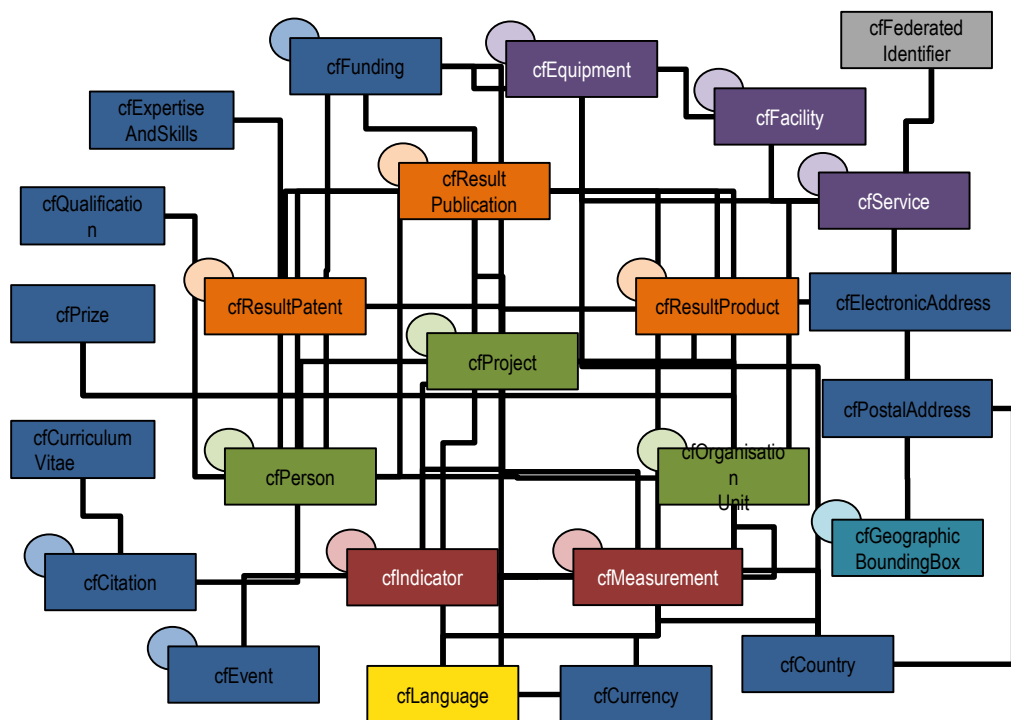


CERIF Tutorial



Jan Dvořák

May 29th, 2017

euroCRIS Spring Membership
Meeting
Dublin, Ireland

Jan Dvořák

jan.dvorak@ff.cuni.cz

euroCRIS

- CERIF TG Leader since 2013
- CRIS 2012 (Prague, June 2012) Org. Committee Chair

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

- Researcher & Lecturer

Czech Technical University, Computing and Information Centre

- IS Analyst

InfoScience Praha

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

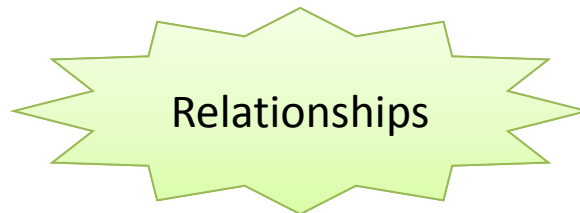
This deck of slides is based on the CERIF Tutorial by Brigitte Jörg

CERIF TG Leader 2004-2012

What is Research Information?

The process of research

- Research projects
- Funding
- Research infrastructures



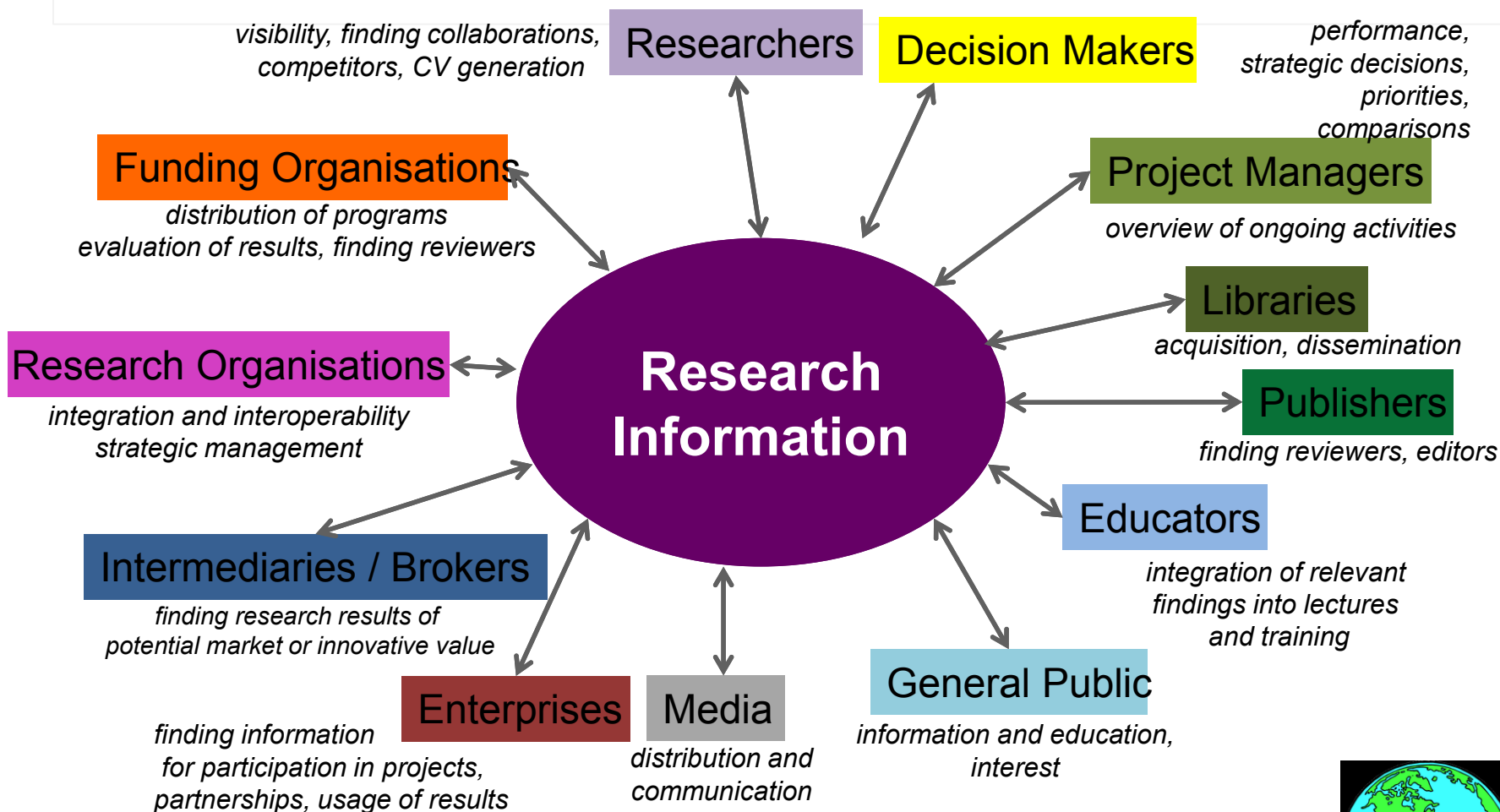
The research actors

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

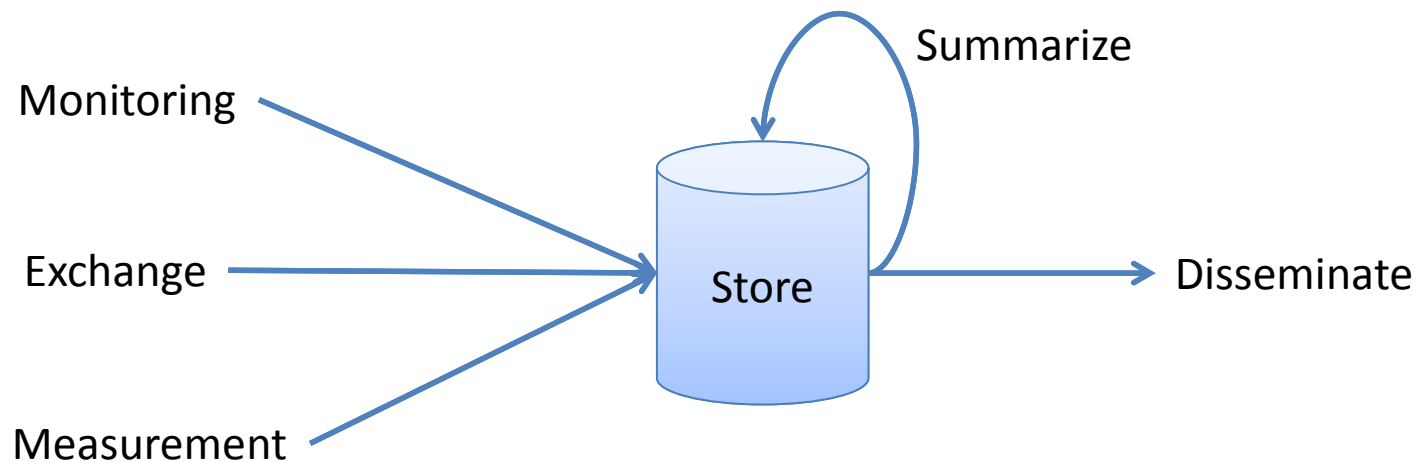
Research results

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

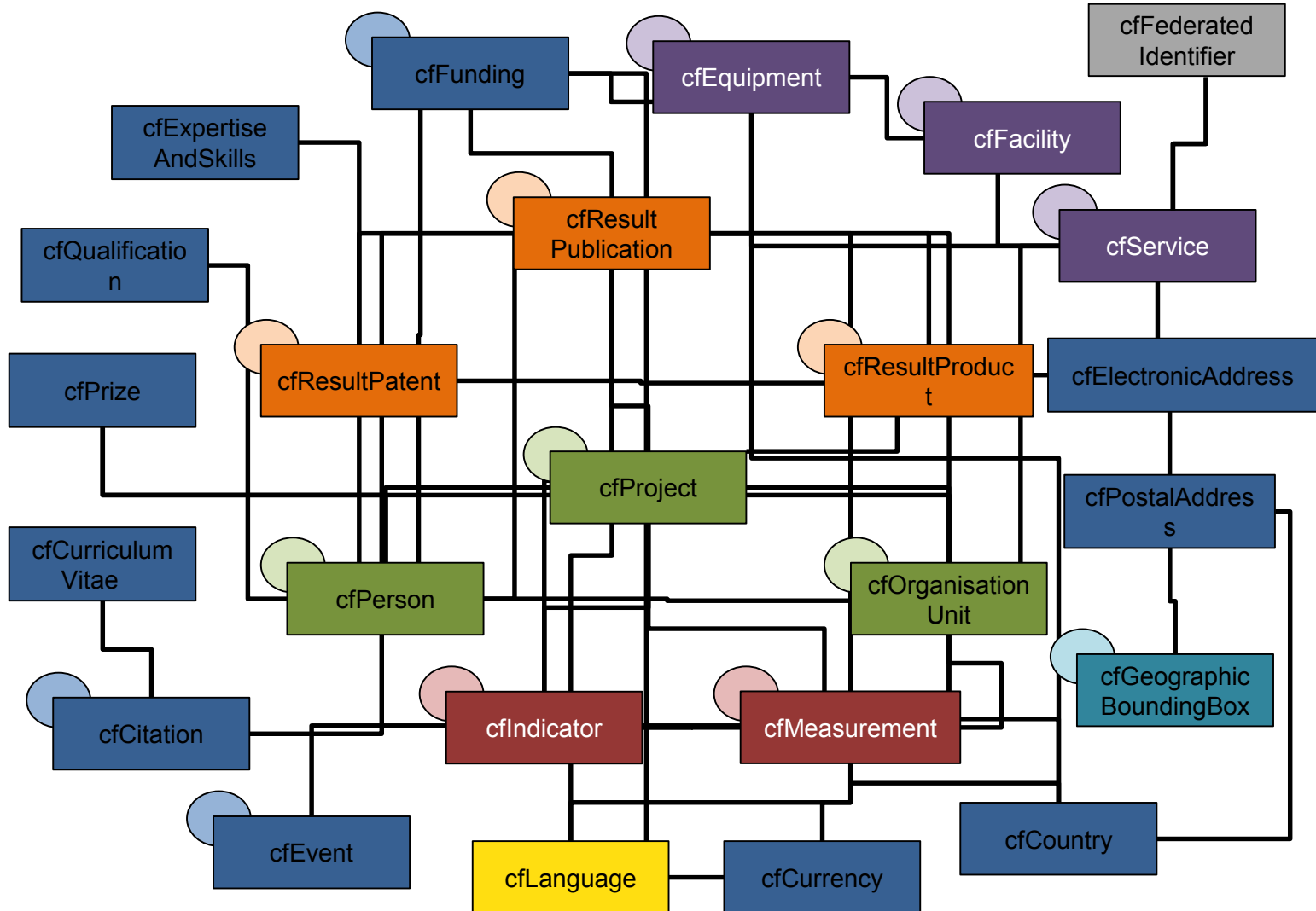
Who needs Research Information?



