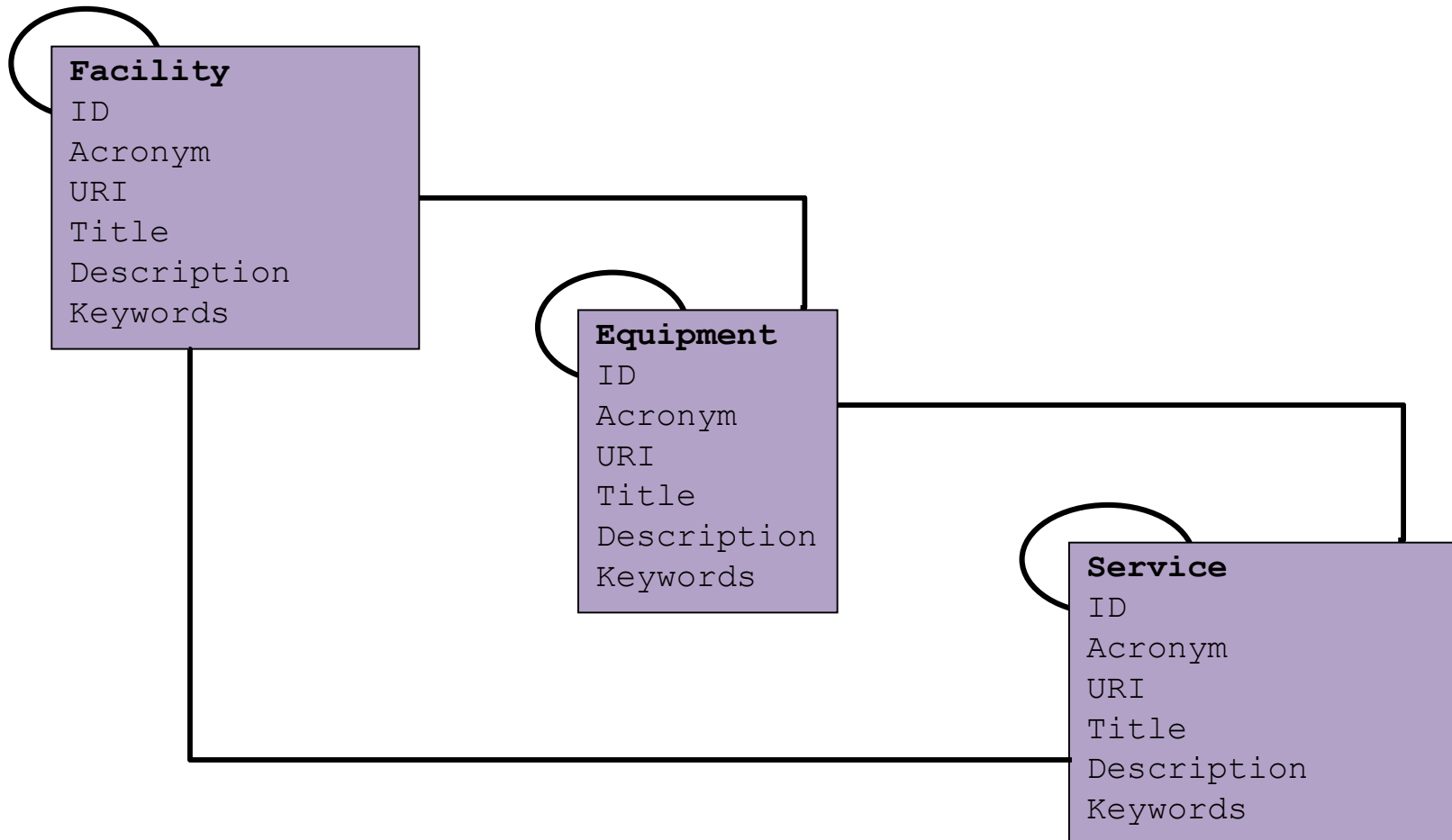


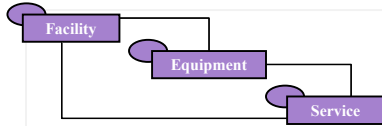
CERIF Infrastructure Entities



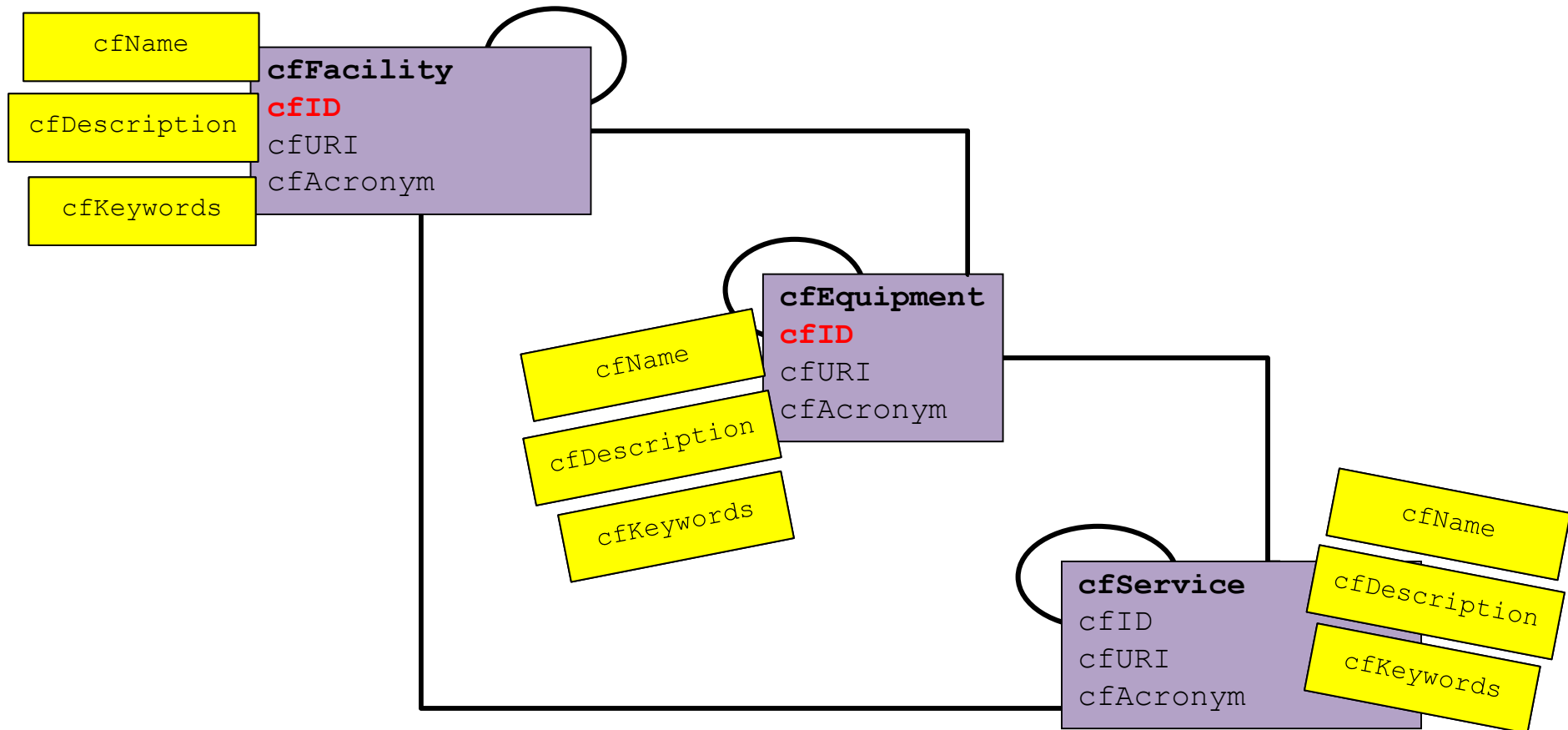


CERIF Infrastructure Entities





CERIF Infrastructure Entities

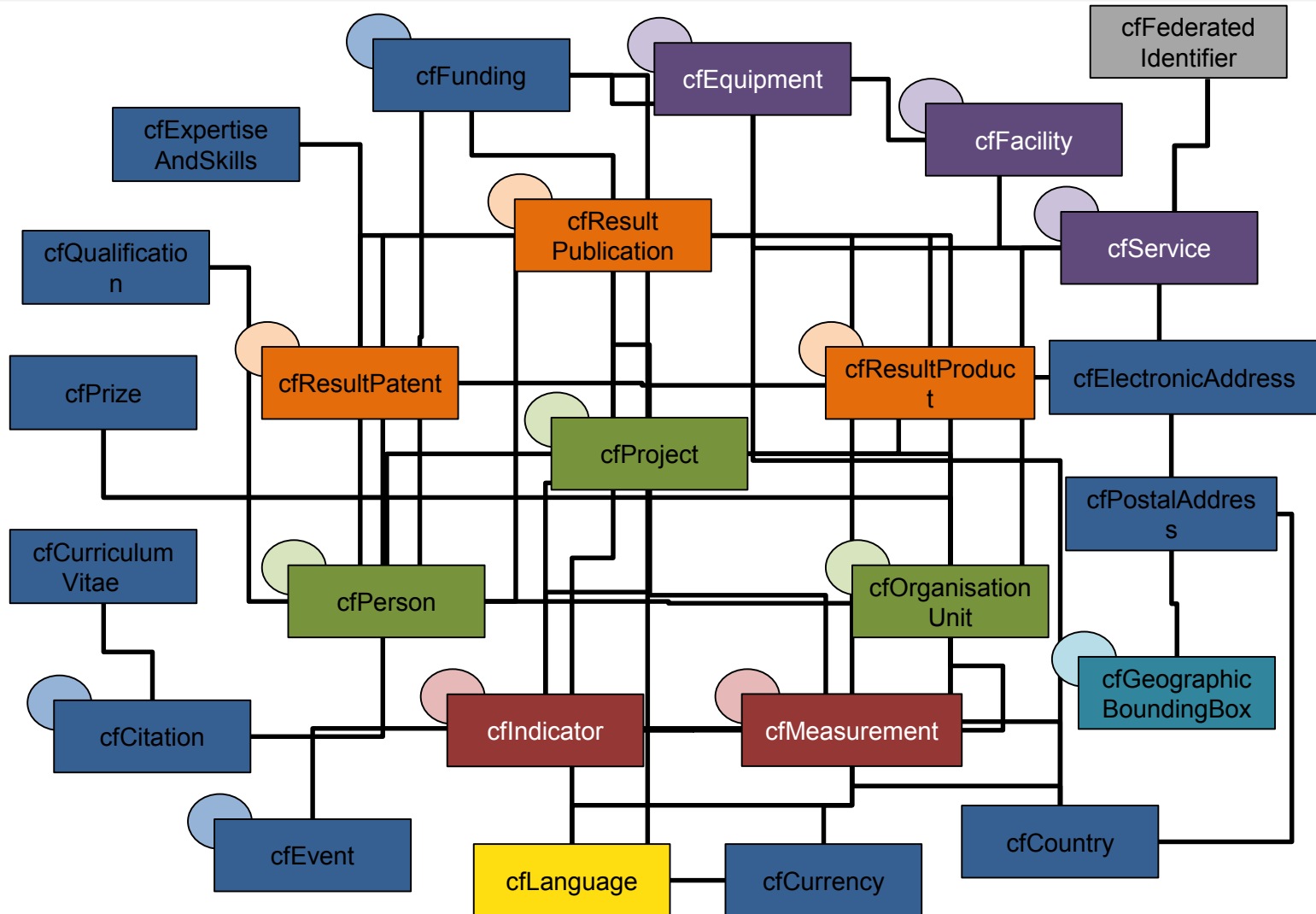


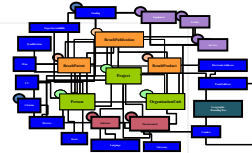
