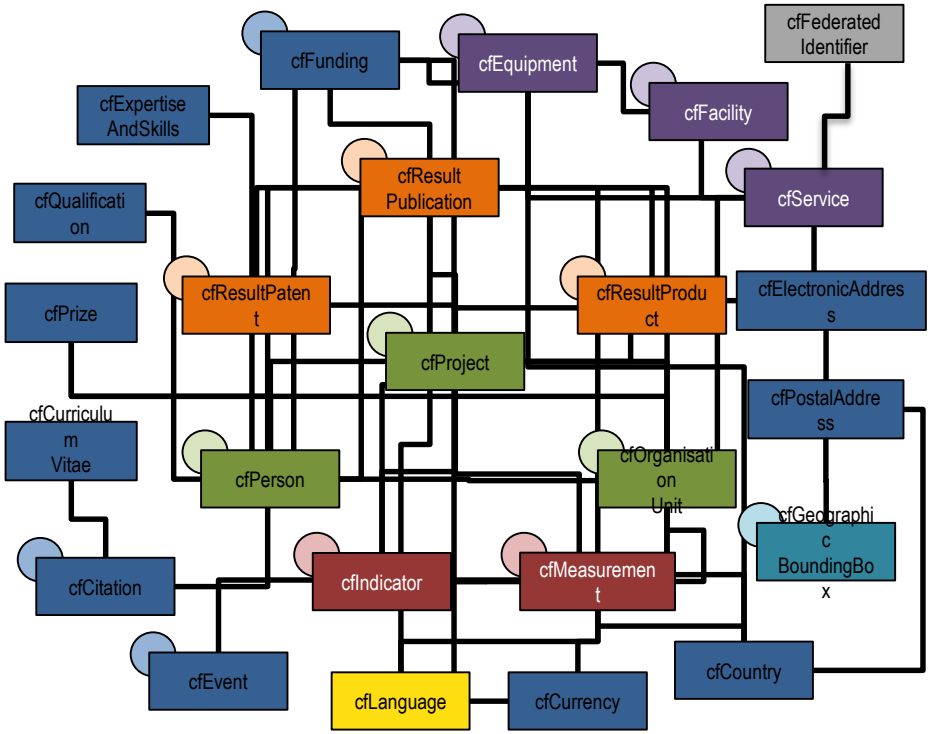
CERIF Tutorial: Part 1



Jan Dvořák

October 2nd, 2018

I encuentro de gestores de información en CTI - Perú
ESAN University, Lima, Perú

Jan Dvořák

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

euroCRIS
- CERIF TG Leader since 2013

Charles University, Faculty of Arts, Institute of Information Studies & Librarianship
- Researcher & Lecturer