Research Information Life-Cycle



Common European Research Information Format



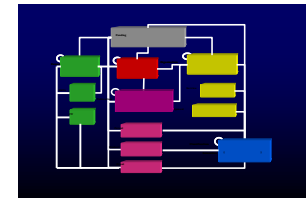
Common European Research Information Format

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

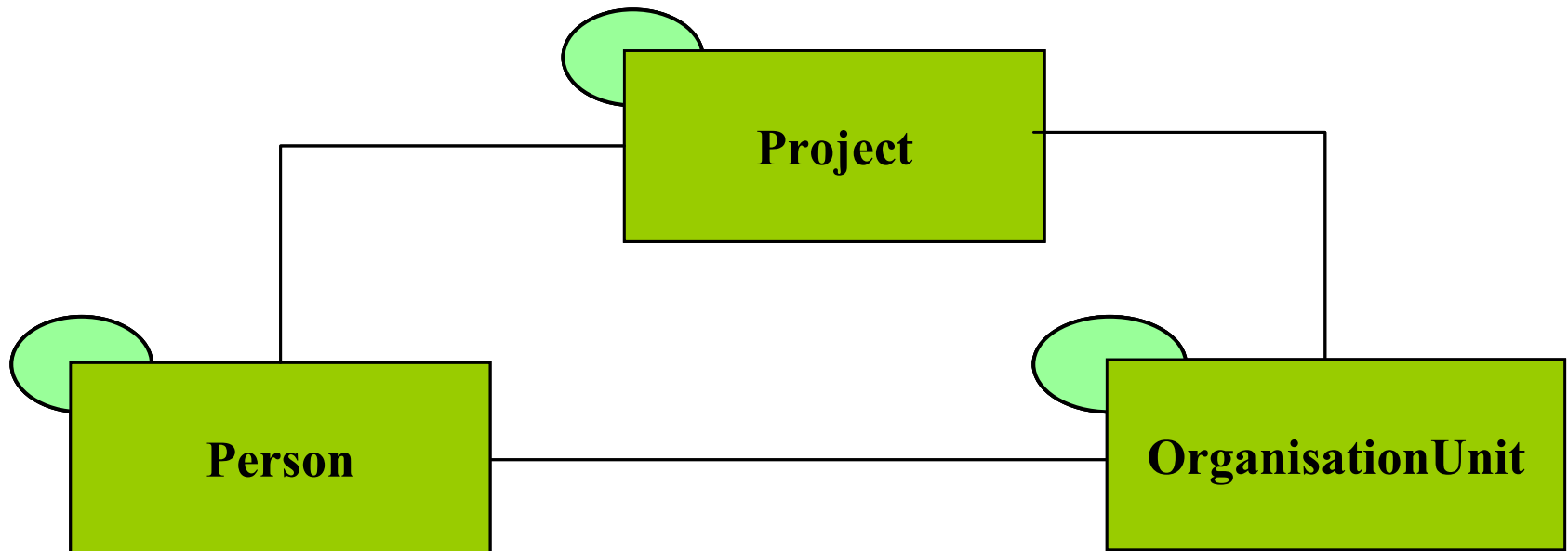
<http://cordis.europa.eu/cerif/>

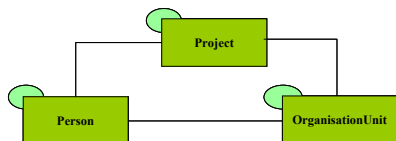
Model Levels

- **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