CERIF General Pattern

A typical CERIF entity has:

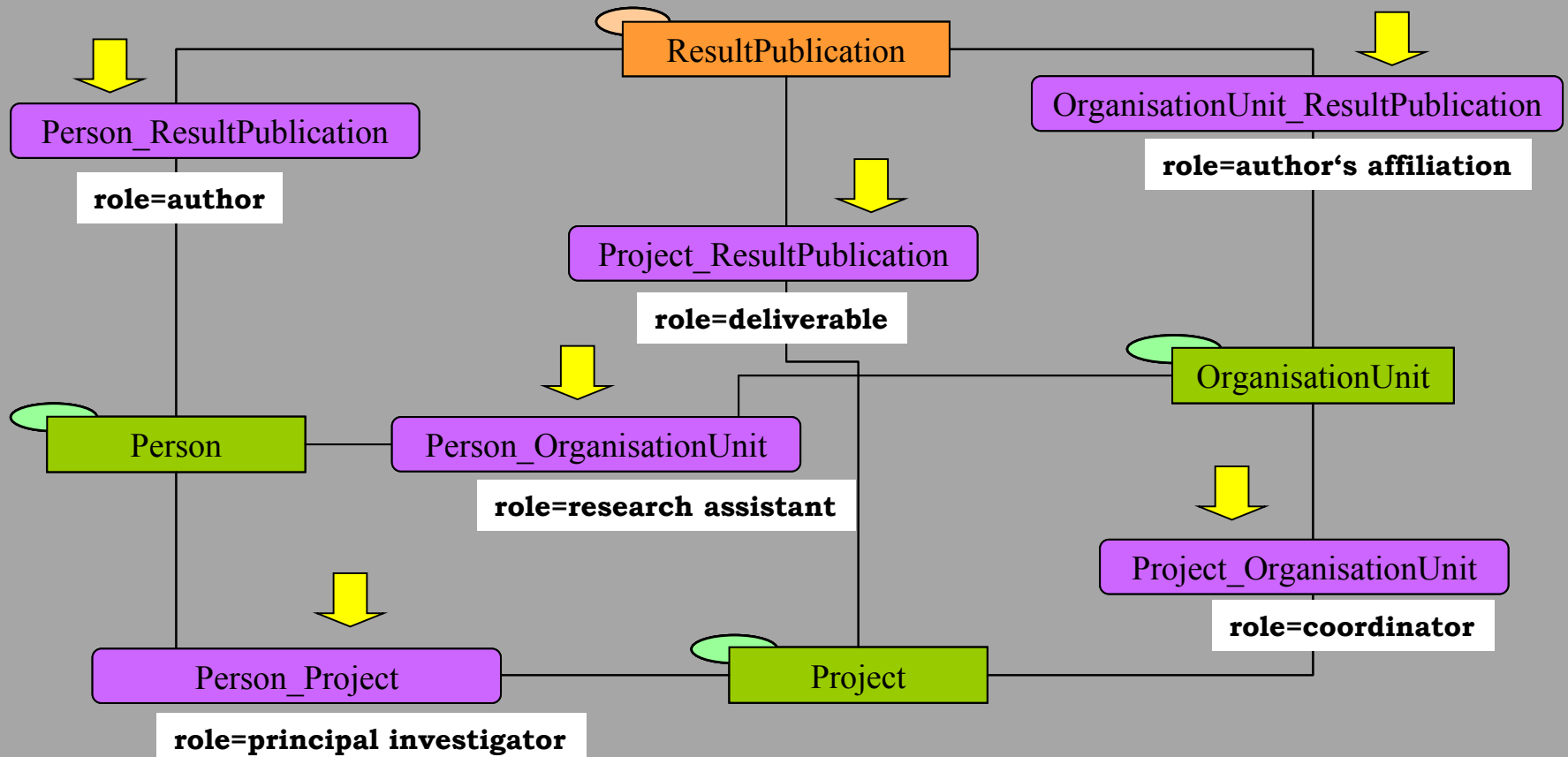
- Identifier (internal)
- Attributes
 - the basic ones
 - the multi-lingual ones
- External Identifiers
- Classifications
 - Type
 - Status
 - Subject area
- Links
 - to other entities
 - recursive