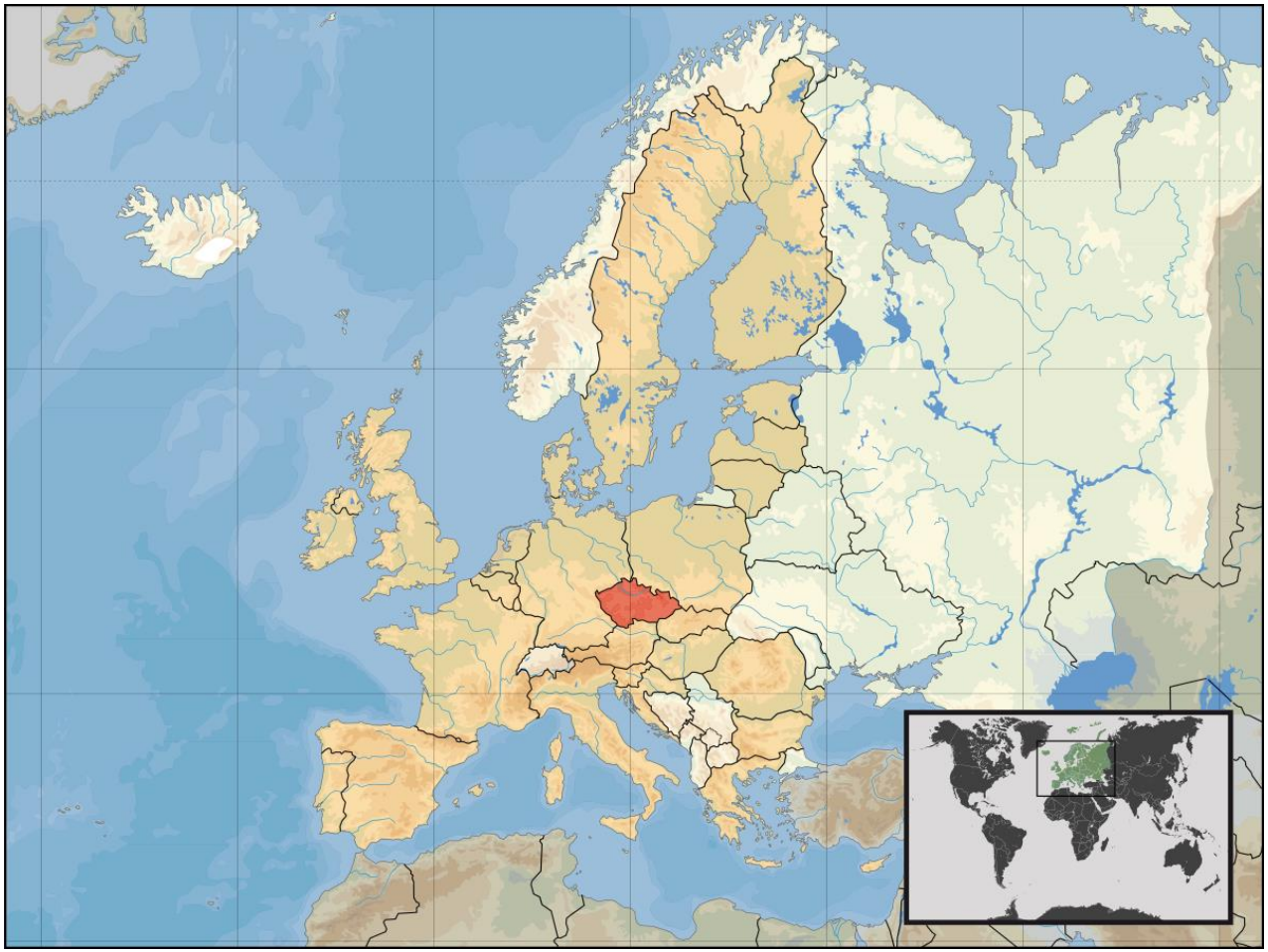
Czech Technical University in Prague, Computing and Information Centre
- IS Analyst: institutional CRIS

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

My country: Czech Republic



Population: ~10M

GDP per capita: ~20K USD
(Peru: ~6.5K USD)