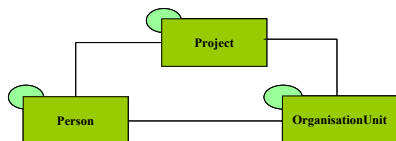
CERIF Base Entities



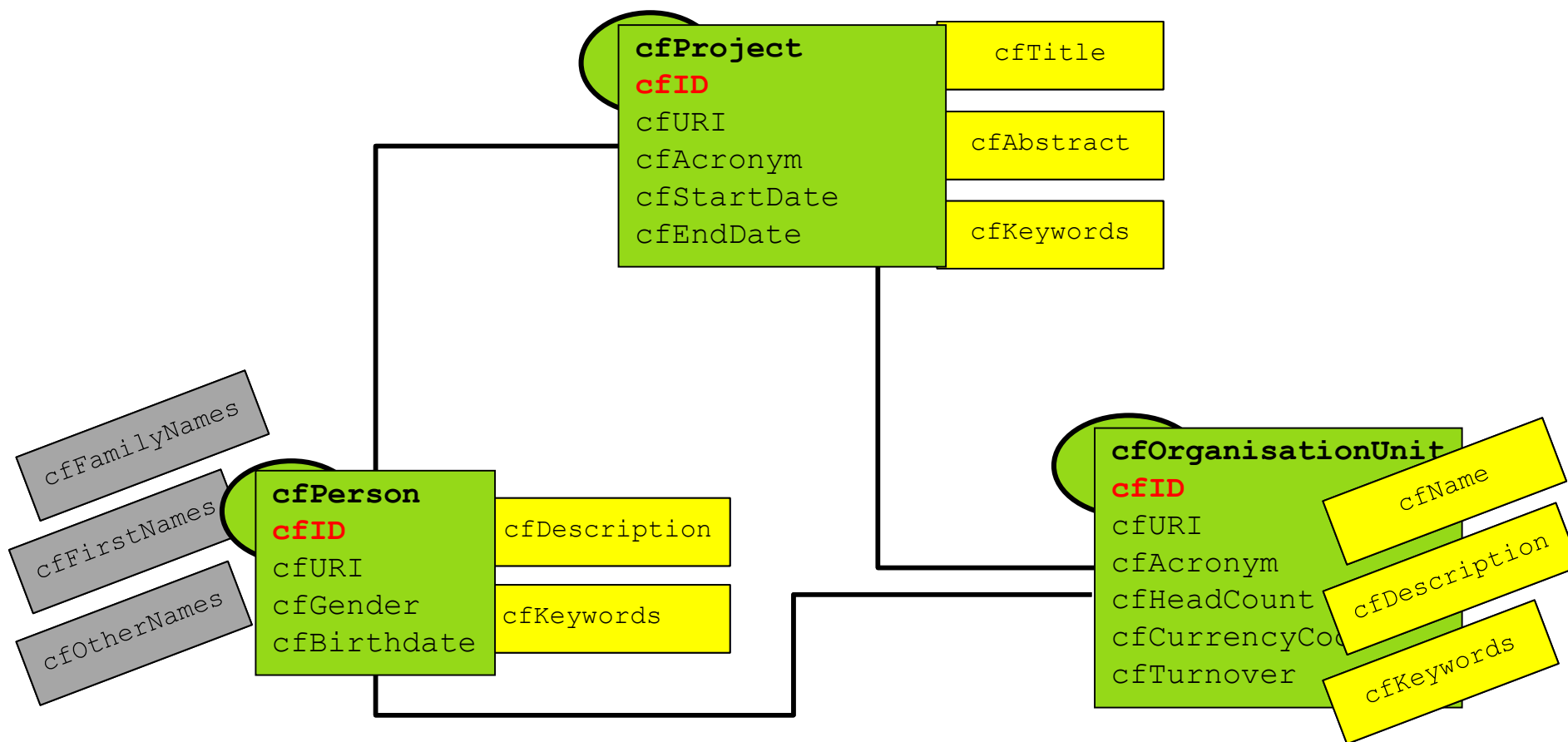


CERIF Base Entities

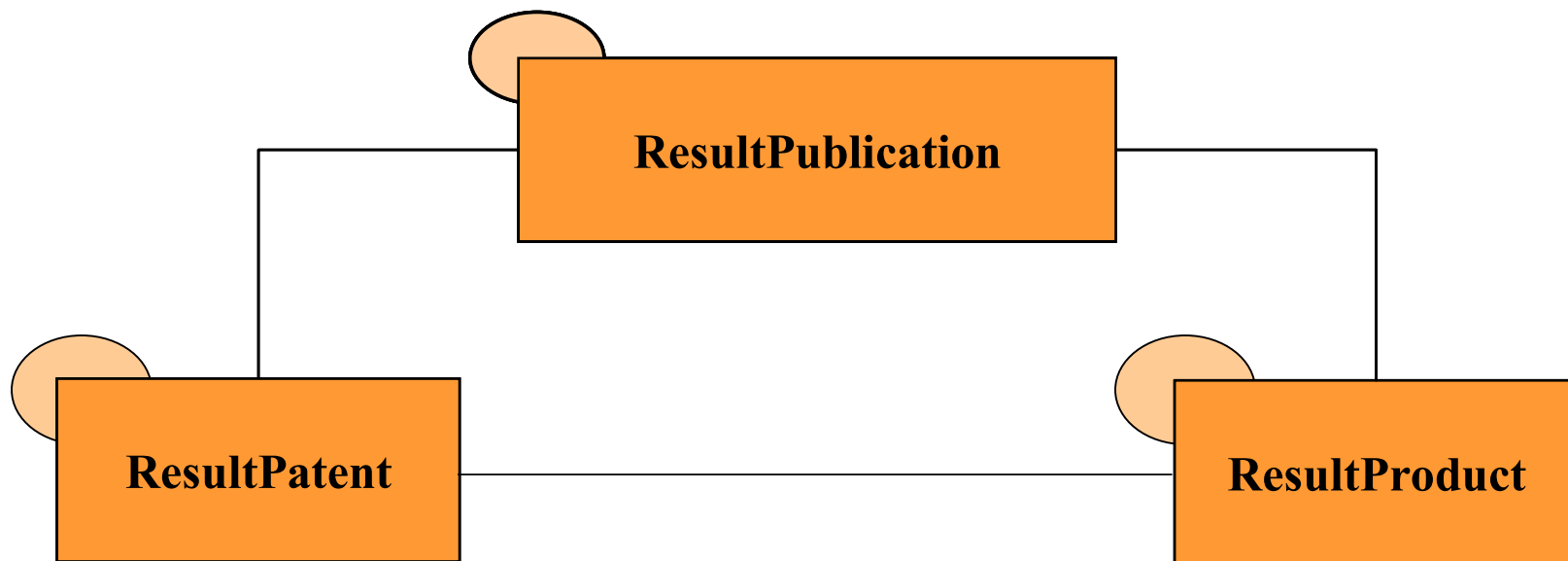


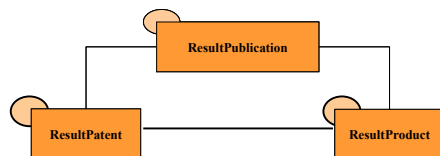


CERIF Base Entities

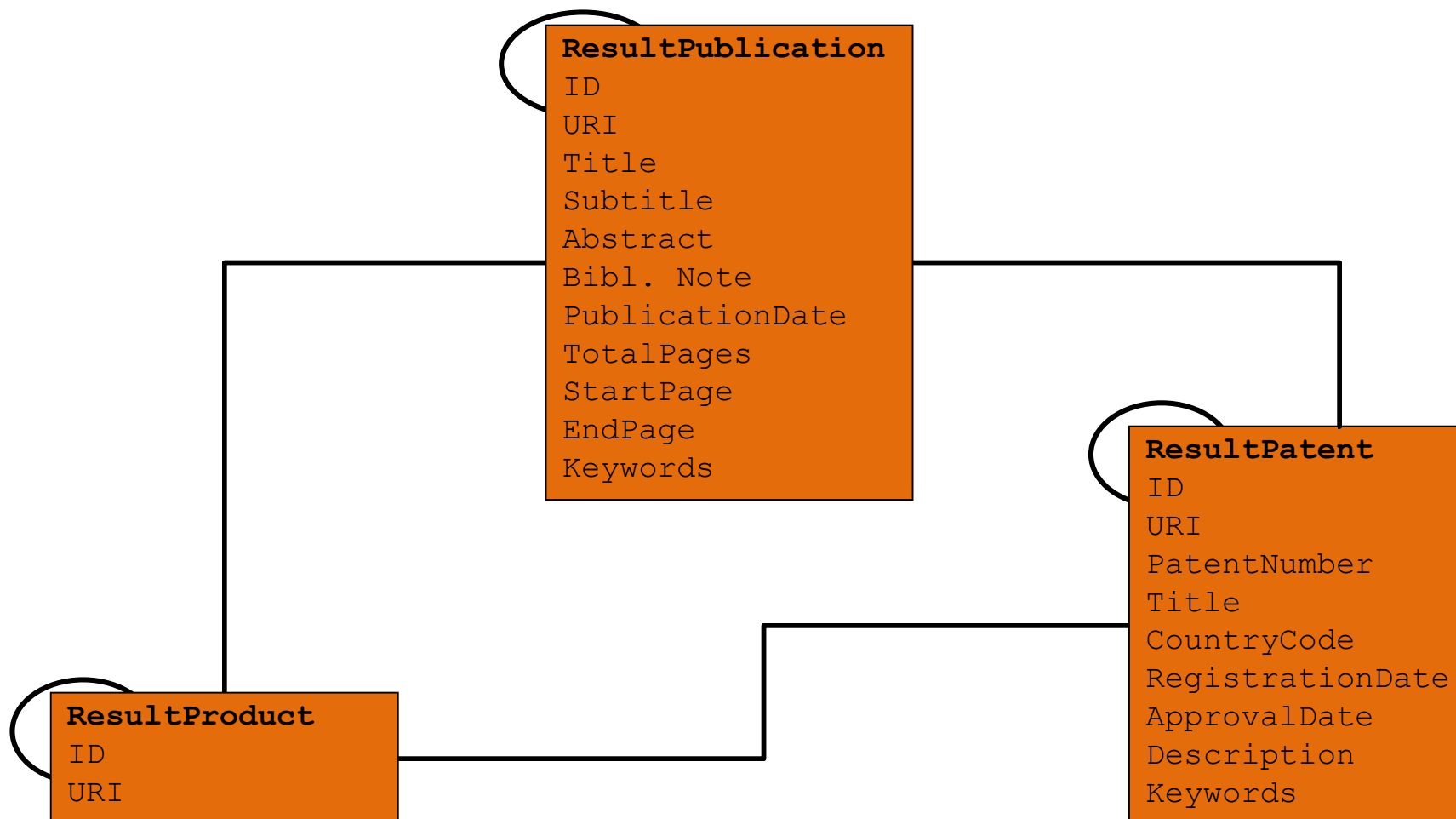


CERIF Result Entities

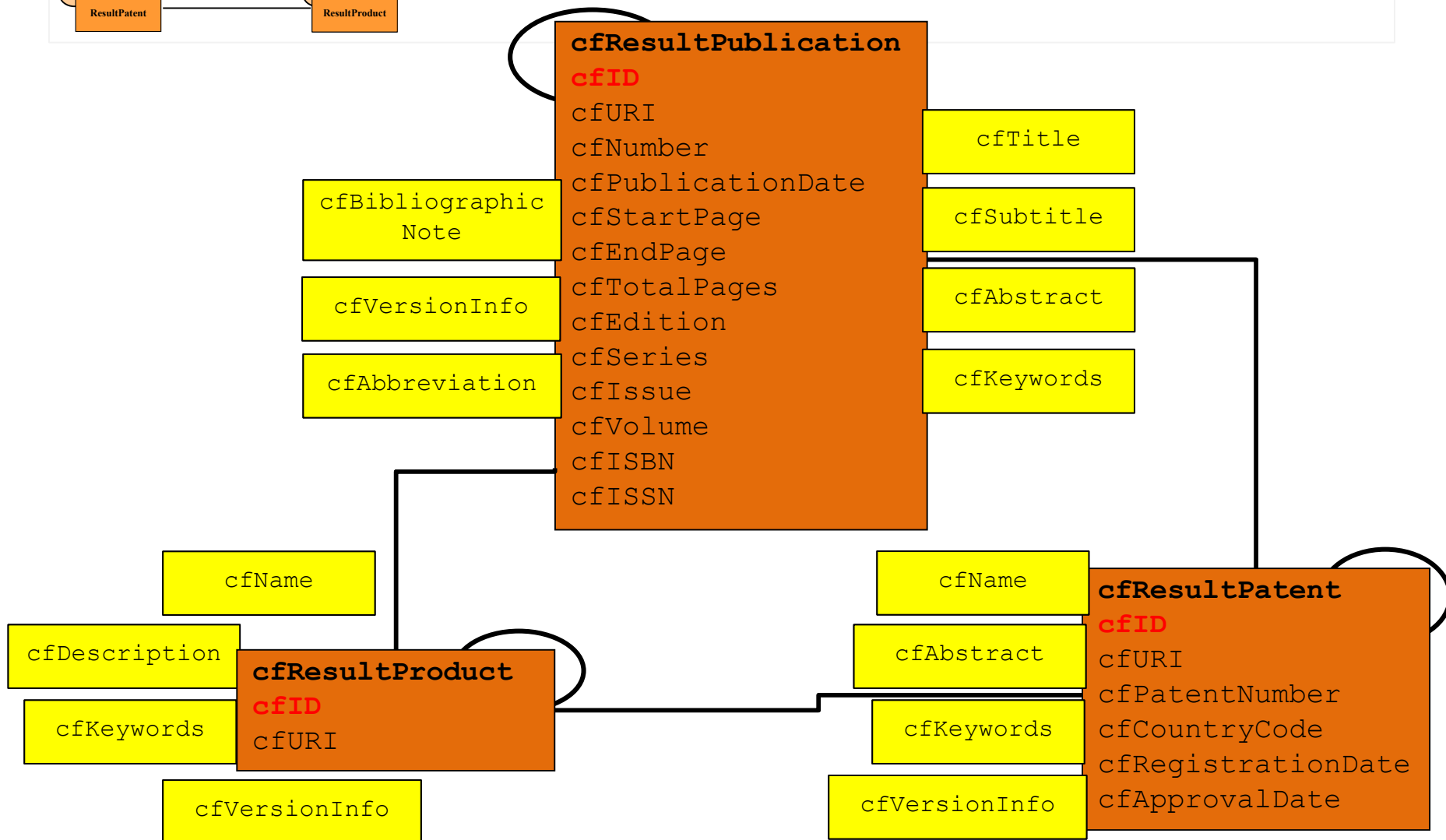
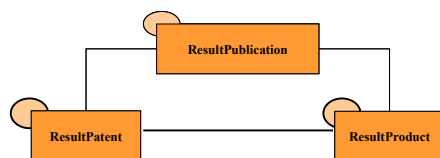