CERIF 1.6

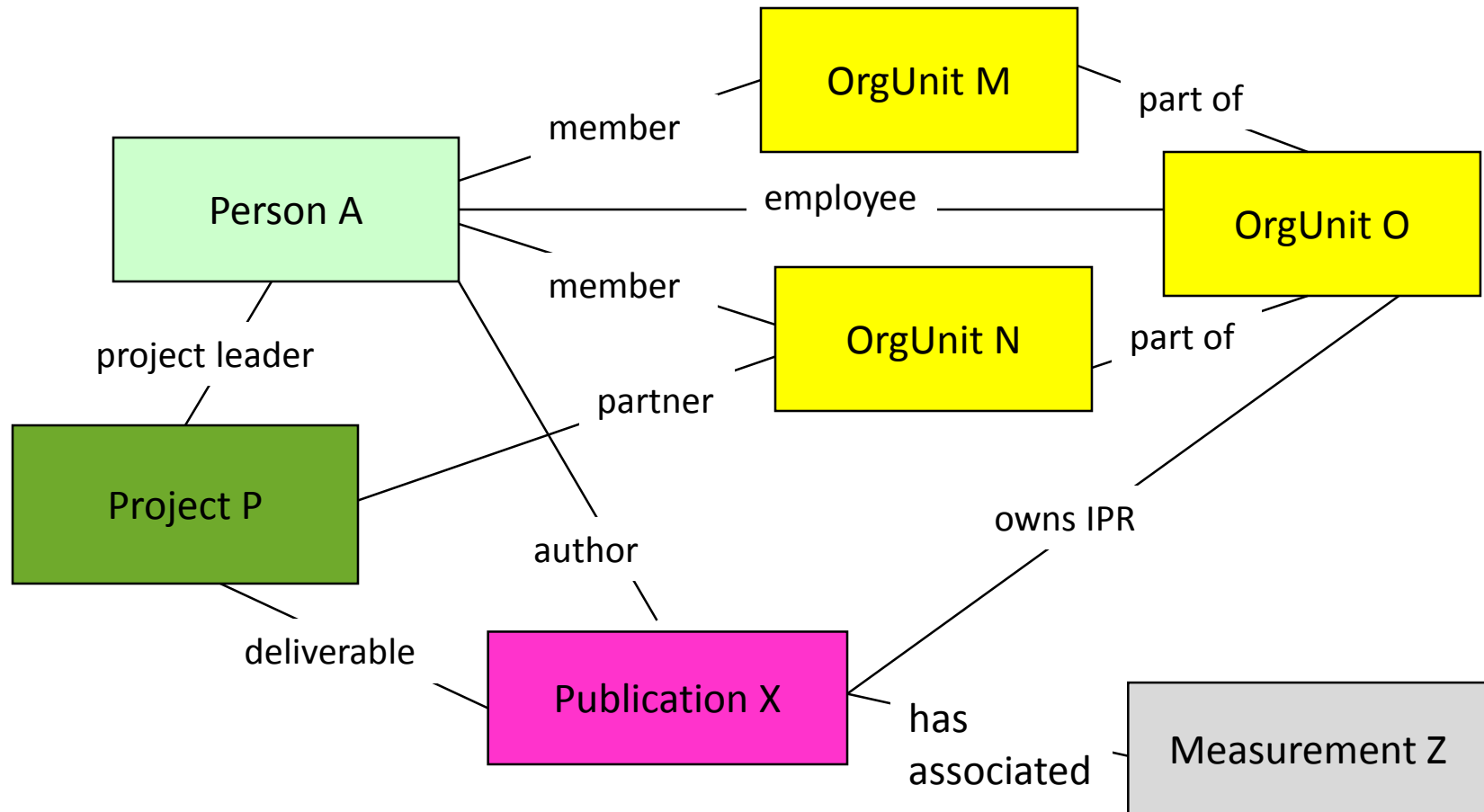




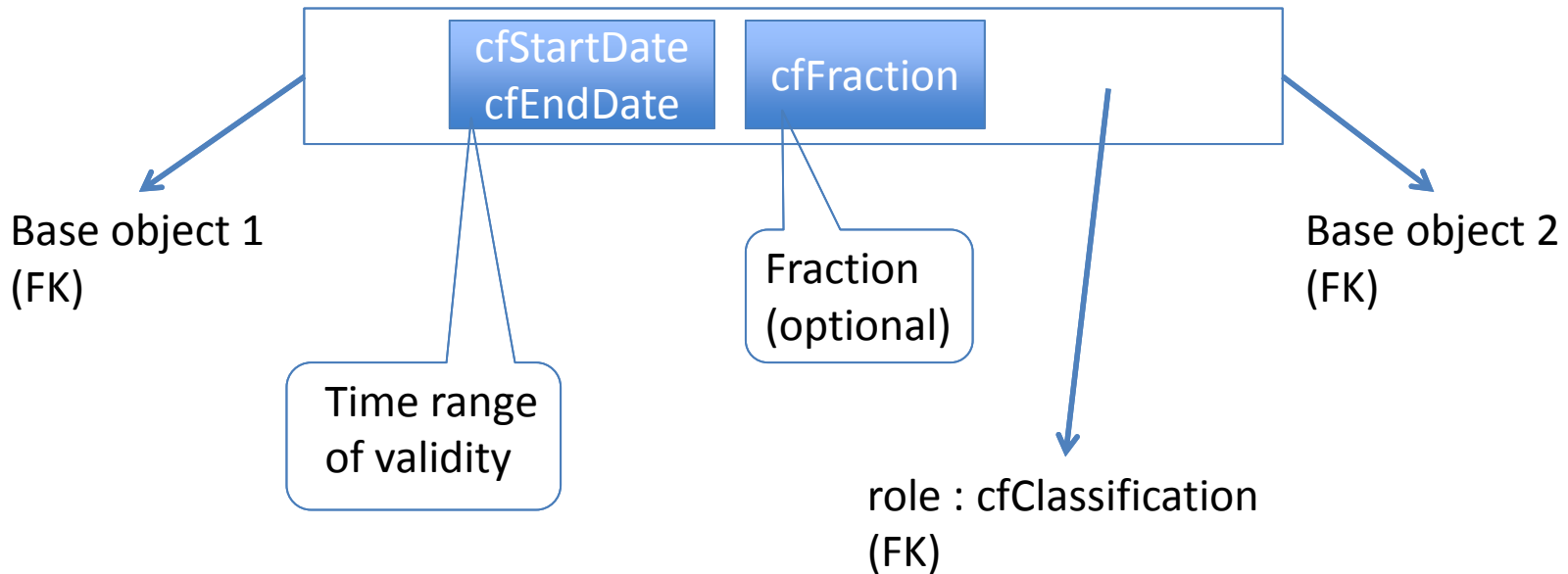
Some CERIF Link Entities



Result_Publication Instance Diagram (slide by Keith Jeffery)



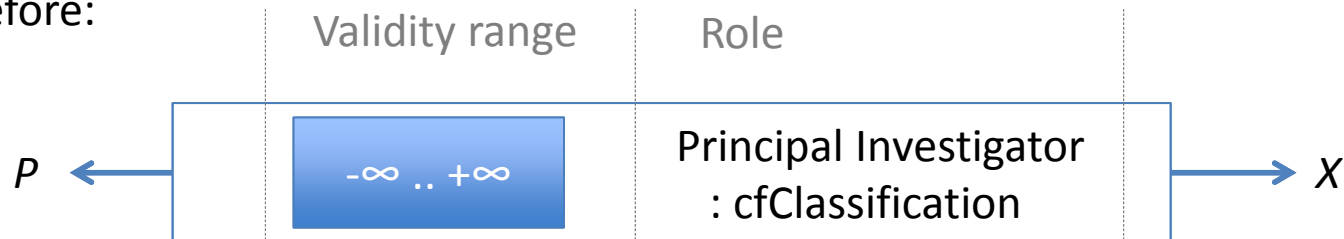