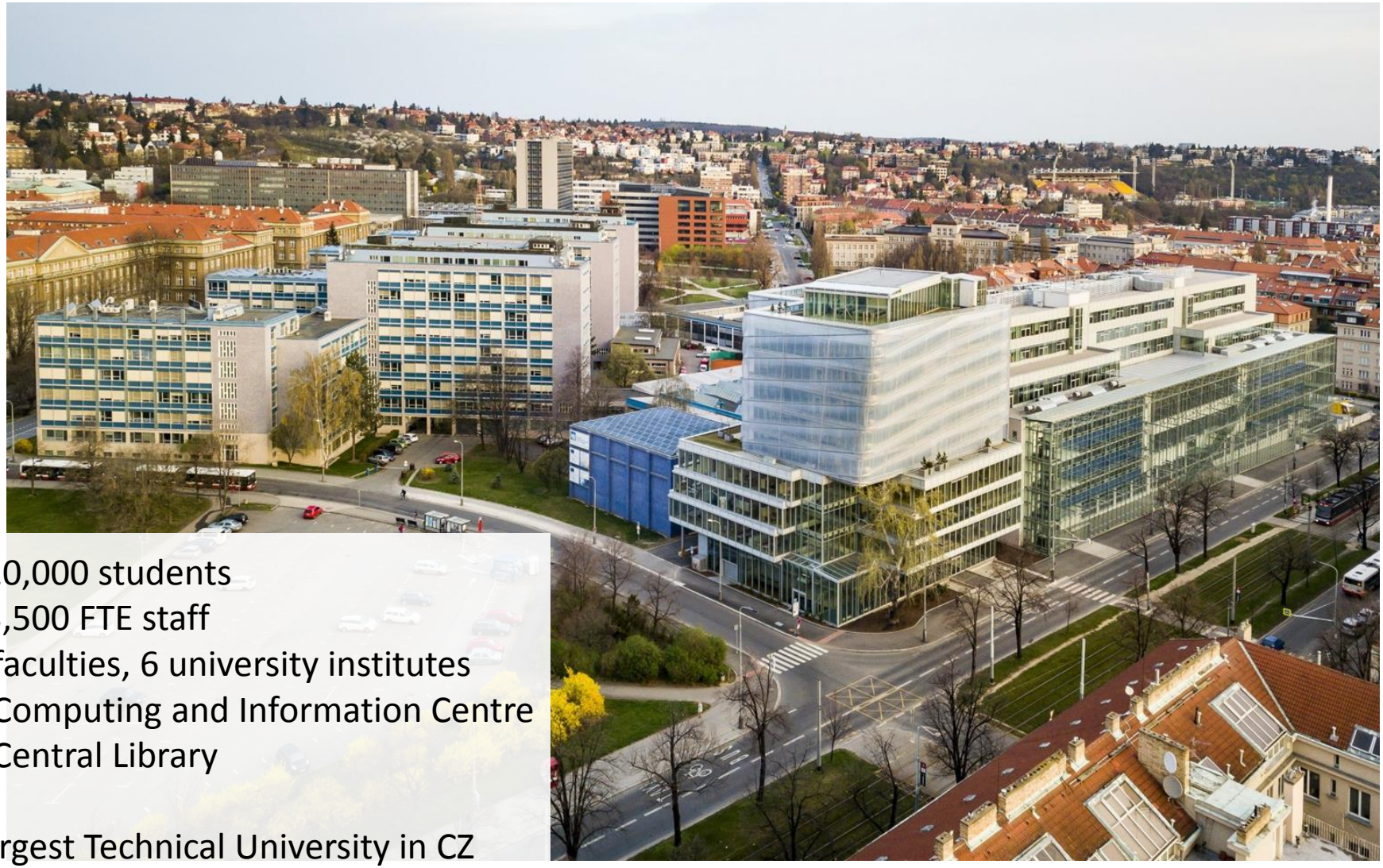
Capital: Prague, Praga, Prága, Praha

The Charles University

~48,000 students
~3,000 staff
17 faculties, 4 university institutes

Largest University in CZ
Est. 1348

The Czech Technical University in Prague



~20,000 students
~3,500 FTE staff
8 faculties, 6 university institutes
+ Computing and Information Centre
+ Central Library

Largest Technical University in CZ
Est. 1707

Research Information

= Information about research

;-)

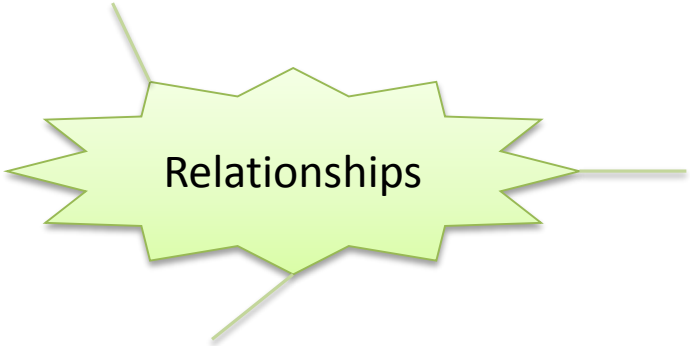
Current Research Information System

- Ongoing
- Past, of current interest
- Currently planned

What is Research Information?