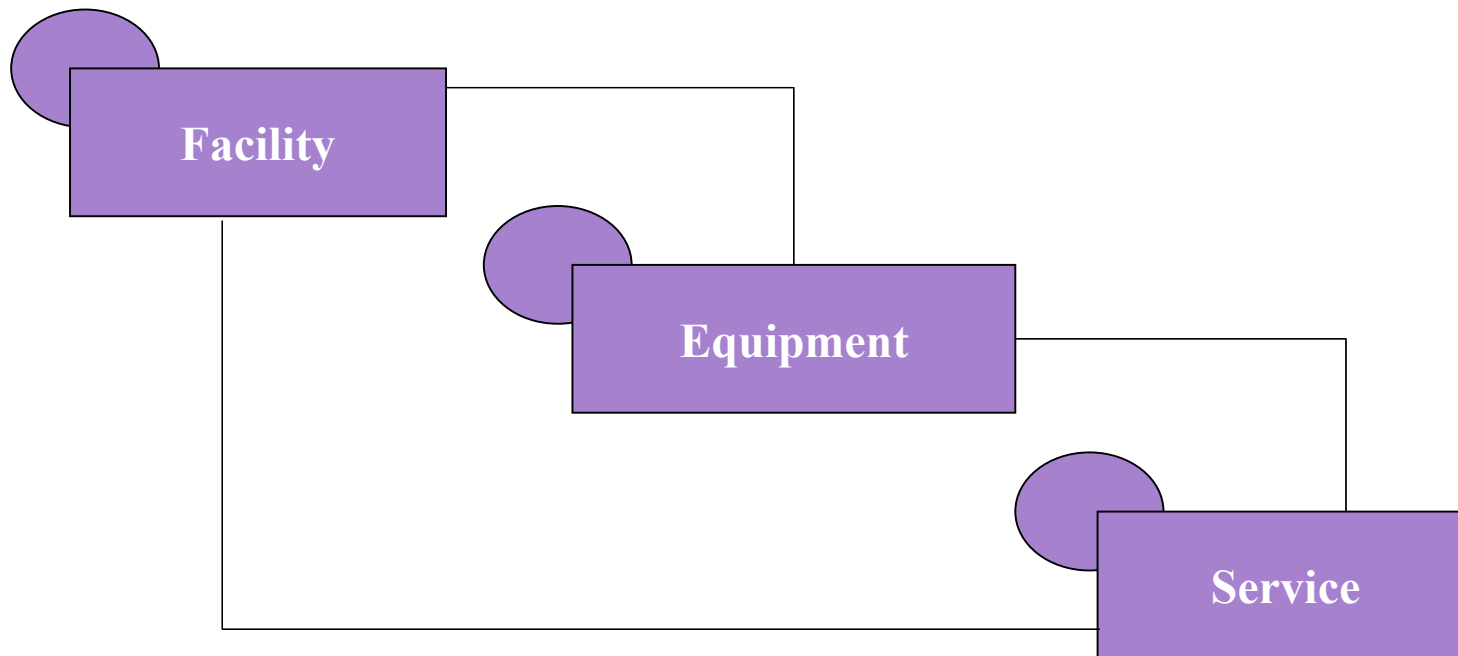
CERIF Result Entities

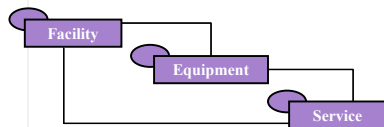


CERIF Result Entities

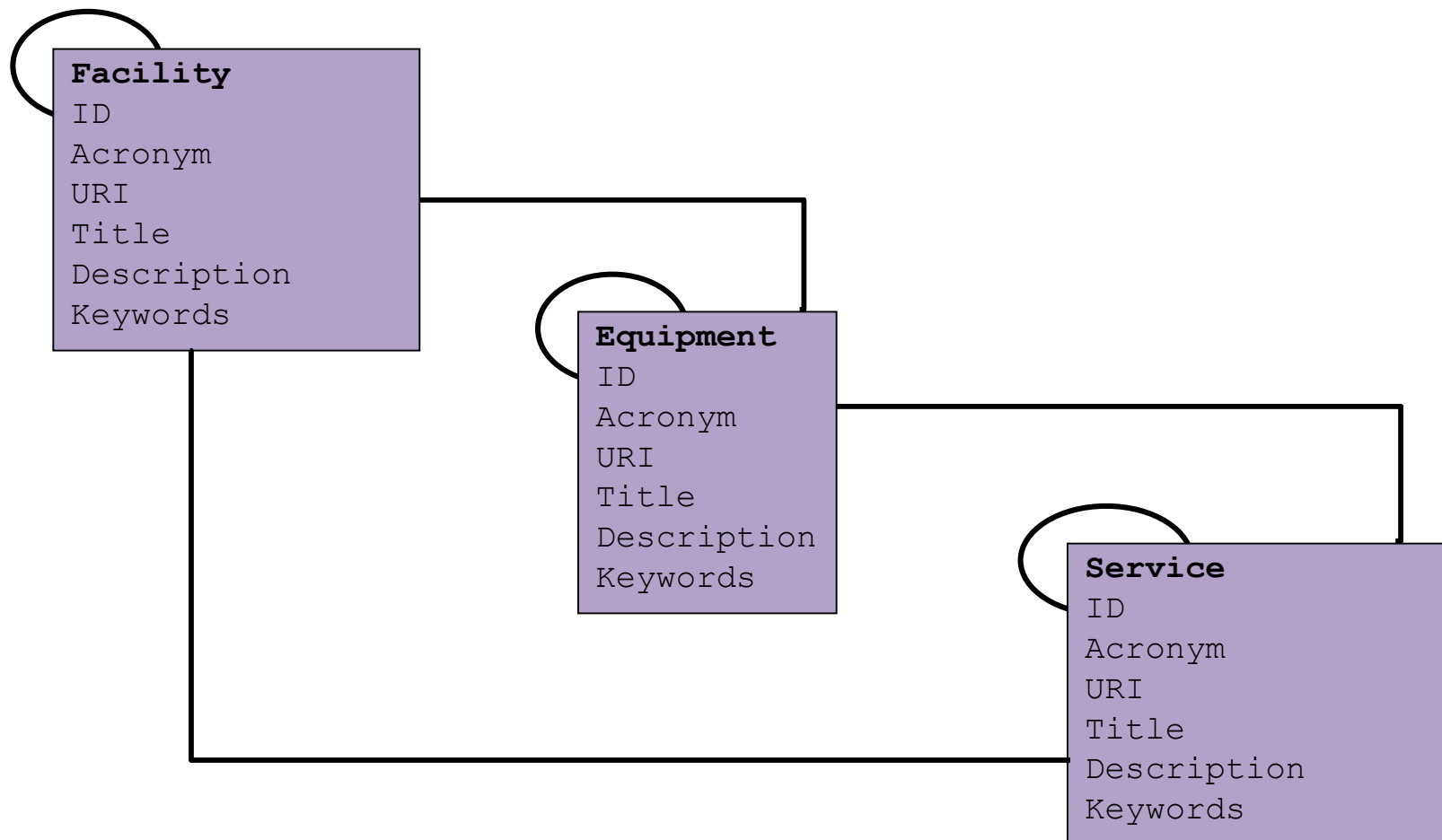


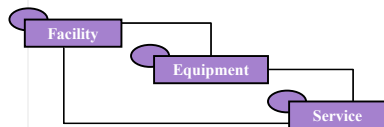
CERIF Infrastructure Entities



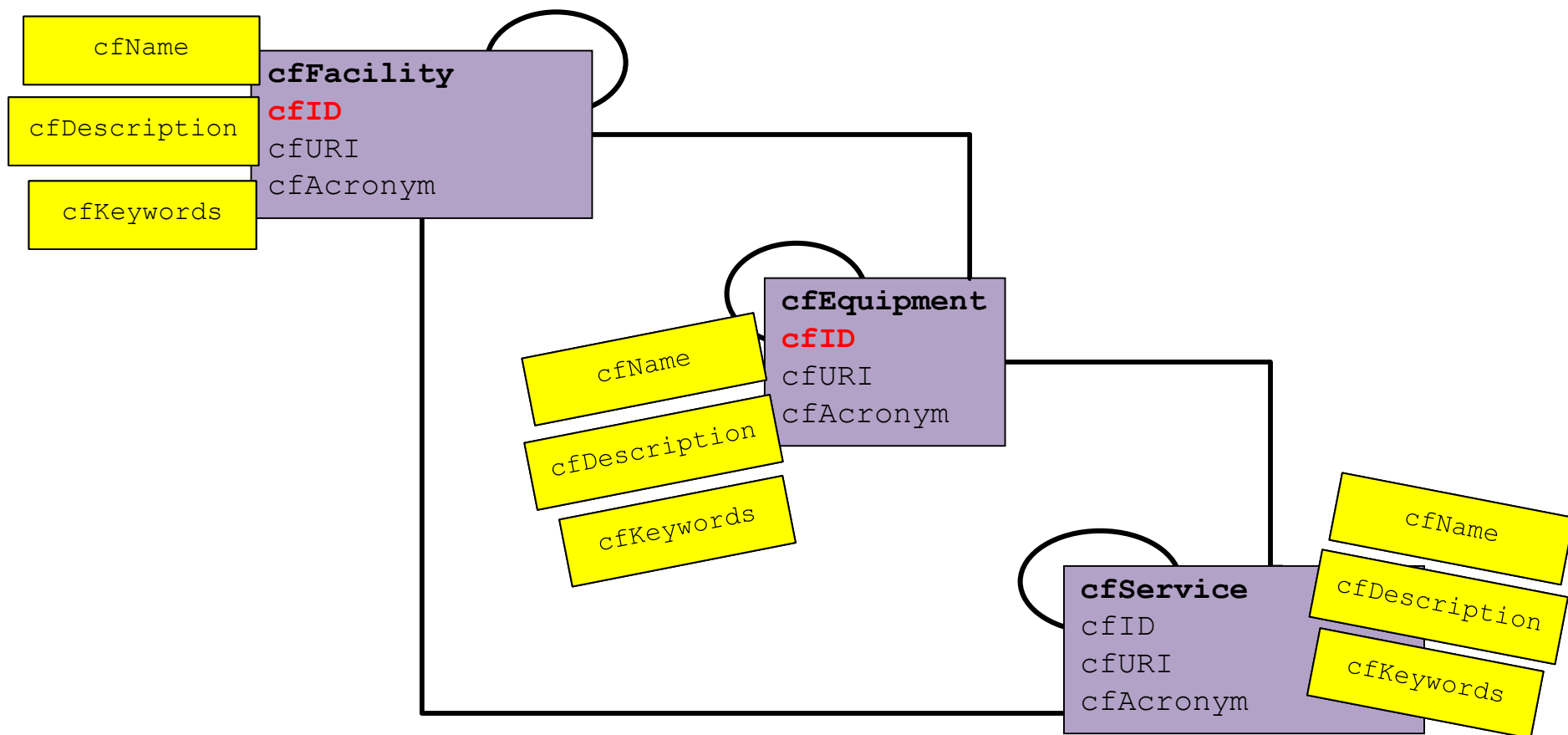


CERIF Infrastructure Entities

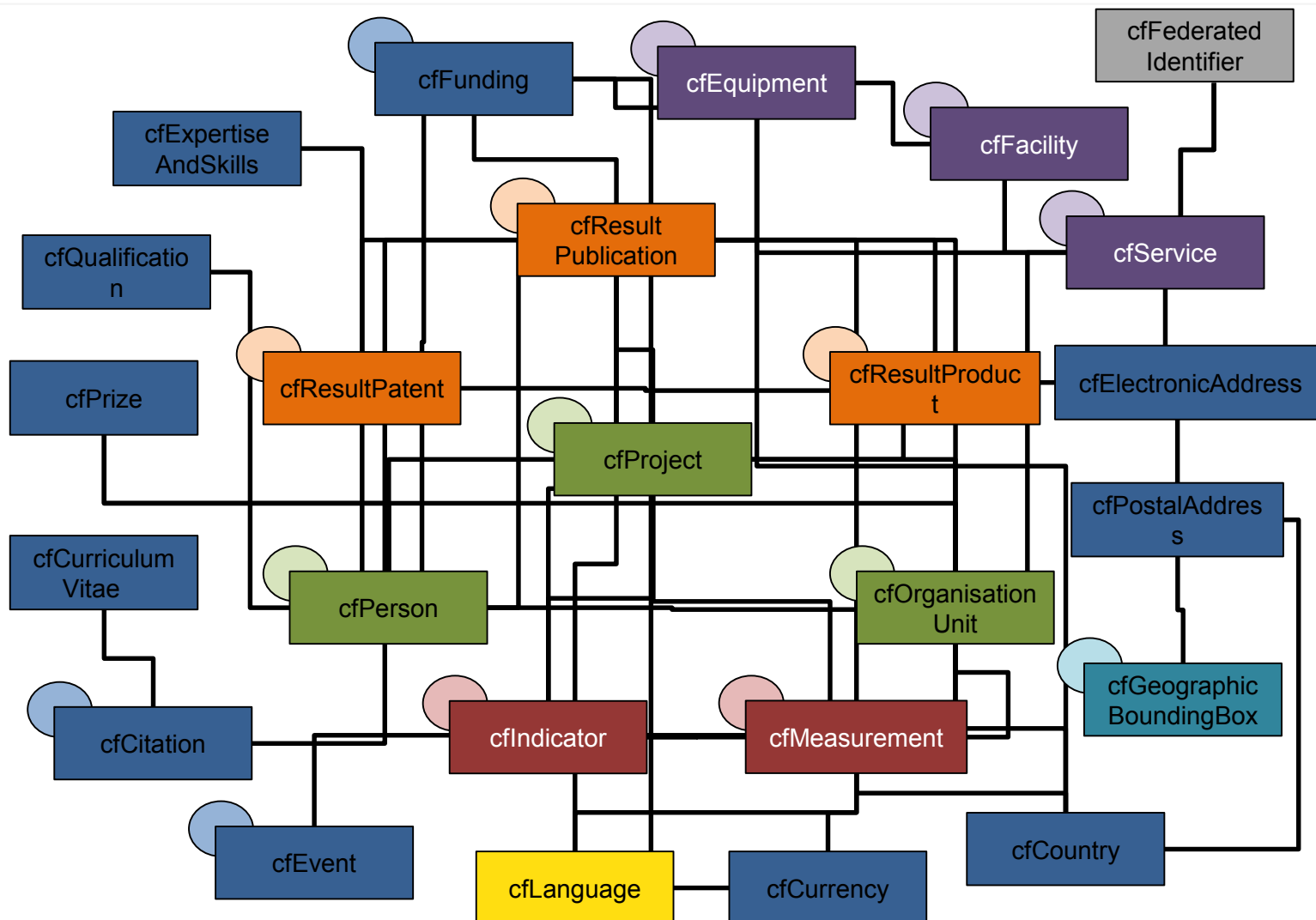


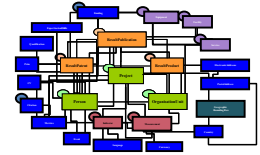


CERIF Infrastructure Entities

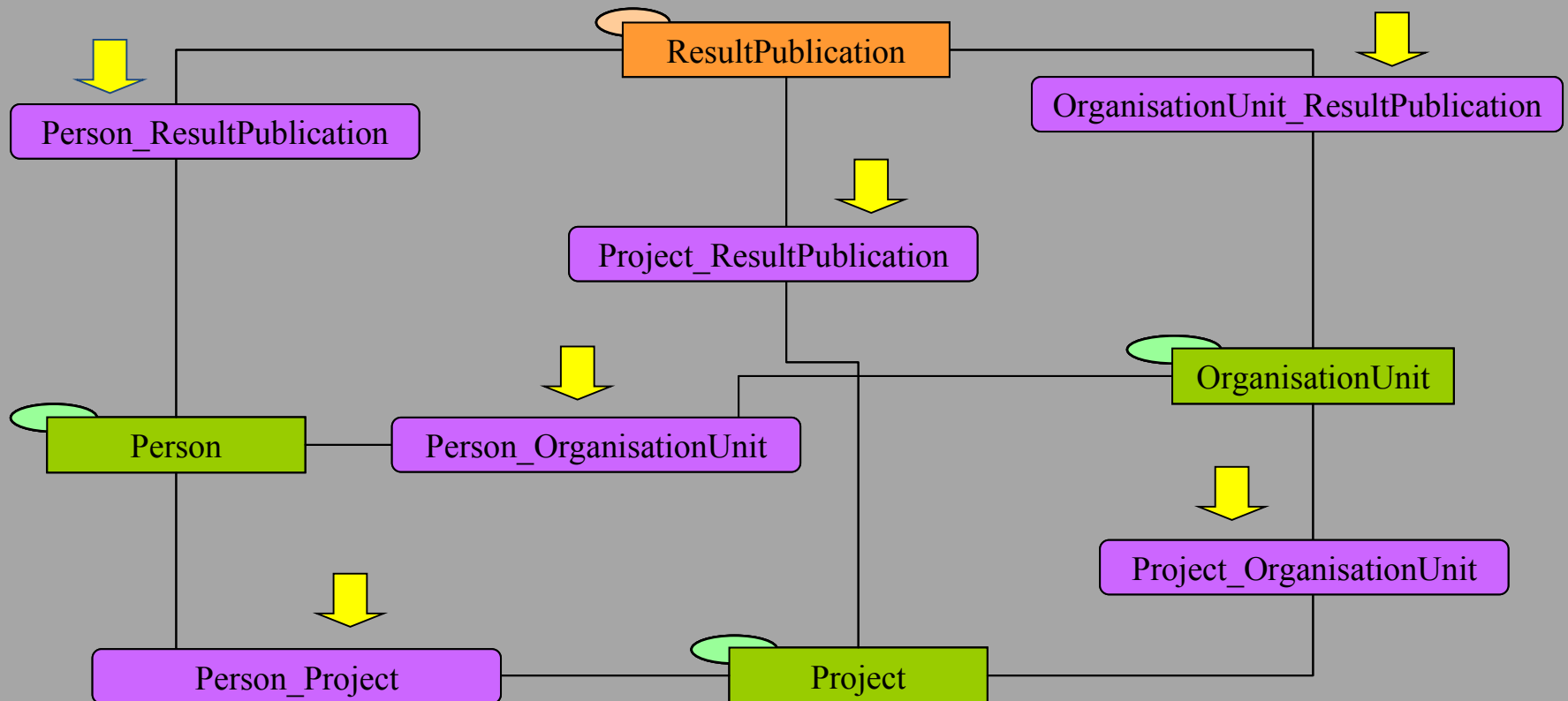


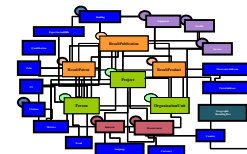