Generic Linking Entity Structure



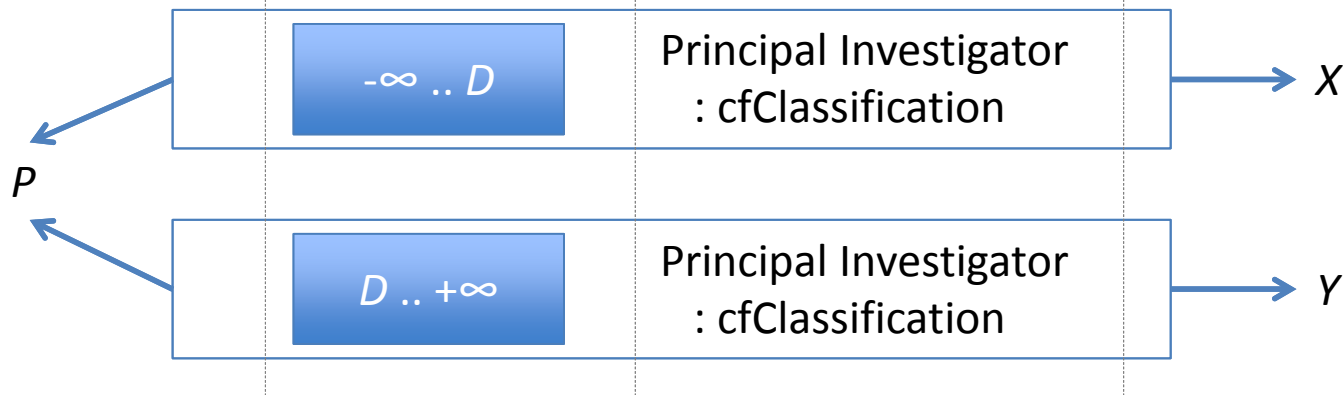
Recording Change in CERIF

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

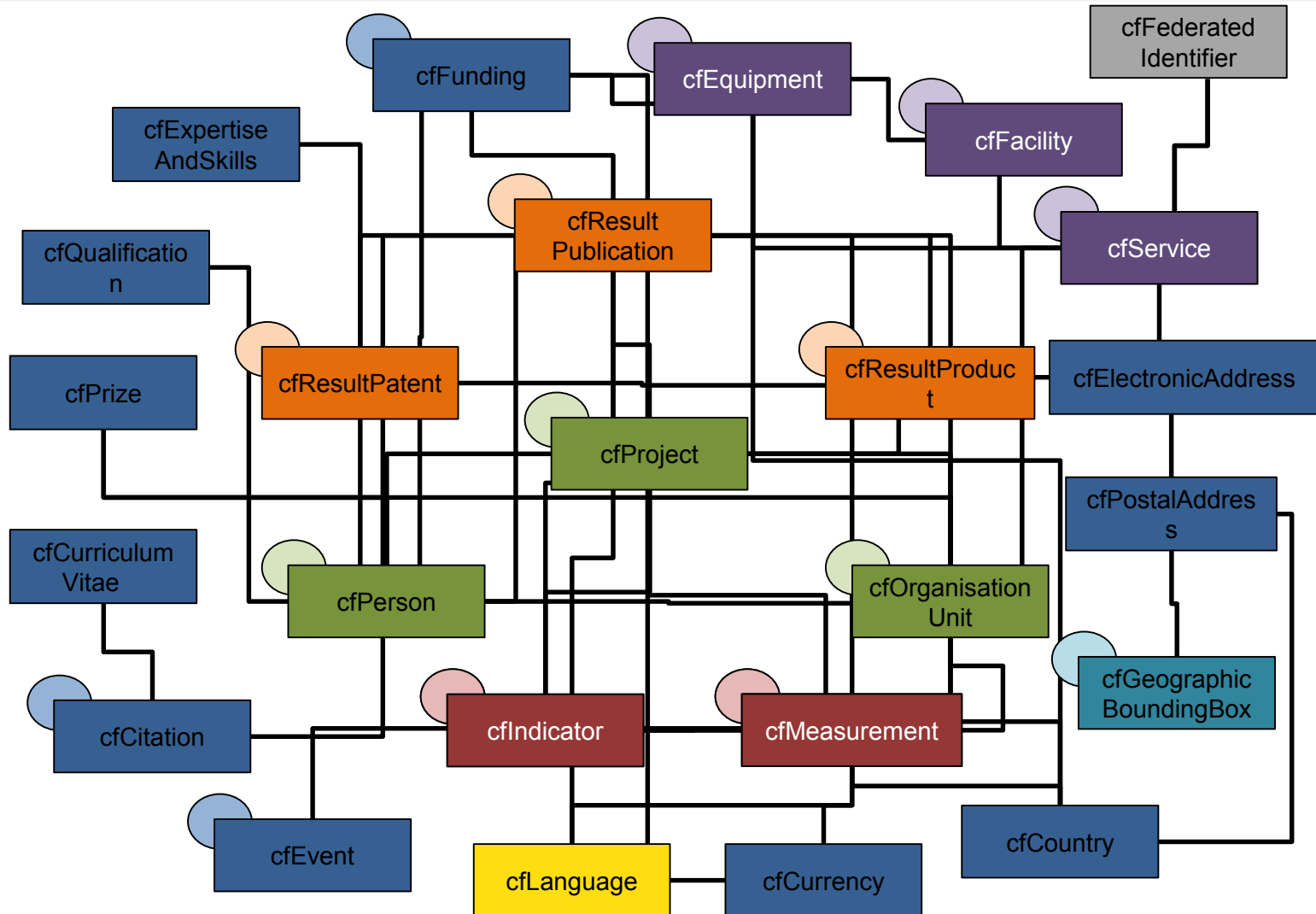
Before:



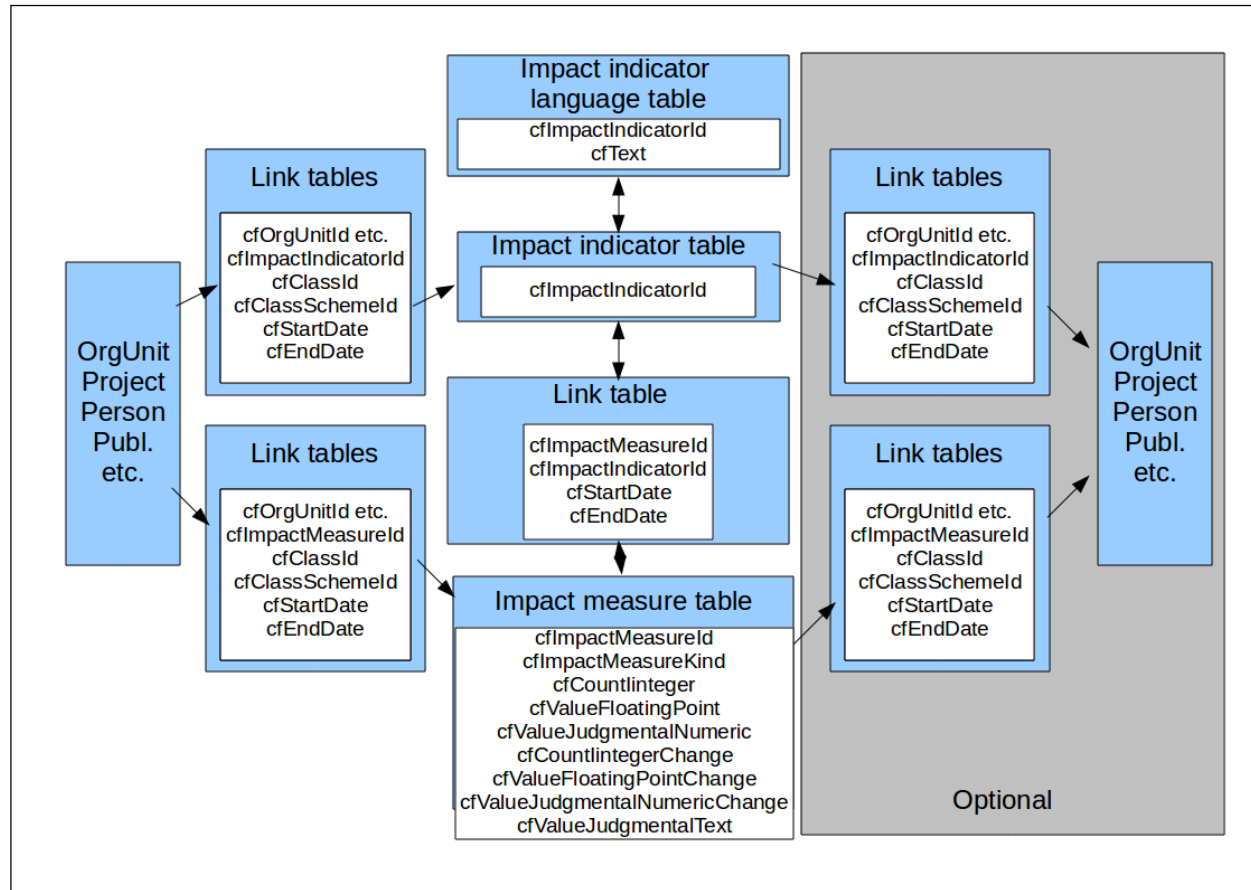
After:



CERIF 1.6



Measuring Impact in CERIF (MICE)



MICE, a JISC-funded Project coordinated by Richard Gartner, Kings College, London, UK

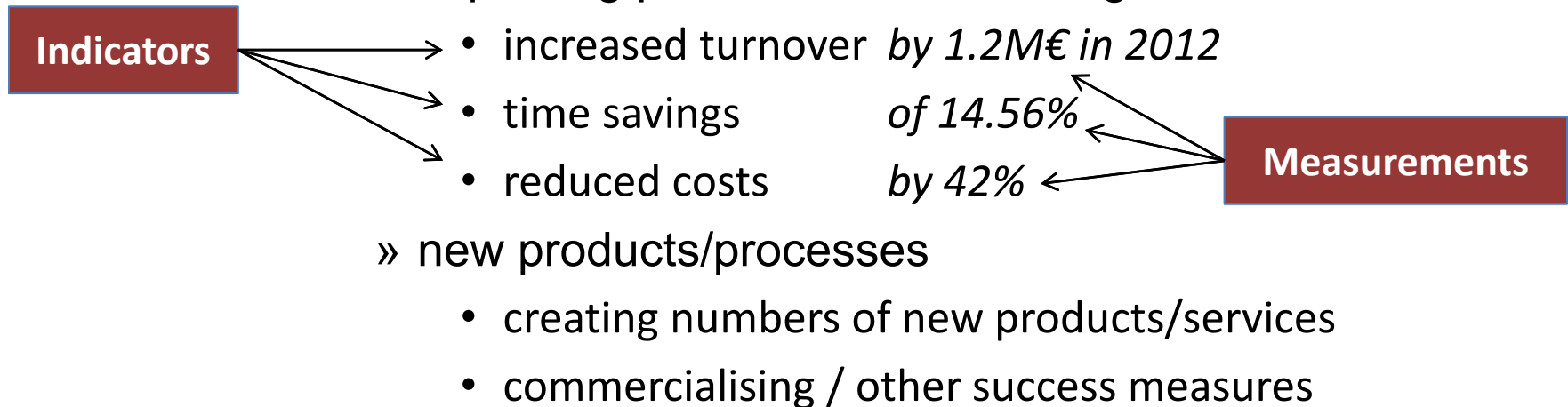
Measurement & Indicator (some examples)