The process of research

- Research projects
- Funding
- Research infrastructures



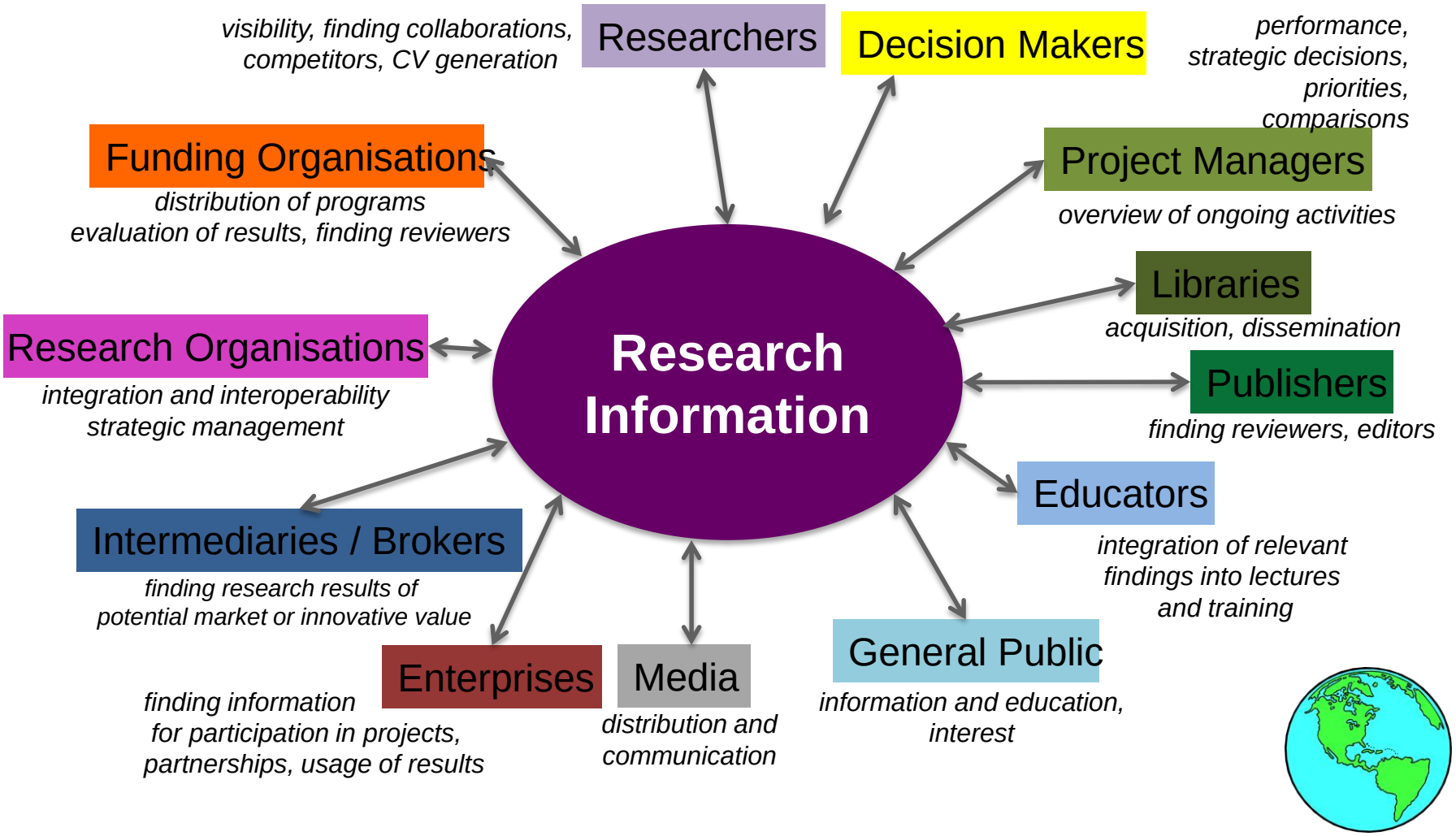
The research actors

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

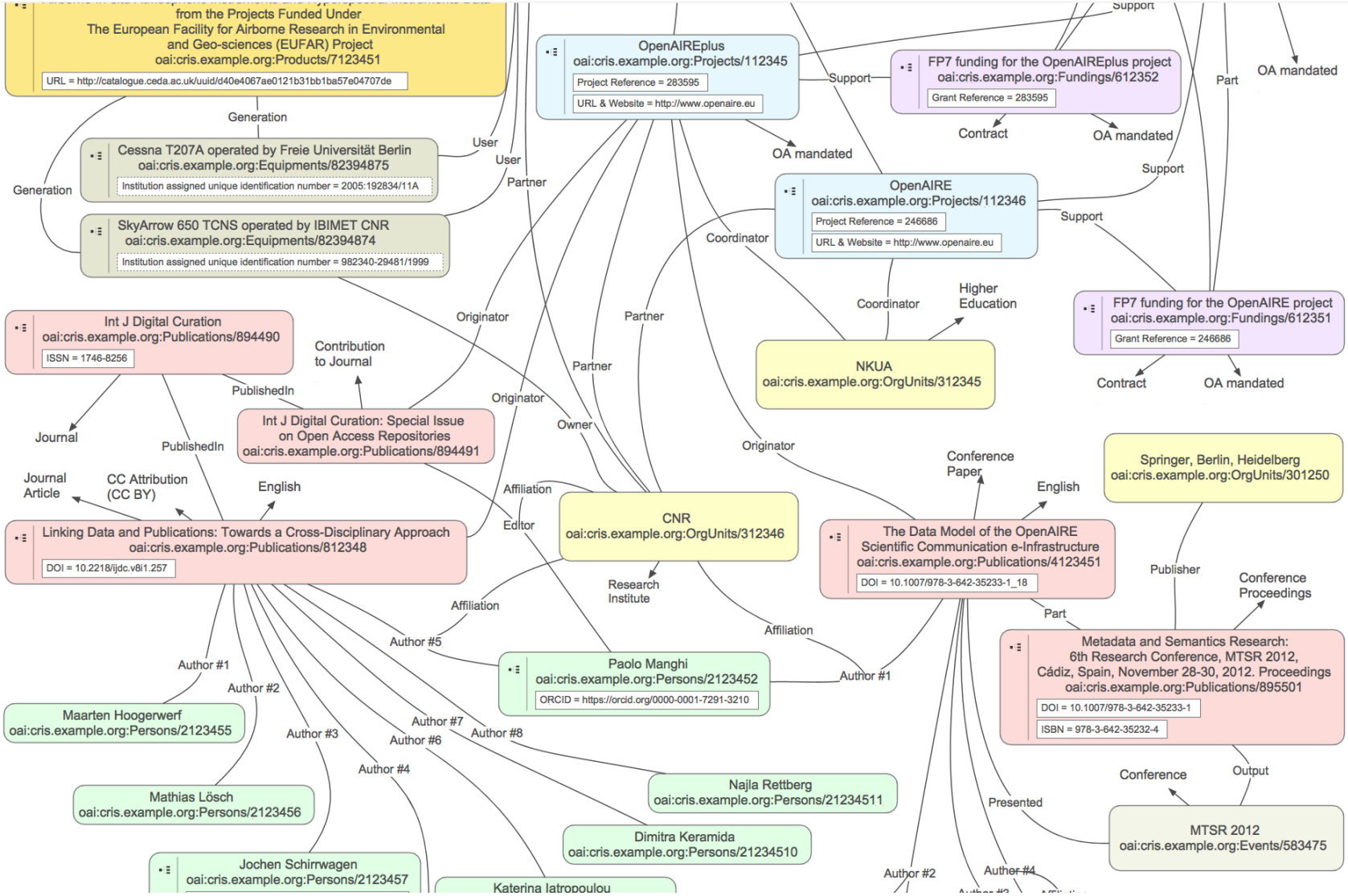
Research results

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

Who needs Research Information?



Research Information is heavily interlinked