CERIF 1.6



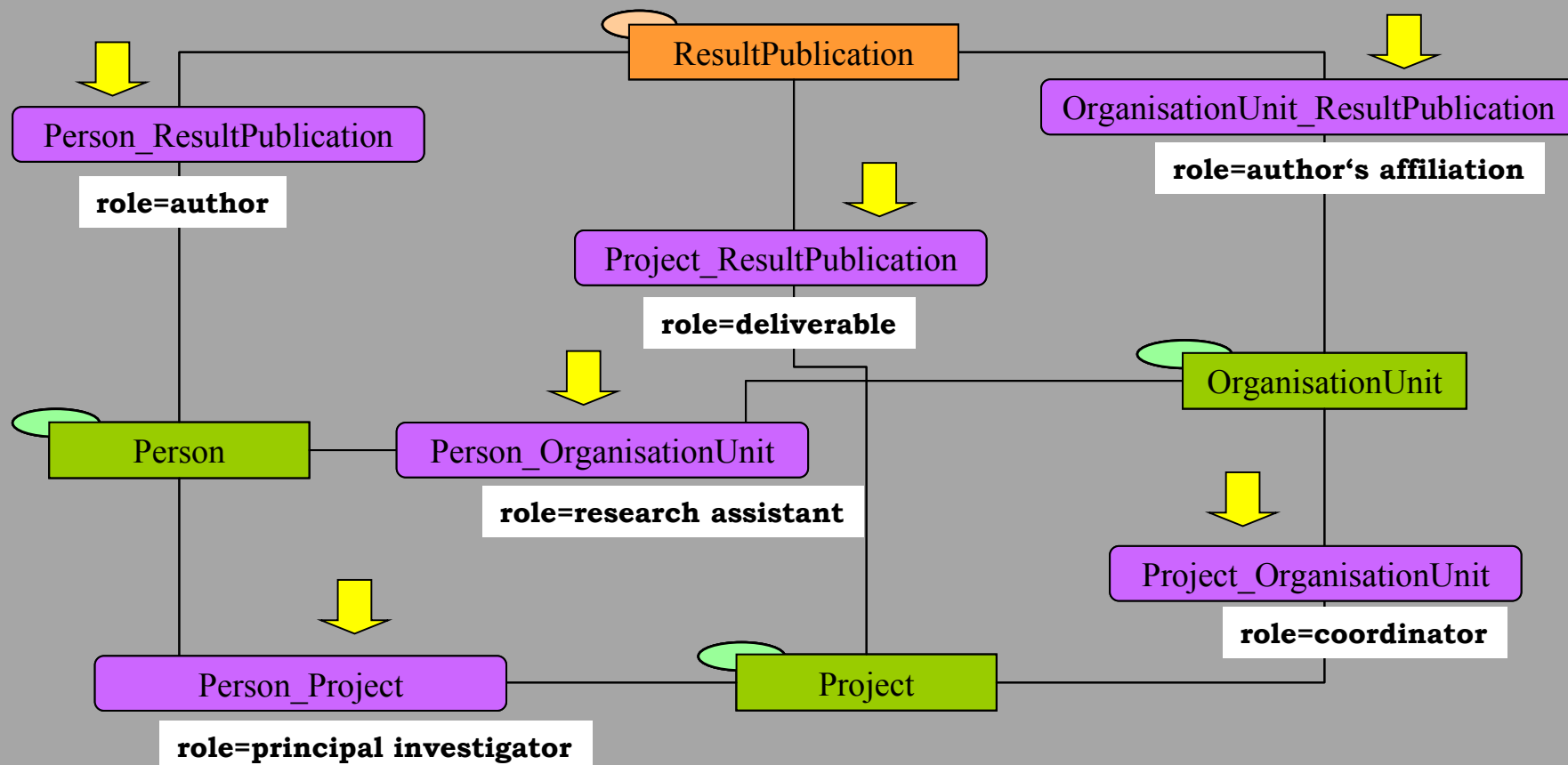


Some CERIF Link Entities



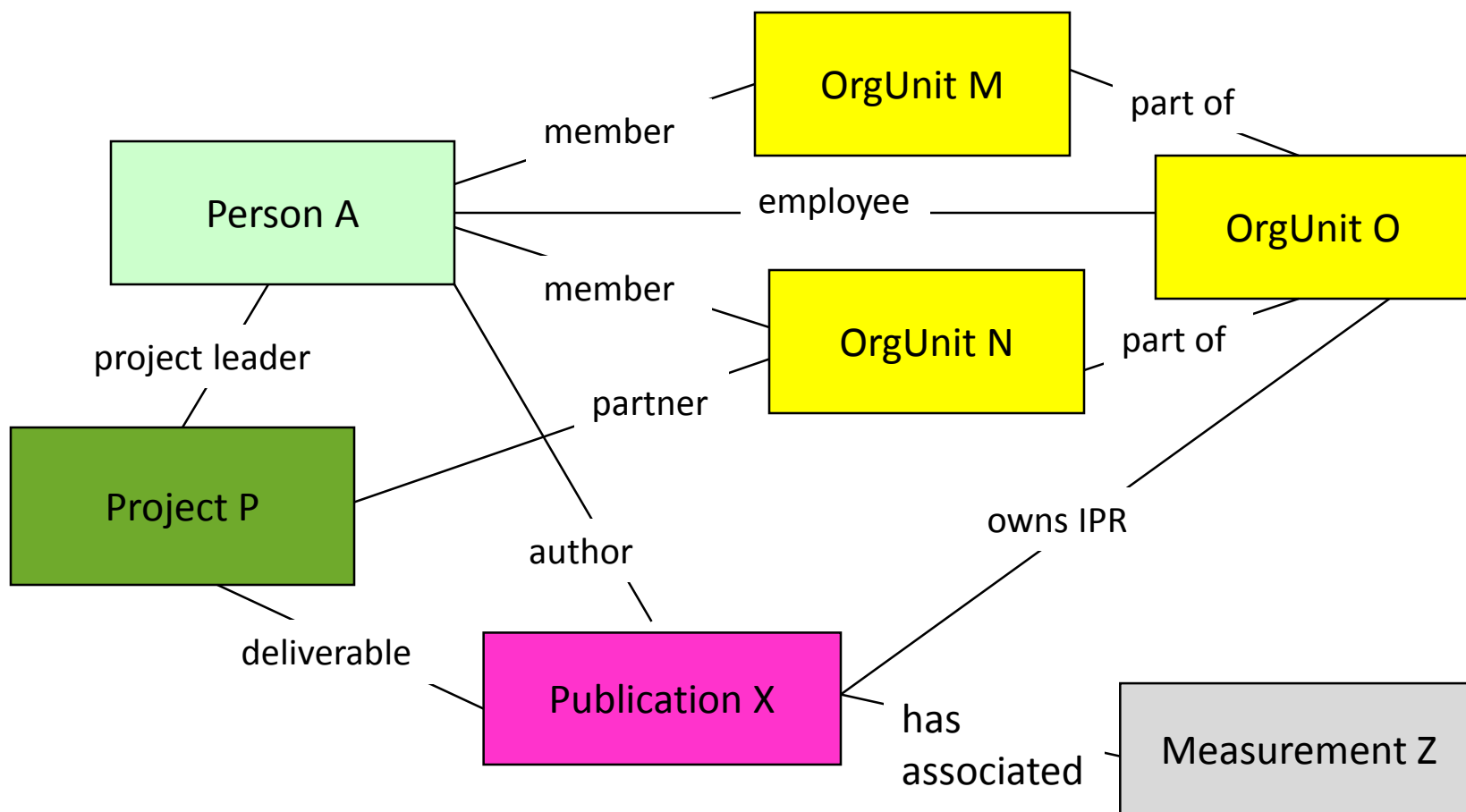


Some CERIF Link Entities



Result_Publication Instance Diagram

(slide by Keith Jeffery)

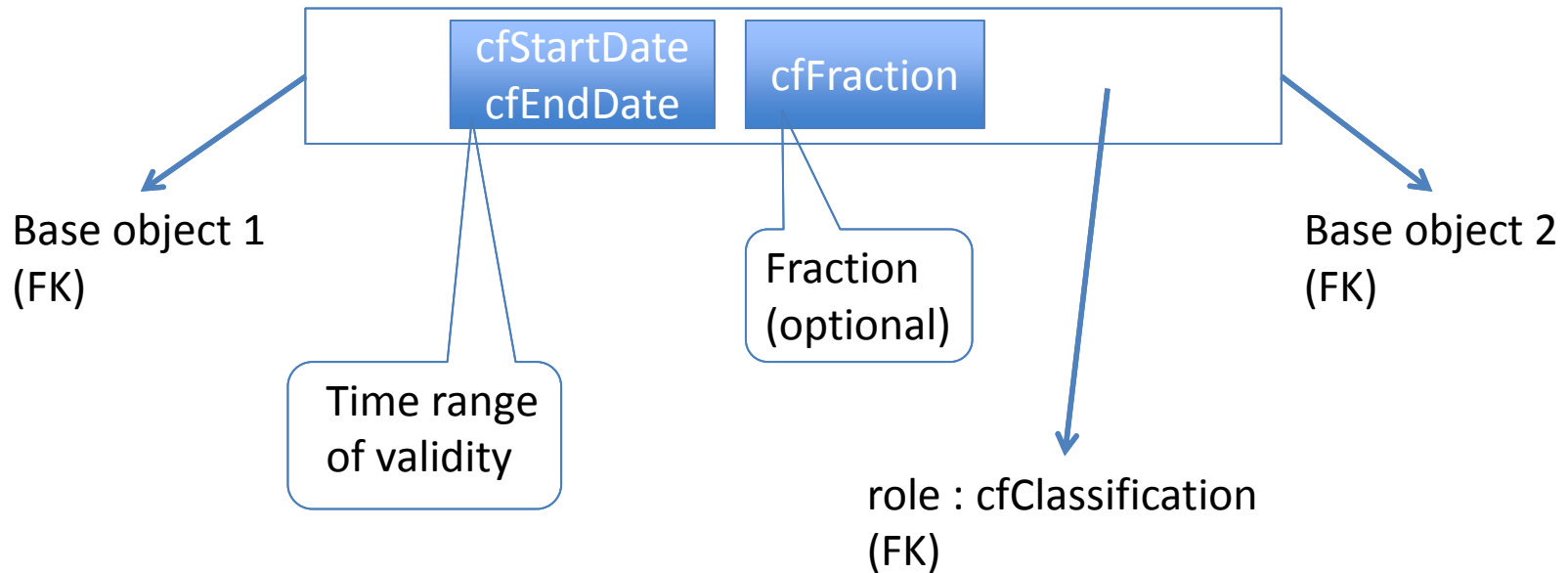


CERIF General Pattern

A typical CERIF entity:

- Identifier
 - internal
- Attributes
 - the basic ones
- Multi-lingual attributes
- Classifications
 - Type
 - Status
 - Subject area
- Links
 - to other entities
 - recursive