– economic and commercial

- economic

- impact on business

- » improving performance of existing businesses

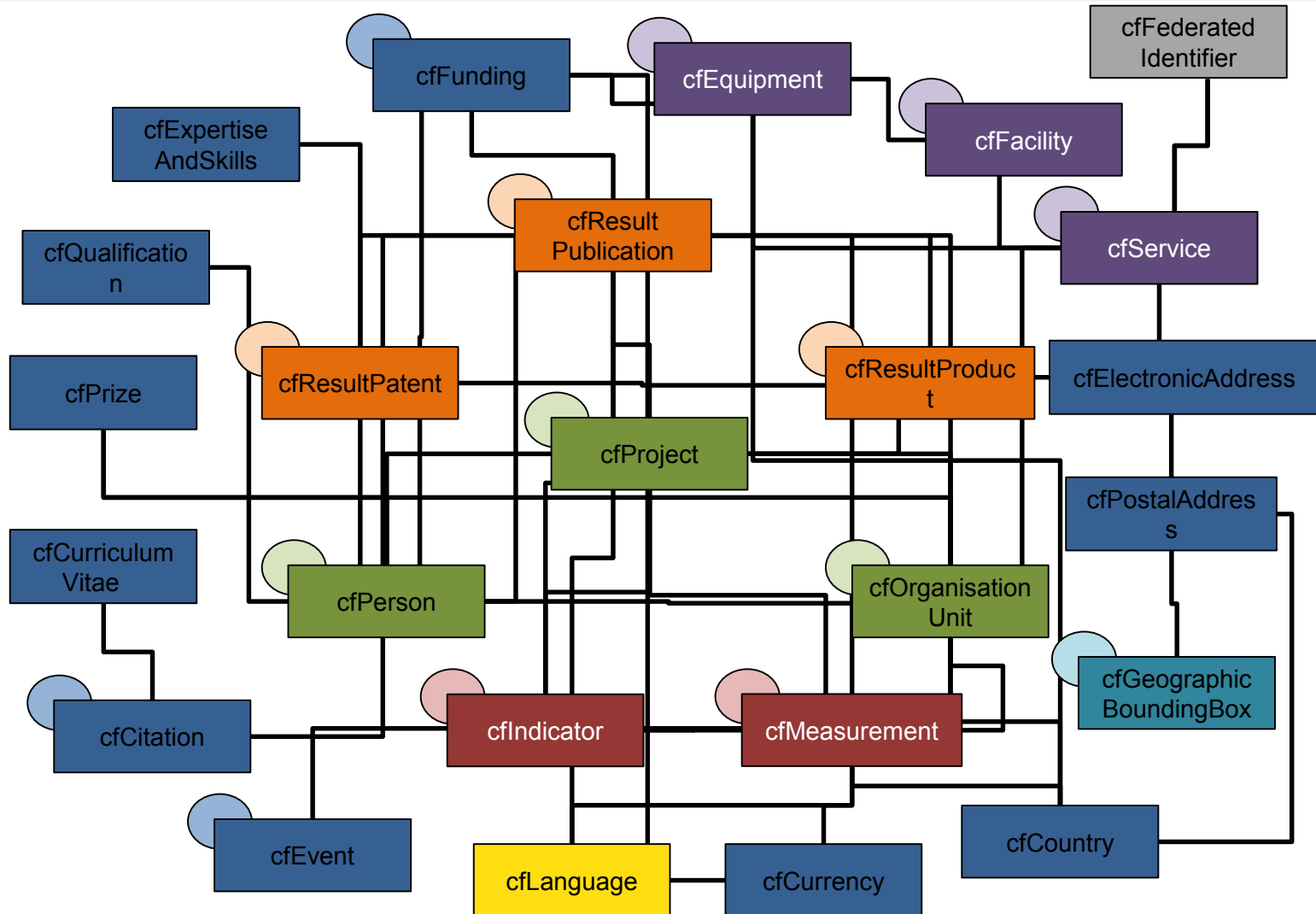


- » new products/processes

- creating numbers of new products/services
 - commercialising / other success measures