Different viewpoints on Research Information

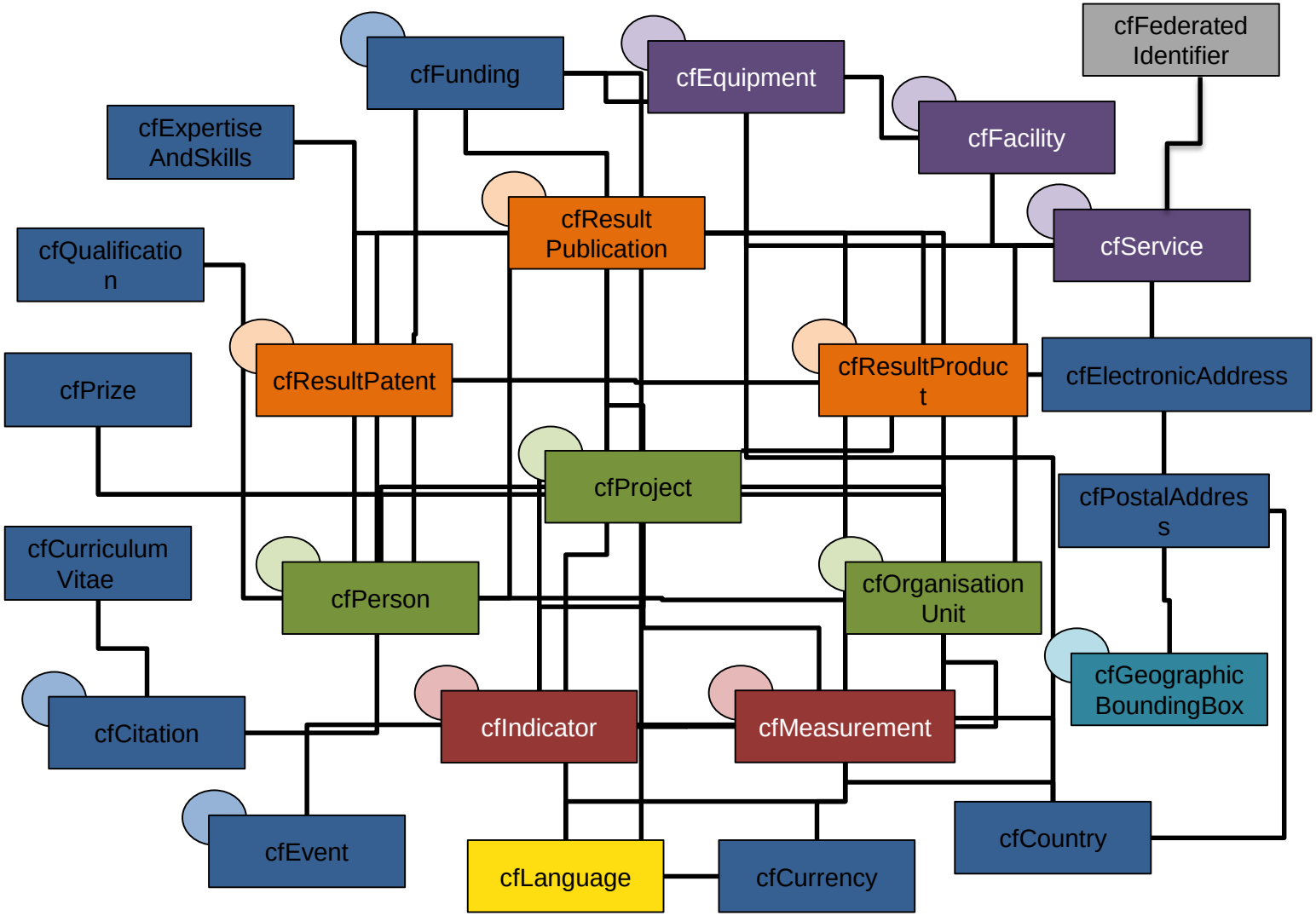
Information perspective	Useful for
Researcher, his/her activities, results, …	Researcher profile / CV system
Project, its consortium, team, results, …	Project webpage / report
Organisational Unit, its activities, results, ...	Unit webpage / report / evaluation report
Publication	Bibliometrics / panel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