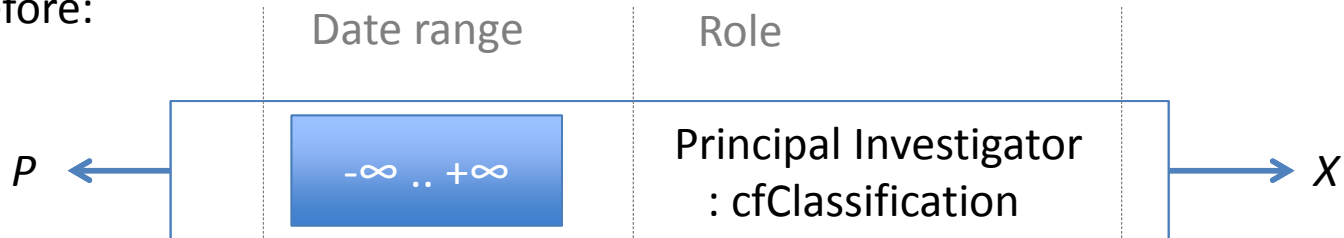
Generic Linking Entity Structure



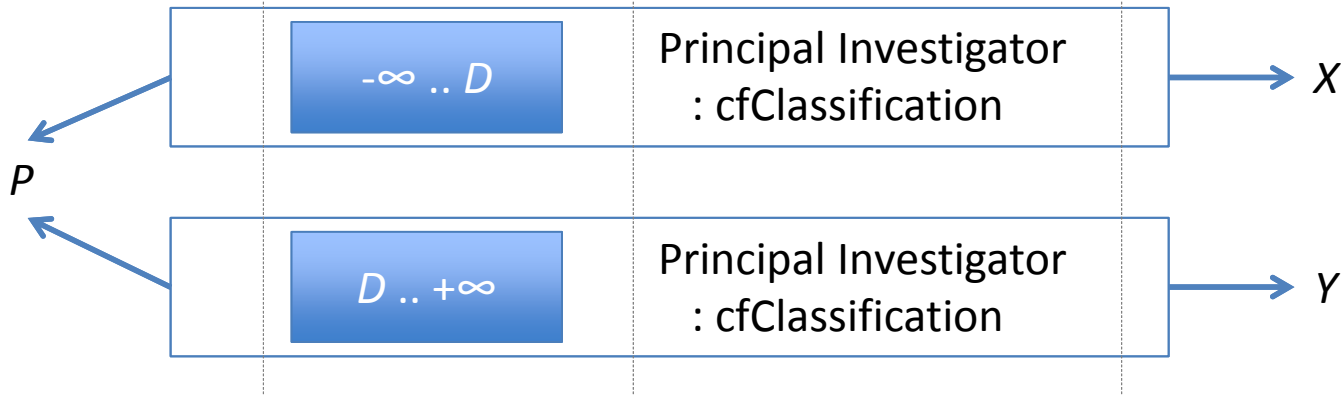
Recording Change in CERIF

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

Before:



After:



Some CERIF Link Entities

Binary
classifications:

- Role

cfPers_Pers			
cfPersId1	ID	NN	(PFK)
cfPersId2	ID	NN	(PFK)
cfClassId	ID	NN	(PFK)
cfClassSchemeld	ID	NN	(PFK)
cfFraction	Float		
cfStartDate	Timestamp(6)	NN	(PK)
cfEndDate	Timestamp(6)	NN	(PK)

cfProj_OrgUnit			
cfProjId	ID	NN	(PFK)
cfOrgUnitId	ID	NN	(PFK)
cfClassId	ID	NN	(PFK)
cfClassSchemeld	ID	NN	(PFK)
cfFraction	Float		
cfStartDate	Timestamp(6)	NN	(PK)
cfEndDate	Timestamp(6)	NN	(PK)

cfOrgUnit_Equip			
cfOrgUnitId	ID	NN	(PFK)
cfEquipId	ID	NN	(PFK)
cfClassId	ID	NN	(PFK)
cfClassSchemeld	ID	NN	(PFK)
cfFraction	Float		
cfStartDate	Timestamp(6)	NN	(PK)
cfEndDate	Timestamp(6)	NN	(PK)
cfCurrCode	Char(3)		(FK)
cfAvailability	Char(64)		
cfConditions	Char(255)		
cfPrice	Float		

cfOrgUnit_Class			
cfOrgUnitId	ID	NN	(PFK)
cfClassId	ID	NN	(PFK)
cfClassSchemeld	ID	NN	(PFK)
cfFraction	Float		
cfStartDate	Timestamp(6)	NN	(PK)
cfEndDate	Timestamp(6)	NN	(PK)

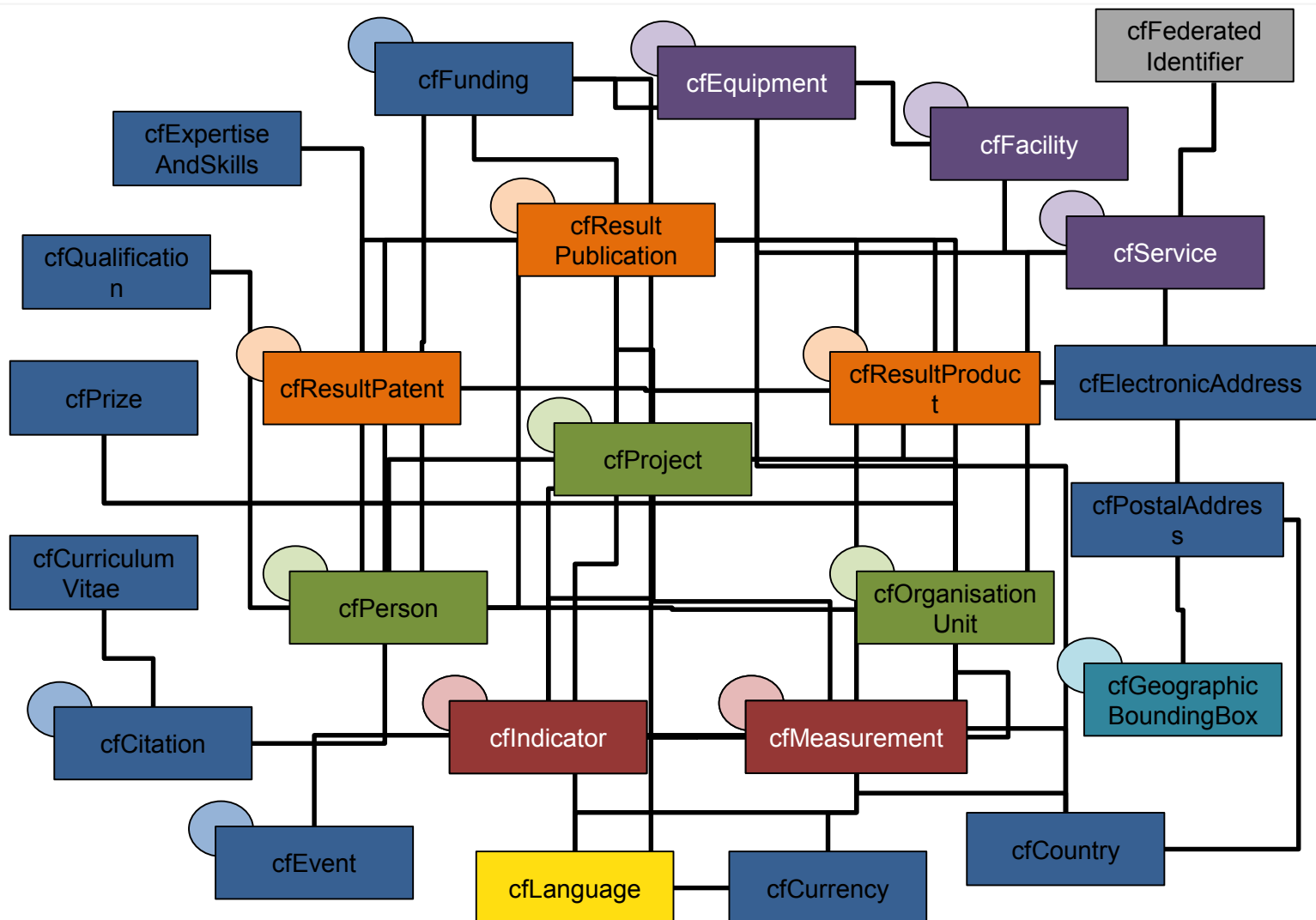
cfClass_Class			
cfClassId1	ID	NN	(PFK)
cfClassId2	ID	NN	(PFK)
cfClassSchemeld1	ID	NN	(PFK)
cfClassSchemeld2	ID	NN	(PFK)
cfClassId	ID	NN	(PFK)
cfClassSchemeld	ID	NN	(PFK)
cfFraction	Float		
cfStartDate	Timestamp(6)	NN	(PK)
cfEndDate	Timestamp(6)	NN	(PK)

cfProj_Pers			
cfProjId	ID	NN	(PFK)
cfPersId	ID	NN	(PFK)
cfClassId	ID	NN	(PFK)
cfClassSchemeld	ID	NN	(PFK)
cfFraction	Float		
cfStartDate	Timestamp(6)	NN	(PK)
cfEndDate	Timestamp(6)	NN	(PK)

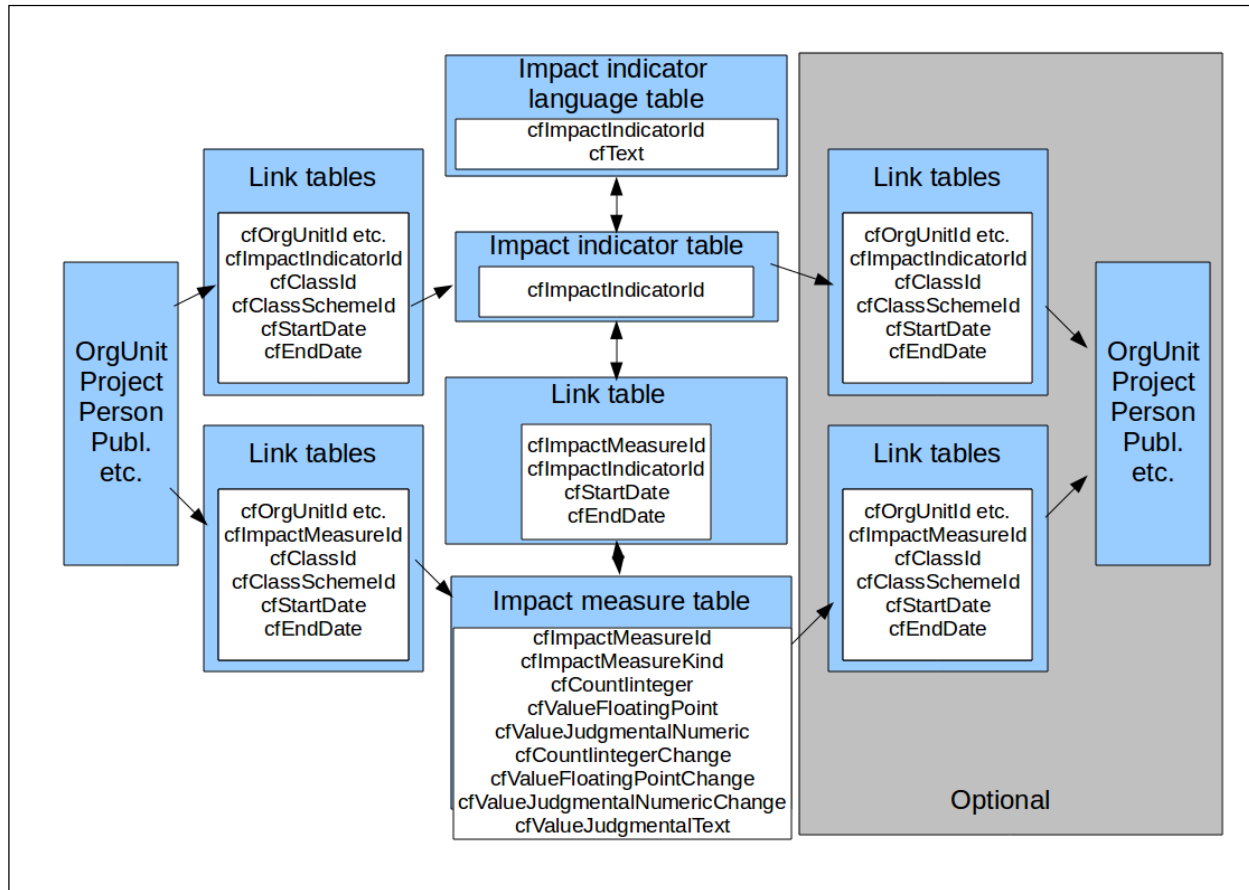
Unary
classification:

- Type
- Status
- Subject
area

CERIF 1.6



Measuring Impact in CERIF (MICE)



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

CERIF Measurement & Indicator



Is an Aggregation Entity

cfMeasureIdentifier
cfCountInteger
cfCountIntegerChange
cfValueFloatingPoint
cfCountFloatingPointChange
cfValueJudgementalNumeric
cfValueJudgementalNumericChange
ge
cfValueJudgementalText
cfValueJudgementalTextChange
cfURI