Extract from the MICE List of Indicators

CERIF 1.6



CERIF Semantic Layer

Allows to capture any Schema or Structure

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

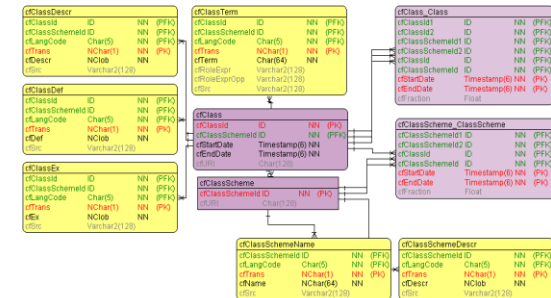
Open / Extensible in all directions

- New Schemas
- New Concepts / Terms
- New Relationships

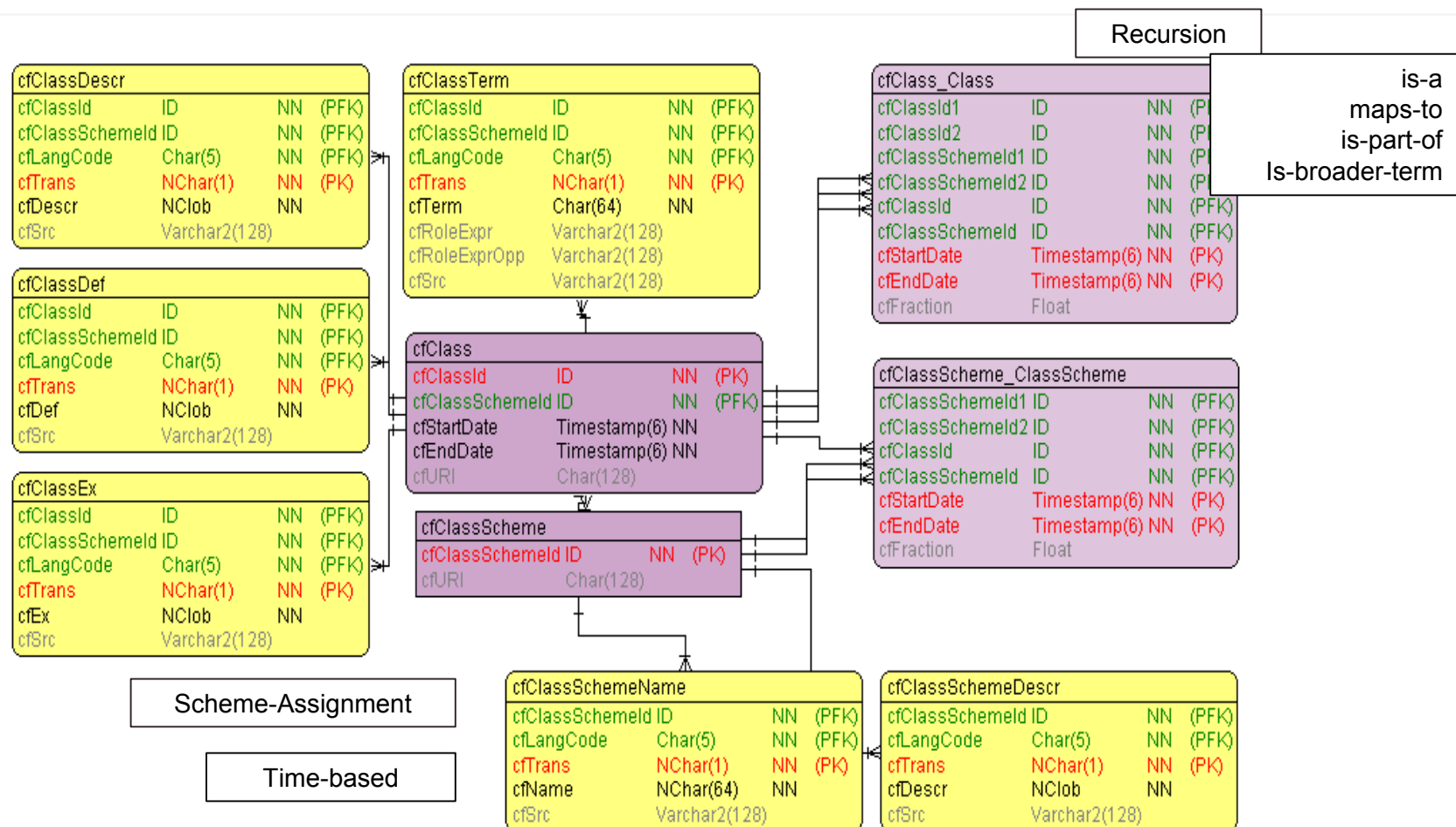
Enables to manage

- Roles / Types Semantics
- Subject Headings
- Archiving (Time component)

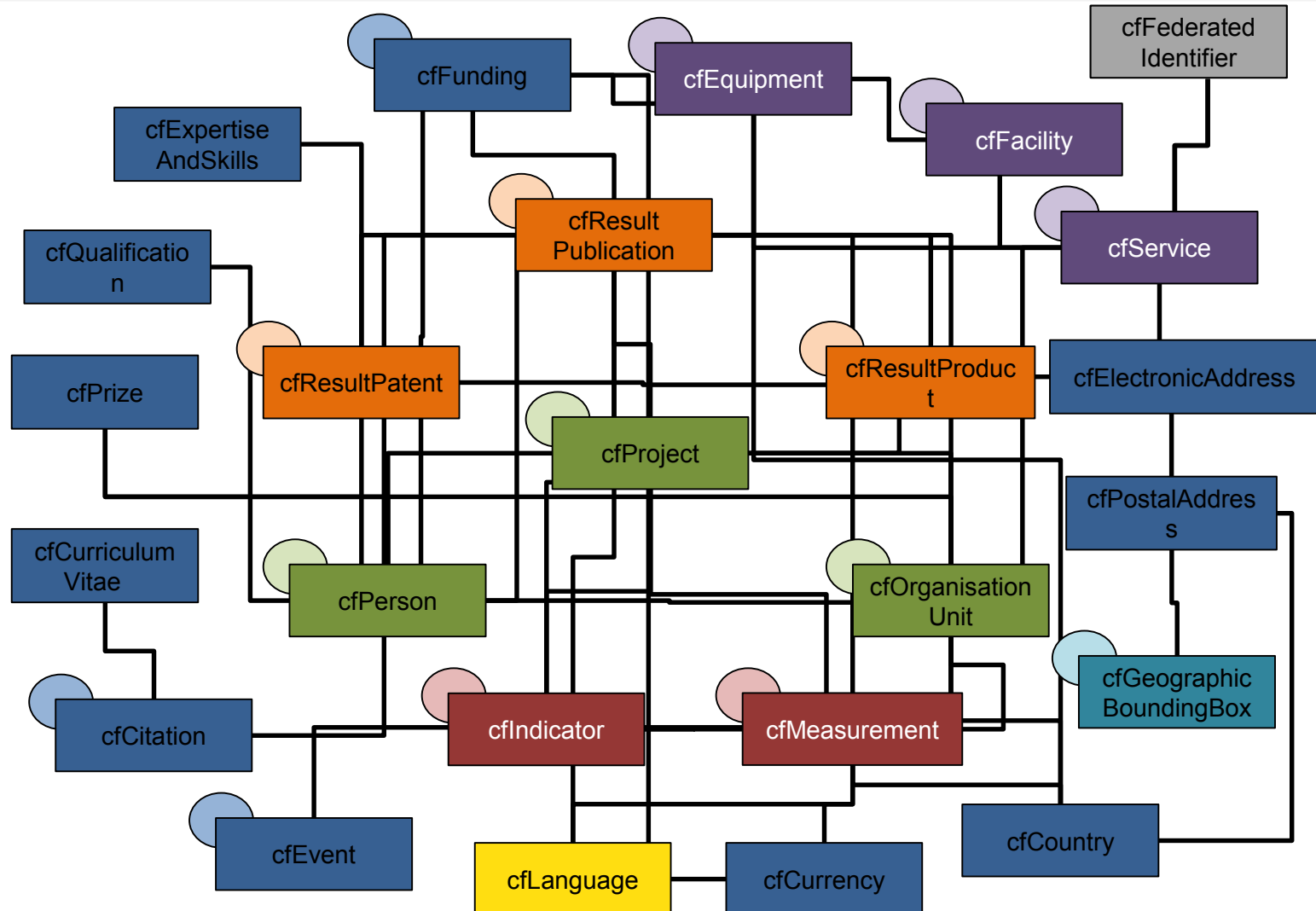
Allows for Mappings between Schemes



CERIF Semantic Layer (Declared Semantics)



CERIF 1.6



CERIF Federated Identifiers

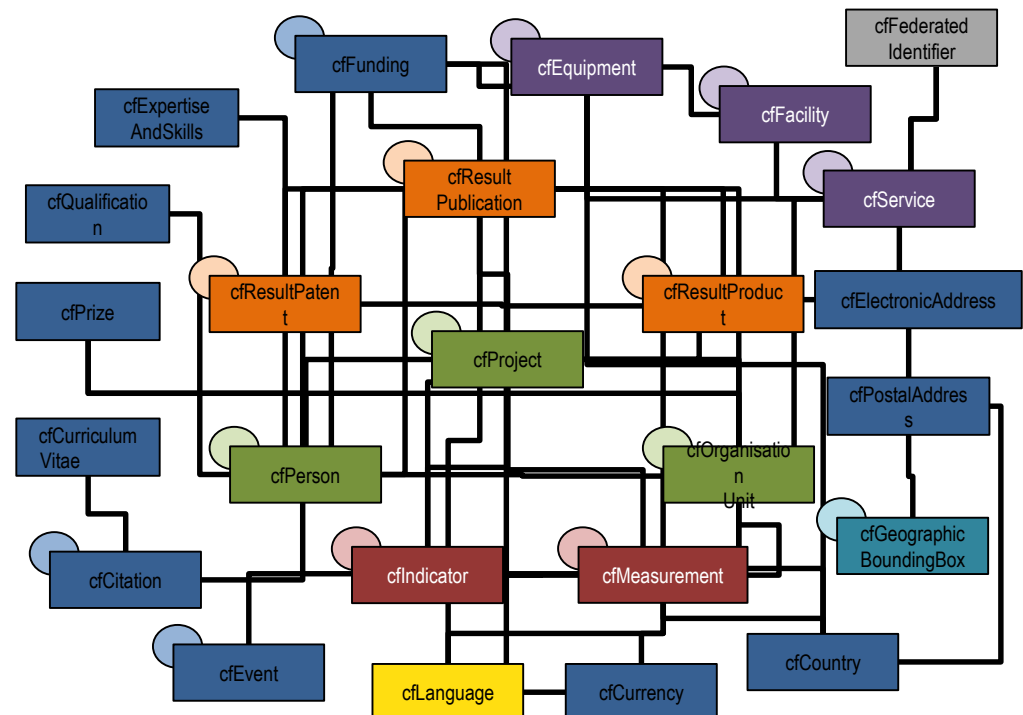
- ResultPublication
 - ISBN
 - ISSN
 - DOI
 - WoS Accession Number
 - Scopus EID
 - PubMed Central ID
- Person
 - Social Security Number
 - Staff Id in HR system
 - Author identifier
 - ORCID, IdRef, DAI, ResearcherID, ScopusID
- Project/Grant
 - Funder's reference number
 - Organisation's reference number
- Organisation
 - VAT Identification Number
 - FundRefID
 - GridID
- Classification
 - External Code