Common European Research Information Format

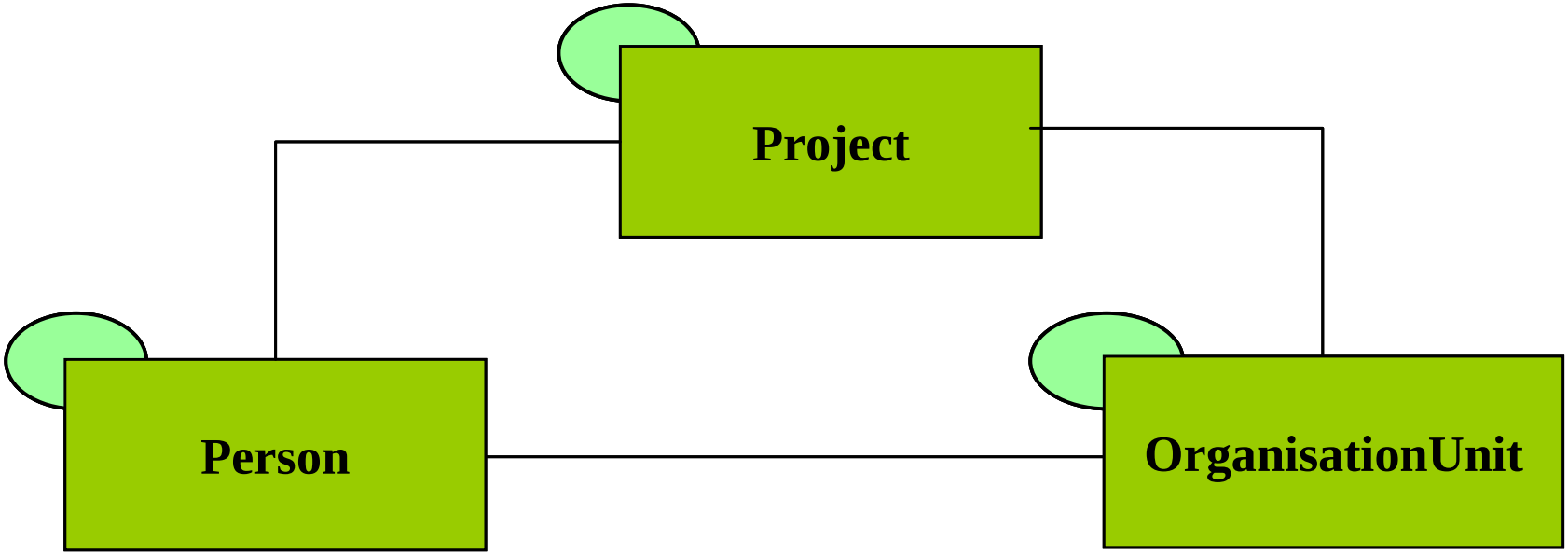
- CERIF is an EU Recommendation to Member States
- The European Commission (EC) has authorised euroCRIS to maintain and develop CERIF and its usage

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

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.

OrgUnit definition

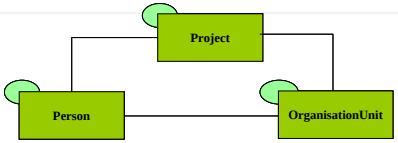
Organisation Unit: an organisation, a unit therein, a committee or any other group of people that has a collective goal. Organisation Units are not necessarily formalized as legal entities.

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.

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.