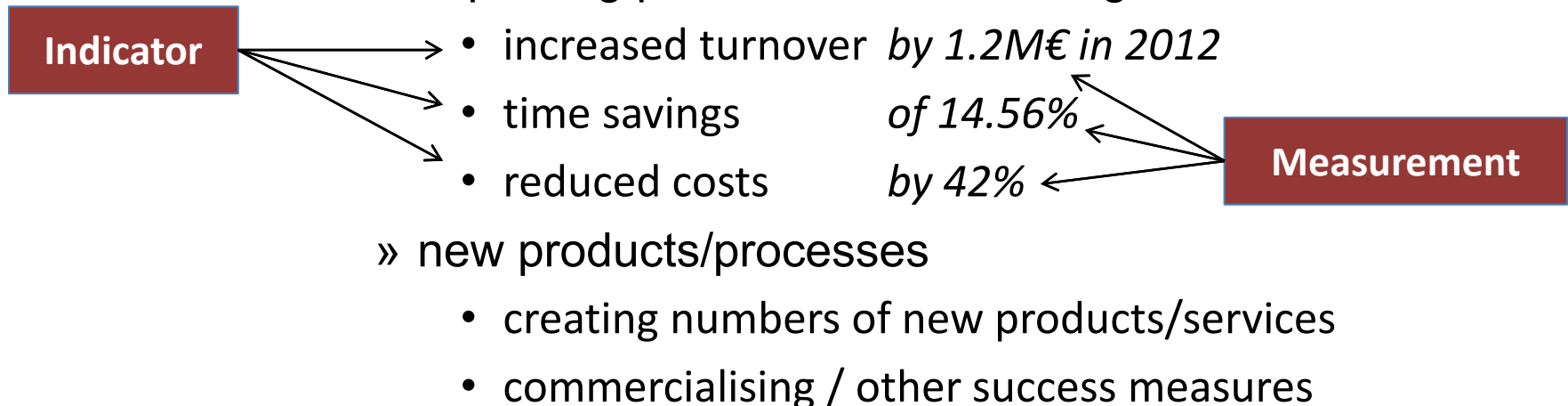
Measurement & Indicator (some examples)

– economic and commercial

- economic

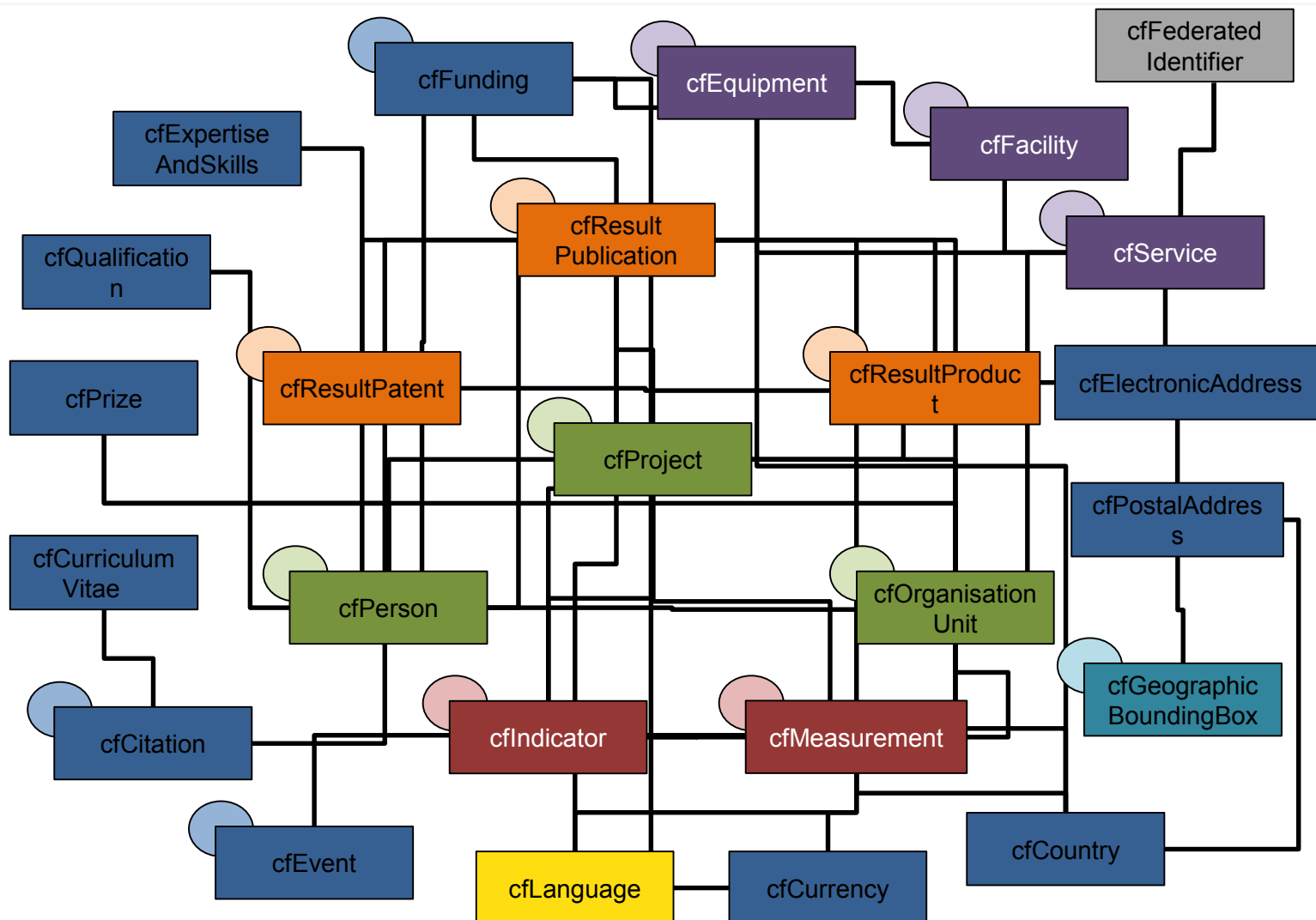
- impact on business

- » improving performance of existing businesses



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

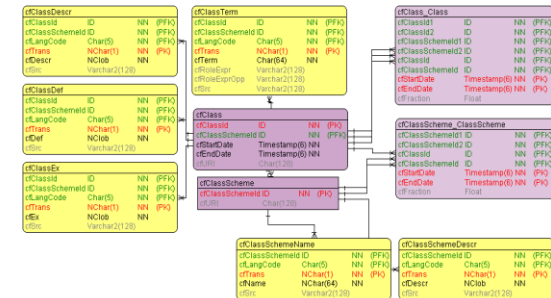
Open / Extensible in all directions

- New Schemas
- New Concepts / Terms
- New Relationships

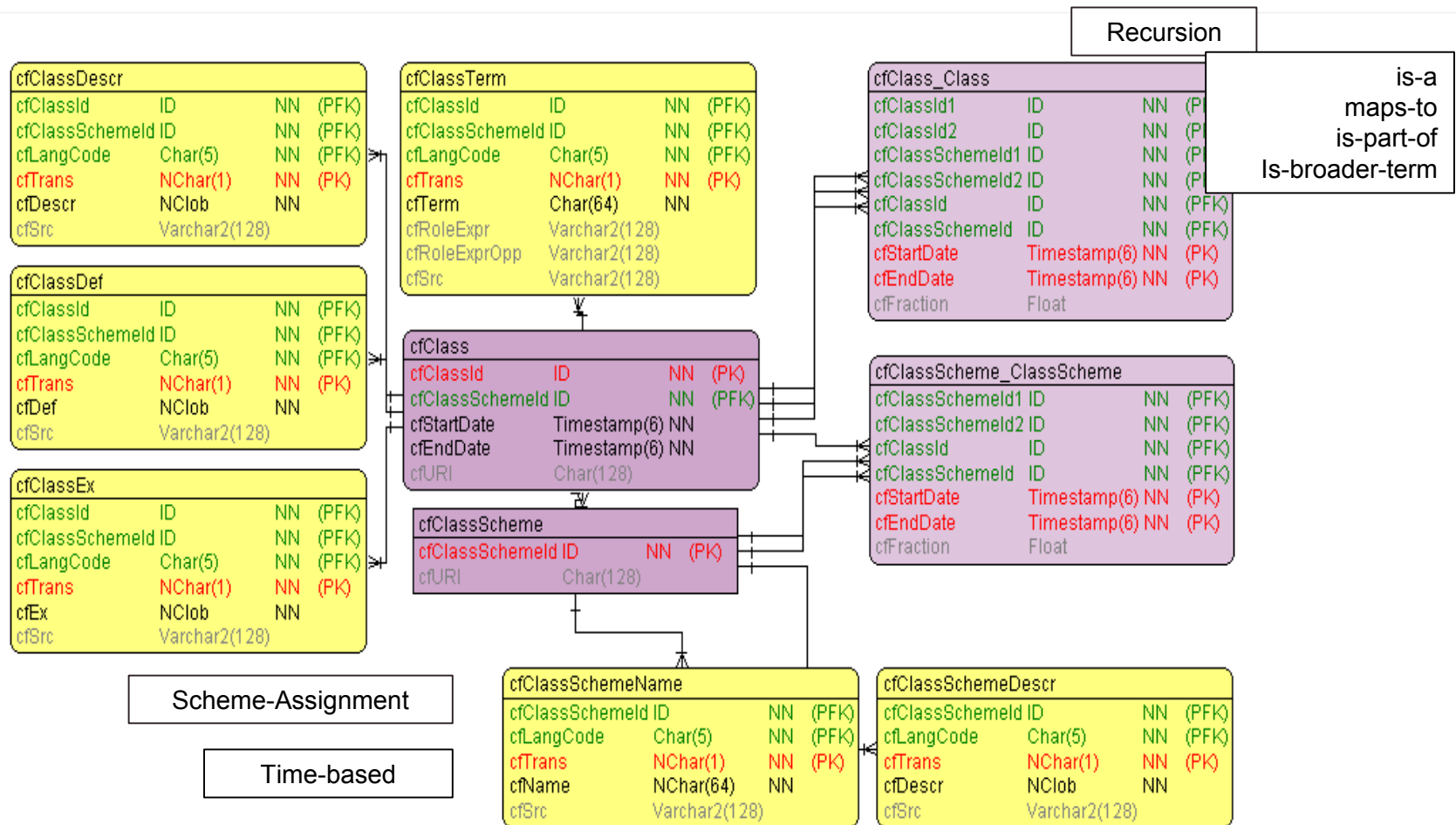
Enables to manage

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

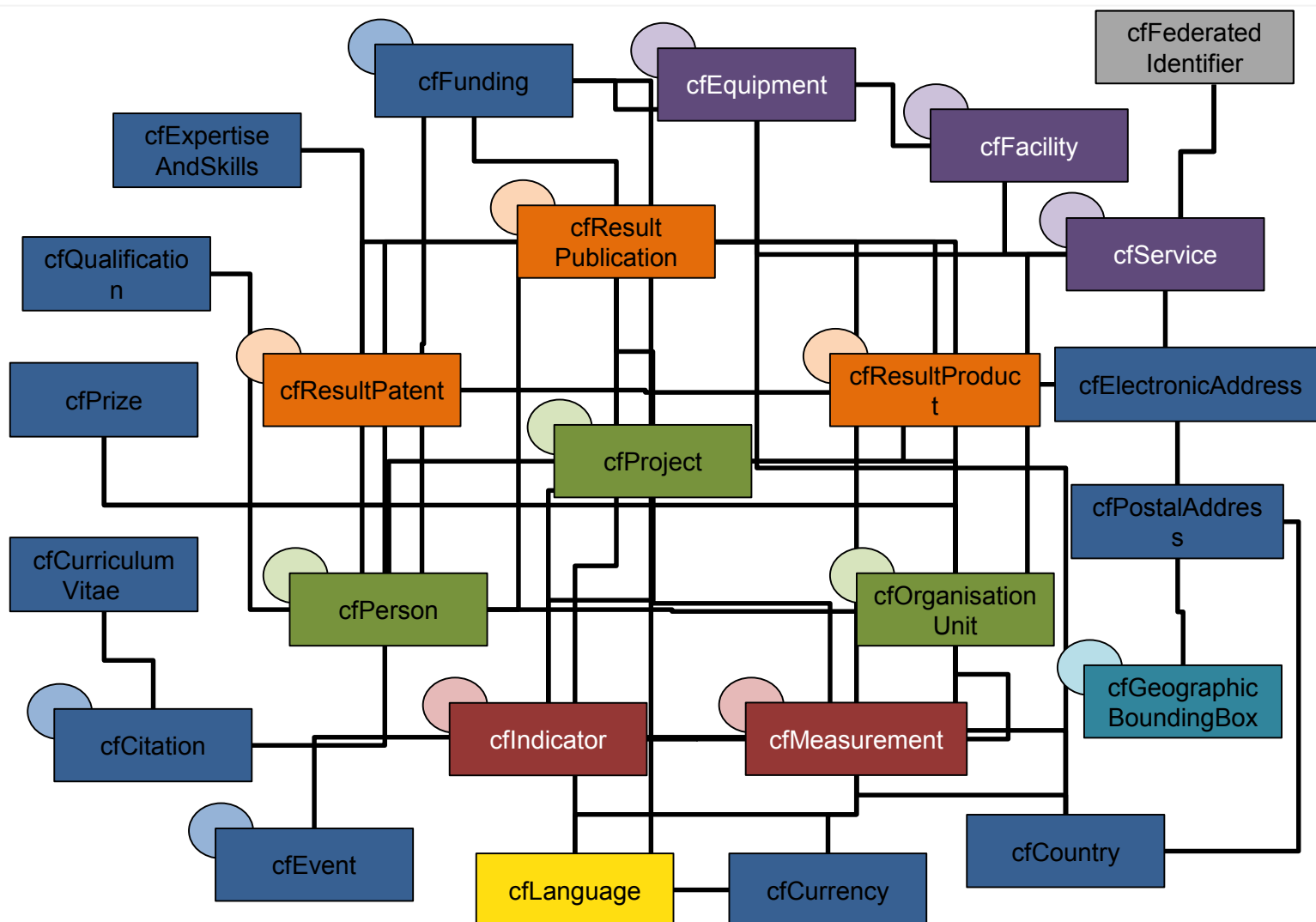
Allows for Mappings between Schemes



CERIF Semantic Layer (Declared Semantics)