CERIF Federated Identifiers

- Records the “tag” by which an object is known elsewhere
- For any Base, Result, Infrastructure, or 2nd Level entity
- “Identifier Types” classification scheme
- (optionally) Connected to a Service representing the issuer of the identifier
 - Usually an information system

CERIF XML Interchange Format

- XML Schema
- Based on the ER model
- Undergone a big update



CERIF Profiles

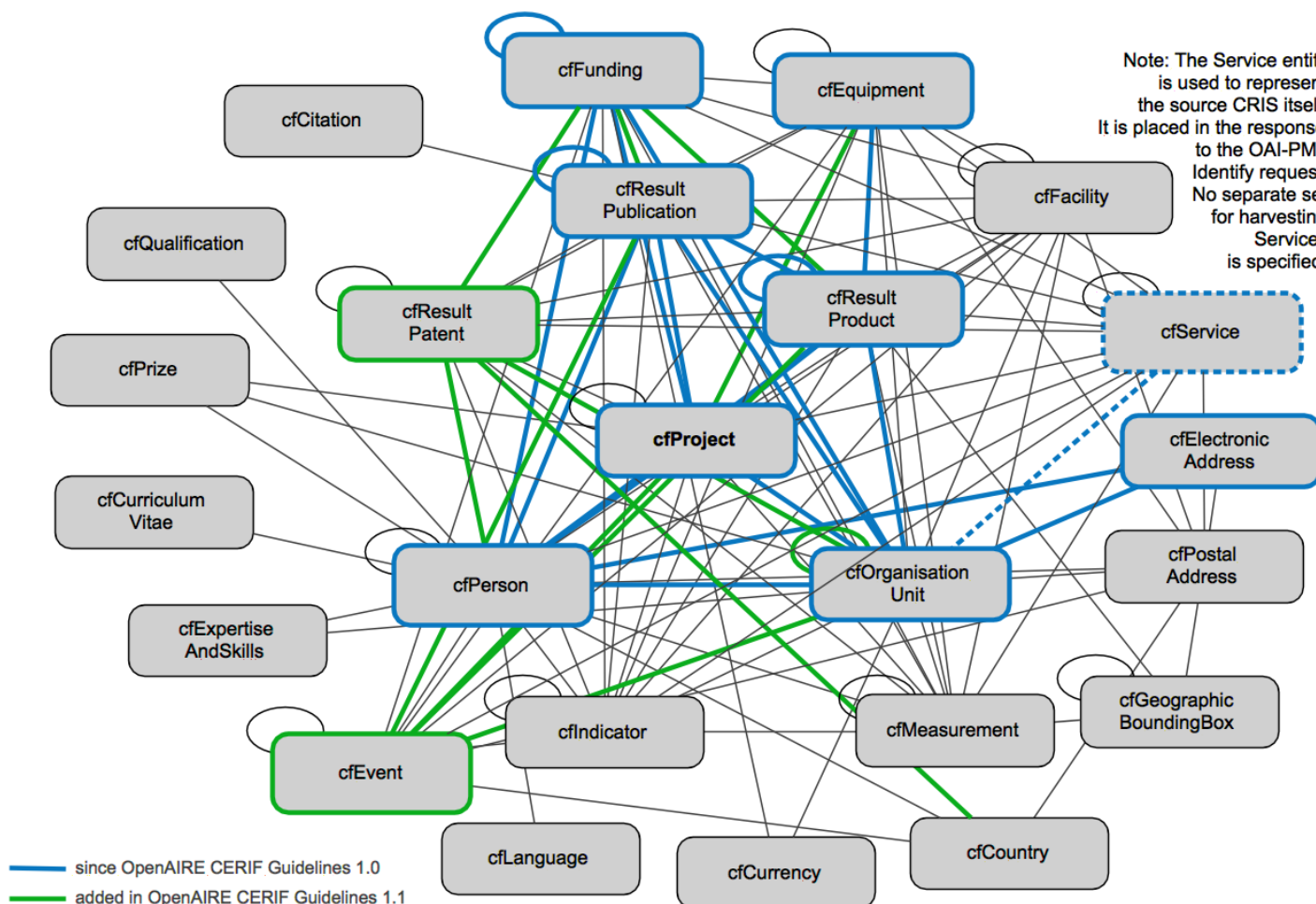
- Entities & attributes:
 - Profile $\subset$ CERIF
- Semantic vocabularies:
 - Profile – more specific
 - Sources: CERIF & beyond
- Integrity constraints:
 - Profile $\supset$ CERIF

Profile data is
CERIF

Producers know
what to include

Consumers know
what to expect

Subset of CERIF entities



```
<Publication xmlns="https://www.openaire.eu/cerif-profile/1.1/" 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>
      <Identifier type="https://www.openaire.eu/cerif-profile/vocab/IdentifierTypes#EISSN">1558-5646</Identifier>
      <Publishers>
        <!-- [ ... ] -->
      </Publishers>
    </Publication>
  </PublishedIn>
  <Volume>66</Volume>
  <Issue>5</Issue>
  <StartPage>1474</StartPage>
  <EndPage>1489</EndPage>
  <Identifier type="https://www.openaire.eu/cerif-profile/vocab/IdentifierTypes#DOI">10.1111/J.1558-5646.2011.01539.X</Identifier>
  <!-- [ ... ] -->
</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>
      <Identifier type="https://www.openaire.eu/cerif-
profile/vocab/IdentifierTypes#DOI">10.5061/DRYAD.4GH6HF5G</Identifier>
      <!-- [...] -->
    </Product>
  </References>
</Publication>
```

CERIF development

By the CERIF Task Group of euroCRIS

Adopting open-source software projects
tools & best practices:

→ <https://github.com/EuroCRIS/CERIF-DataModel>

→ 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			

CERIF highlights

- Right level of abstraction
- Normalized model
 - Record information only once
 - Reference rather than copy
- Versatile Semantic Layer
- Time-based relationships
- Clean design, regular structure

Metadata Layers

Discovery metadata

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

Contextual metadata

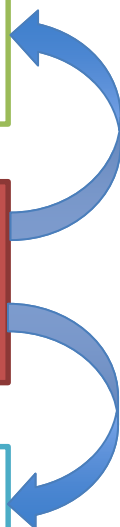
CERIF

Detailed metadata

Domain-specific standards

Generate

Reference





ALLEA