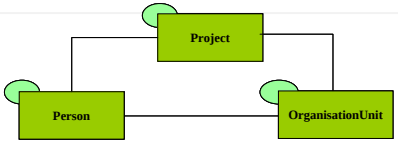


CERIF Base Entities

Project
ID
URI
Acronym
StartDate
EndDate
Title
Abstract
Keywords
Type
Status
Subjects

Person
ID
URI
Gender
FirstNames
OtherNames
FamilyNames
NameVariants
ResearchInterest
Keywords

OrganisationUnit
ID
URI
Acronym
Name
ResearchActivity
Keywords
Type
Status



CERIF Base Entities

cfProject
cfID
cfURI
cfAcronym
cfStartDate
cfEndDate

cfTitle
cfAbstract
cfKeywords

Type
Subject
Status

cfPerson
cfID
cfURI
cfGender
cfBirthdate

cfDescription
cfKeywords

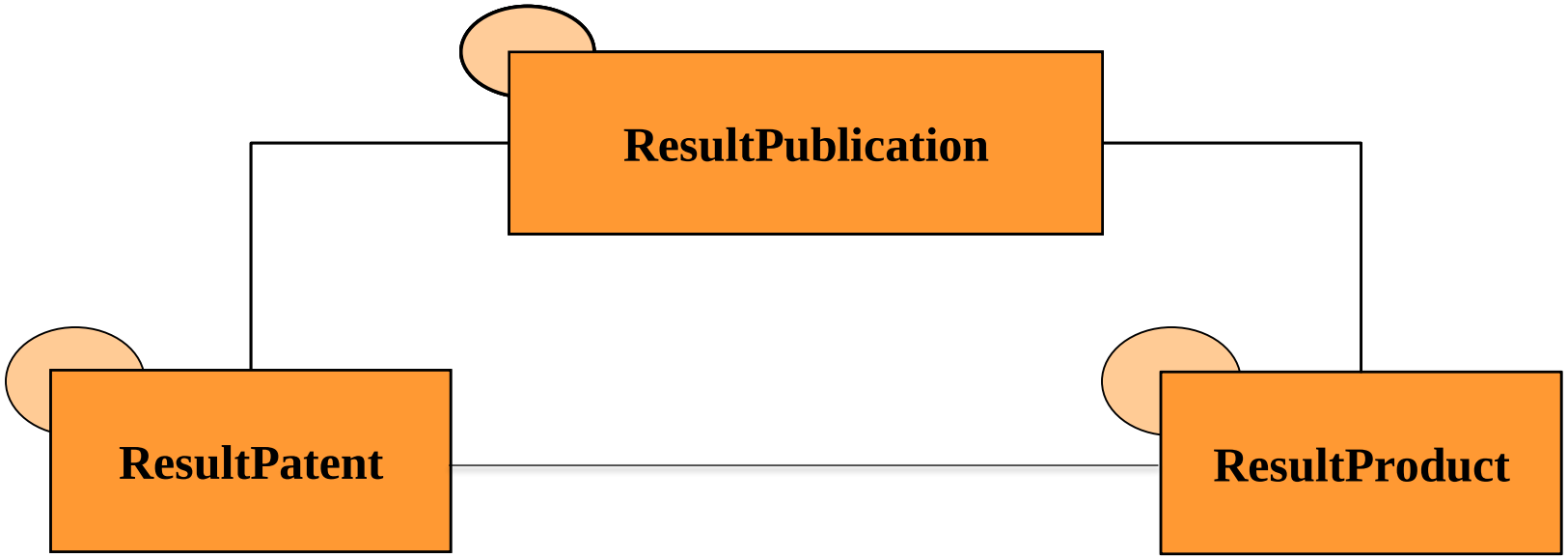
cfFamilyNames
cfFirstNames
cfOtherNames

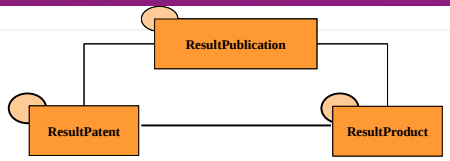
cfOrganisationUnit
cfID
cfURI
cfAcronym

cfName
cfDescription
cfKeywords

Type
Status

CERIF Result Entities



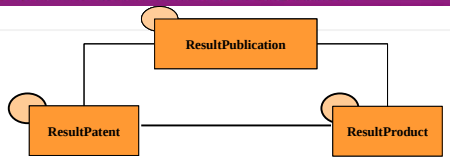


CERIF Result Entities

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

cfResultPublication
cfID
cfURI
cfNumber
cfPublicationDate
cfStartPage
cfEndPage
cfTotalPages
cfEdition
cfSeries
cfIssue
cfVolume

cfBibliographic Note
cfVersionInfo
cfAbbreviation
cfTitle
cfSubtitle
cfAbstract
cfKeywords