CERIF 1.6



CERIF Federated Identifiers

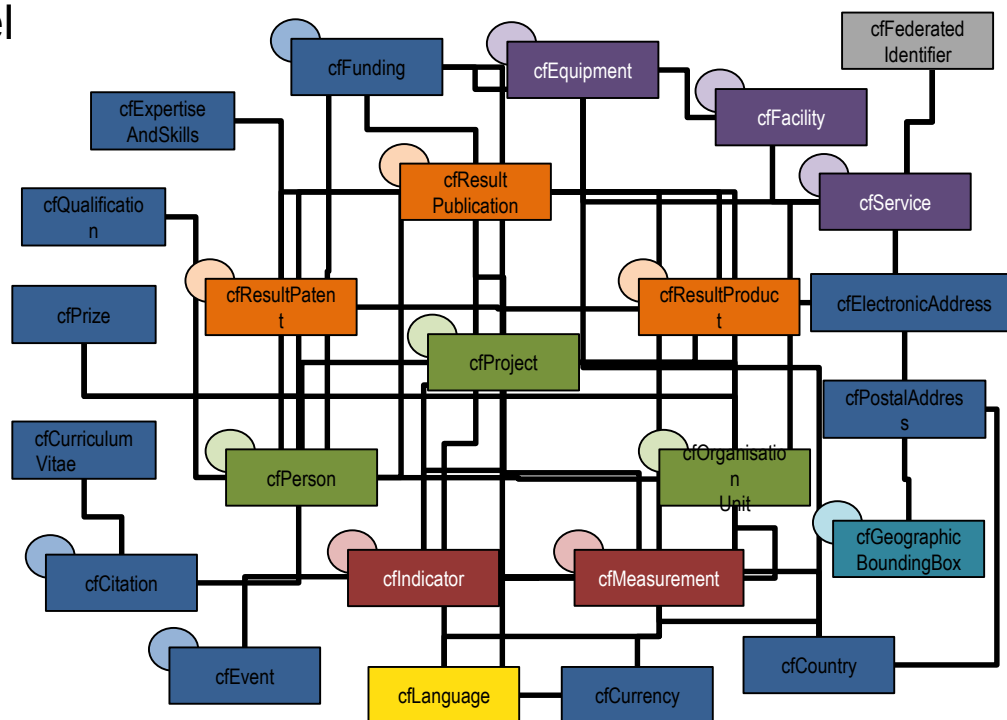
- ResultPublication
 - ISBN
 - ISSN
 - DOI
 - WoS Accession Number
 - Scopus EID
 - PubMed Central ID
- Person
 - Social Security Number
 - Staff Id in HR system
 - Author identifier
 - ORCID, IdRef, DAI, ResearcherID, ScopusID
- Project/Grant
 - Funder's reference number
 - Organisation's reference 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 1.6 Interchange Format

- For point-to-point interchange
- XML namespace
- XML Schema
 - Based on the ER model



CERIF 1.6 XML Interchange Format

```
<CERIF xmlns="urn:xmlns.org:eurocris:cerif-1.6-2">
  <cfProj>
    <cfProjId>internal-project-identifier</cfProjId>
    <cfAcro>ACRO</cfAcro>
    <cfURI>http://www.project-url.ac.uk/acro.html</cfURI>
    <cfTitle cfLangCode="en" cfTrans="o">The title of the project</cfTitle>
    <cfAbstr cfLangCode="en" cfTrans="o">The goals of the project</cfAbstr>
    <cfProj_Class>
      <cfClassId>infrastructure-project-uuid</cfClassId>
      <cfClassSchemeld>-project-types-scheme-uuid</cfClassSchemeld>
    </cfProj_Class>
    <cfFedId>
      <cfFedId>PROJECT NUMBER</cfFedId>
      <cfClassId>project-number-uuid</cfClassId>
      <cfClassSchemeld>-federated-identifier-type-uuid</cfClassSchemeld>
    </cfFedId>
    <cfProj_OrgUnit>
      <cfOrgUnitId>orgunit-1-identifier</cfOrgUnitId>
      <cfClassId>coordinator-uuid</cfClassId>
      <cfClassSchemeld>orgunit-project-roles-scheme-uuid</cfClassSchemeld>
      <cfStartDate>from-datetime</cfStartDate>
      <cfEndDate>to-datetime</cfEndDate>
    </cfProj_OrgUnit>
  </cfProj>
</CERIF>
```

CERIF 1.6 XML Interchange Format

XML Schema-based

Separate namespace

`urn:xmlns:org:eurocris:cerif-1.6-2` for CERIF 1.6

Used in:

OpenAIRE Guidelines for CRIS managers 1.0

CERIF API specification (-> Arch TG)

euroCRIS CERIF CRIS Reference Implementation

Ongoing work: CERIF XML Update

- More readable XML
- Better connection with the research domain
- Reduce fragmentation

→ presentation

CERIF development

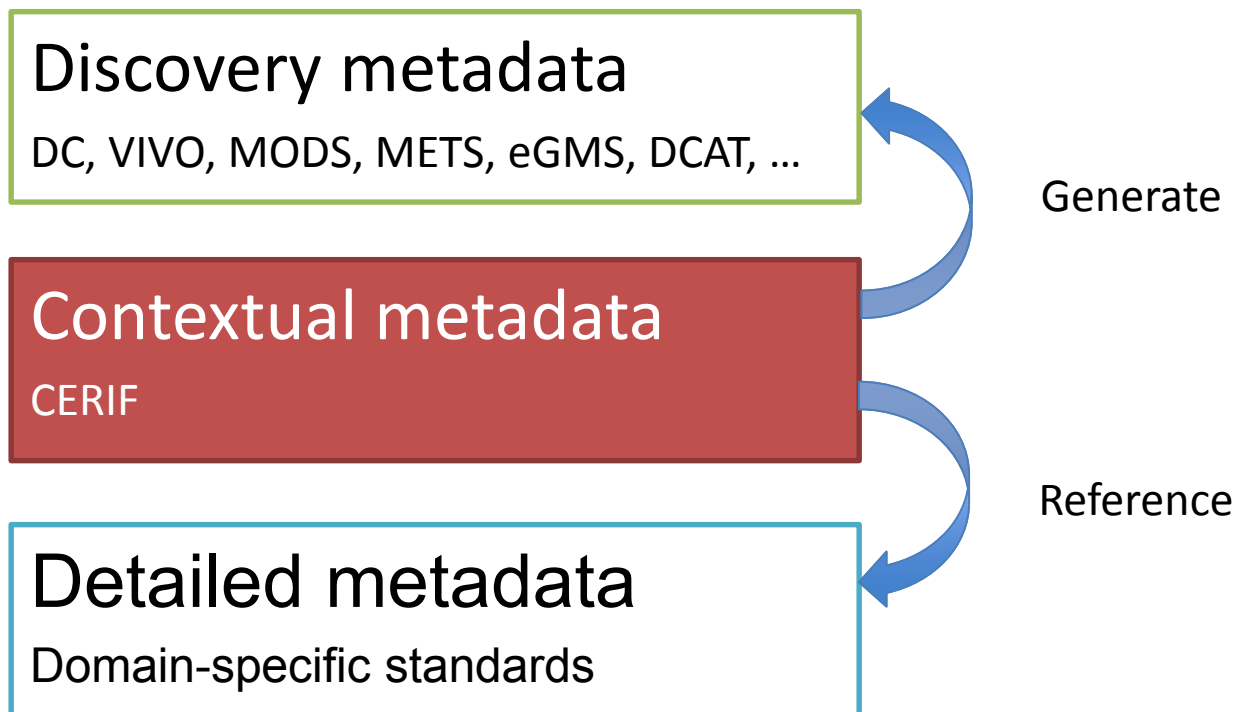
By the CERIF Task Group of euroCRIS

Adopting open-source software projects
tools & best practise

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



The CERIF Evolution

Similar Ideas
UN/UNESCO
OECD
CODATA

EU
Working Group
on Research
Databases
Workshop

CERIF 91

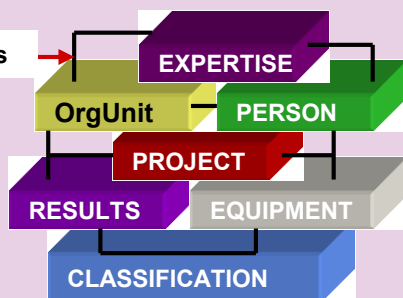
PROJECT

Acronym: ERGO
Participant:
Keith Jeffery, Anne Asser
son, many more
Organisations:
Rutherford Appleton, Uni-
versity of Bergen,

- Networking of DBs
- Exchange of Records
- EC Recommendation to Member States

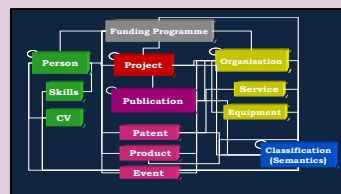
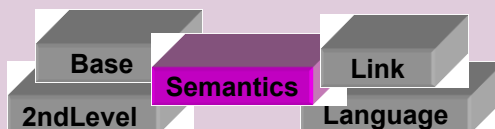
CERIF 2000 Model

Roles

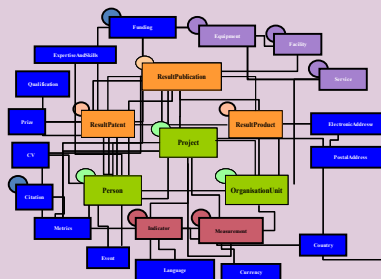
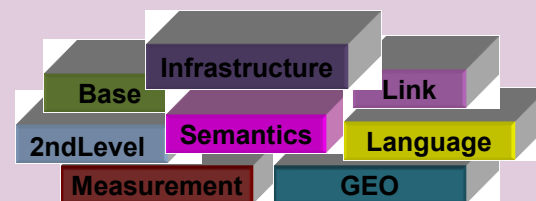


- Data Model
- Multilinguality
- Controlled Vocabulary
- Roles / Types
- User-driven
- EC Recommendation to Member States

CERIF 2006 / 2008 Model



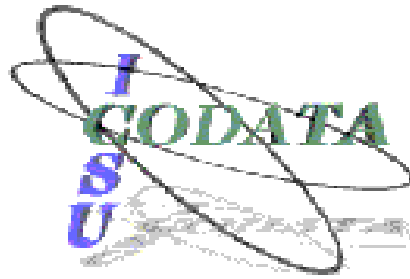
- Data Model
- Model Normalization
- Robust/Consistent Structure
- Extensible Structure
- Semantic Layer
- XML Exchange Specification
- Elaboration on Publication
- CERIF Core Semantics (2008 1.2)



- Data Model
- Infrastructure
- Facility, Equipment, Service
- Measurement & Indicator
- Entities and Link Tables
- Geographic Bounding Box
- CERIF 1.3 Vocabulary
- UUIDs
- Terms
- Schemes
- CERIF 1.4 new XML format
- CERIF 1.5 Federated Identifiers
- CERIF 1.6 Data model

+ Linked
Data

P
r
o
f
i
l
e
s



*International Council for Science;
Commission on Data Access*



*European Association of Research
Managers and Administrators*

ORCID



www.eurocris.org



← Alliance for Permanent Access →