Type
Subject
Status

cfResultProduct
cfID
cfURI

cfName
cfDescription
cfKeywords
cfVersionInfo

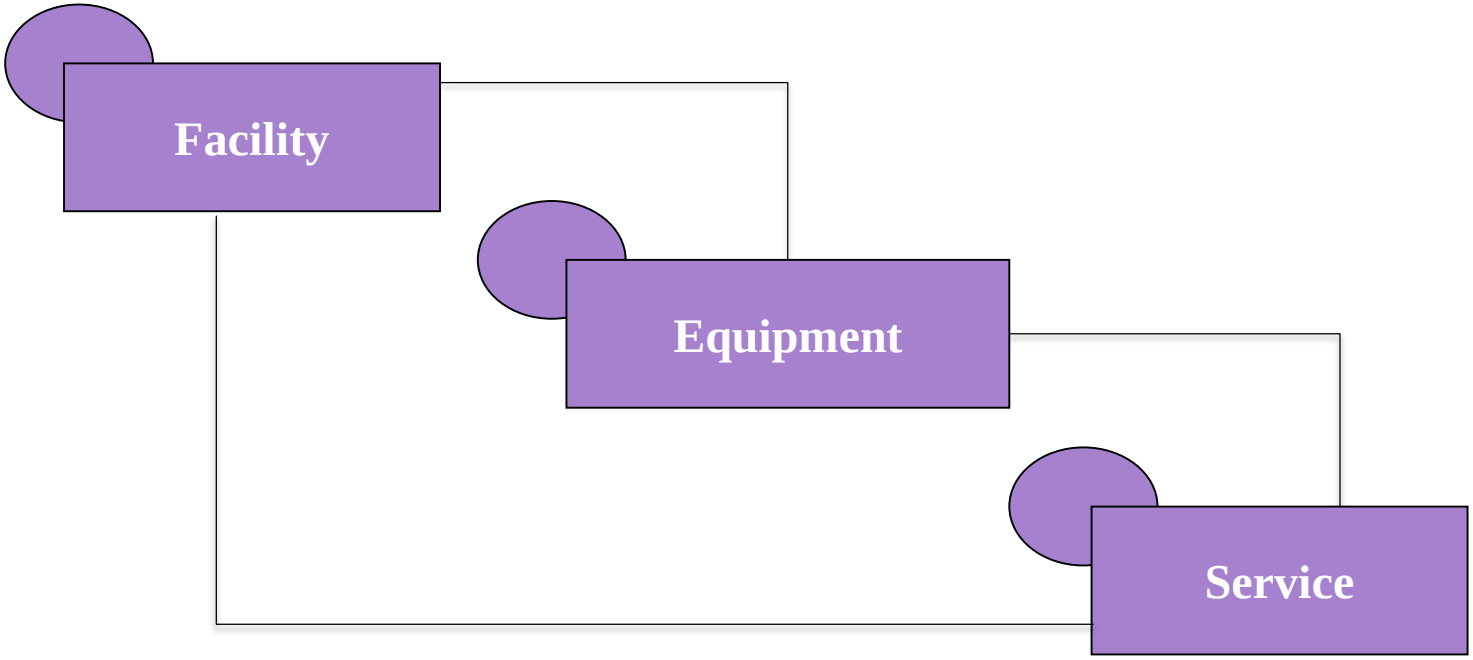
Type
Subject
Status

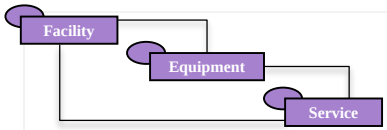
cfResultPatent
cfID
cfURI
cfPatentNumber
cfCountryCode
cfRegistrationDate
cfApprovalDate

cfName
cfAbstract
cfKeywords
cfVersionInfo

Type
Subject
Status

CERIF Infrastructure Entities



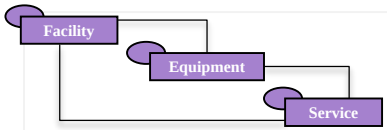


CERIF Infrastructure Entities

Facility
ID
Acronym
URI
Title
Description
Keywords
Type
Status
Subject

Equipment
ID
Acronym
URI
Title
Description
Keywords
Type
Status
Subject

Service
ID
Acronym
URI
Title
Description
Keywords
Type
Status
Subject



CERIF Infrastructure Entities

cfName
cfDescription
cfKeywords
cfFacility
cfID
cfURI
cfAcronym
Type
Status
Subject

cfName
cfDescription
cfKeywords
cfEquipment
cfID
cfURI
cfAcronym
Type
Status
Subject

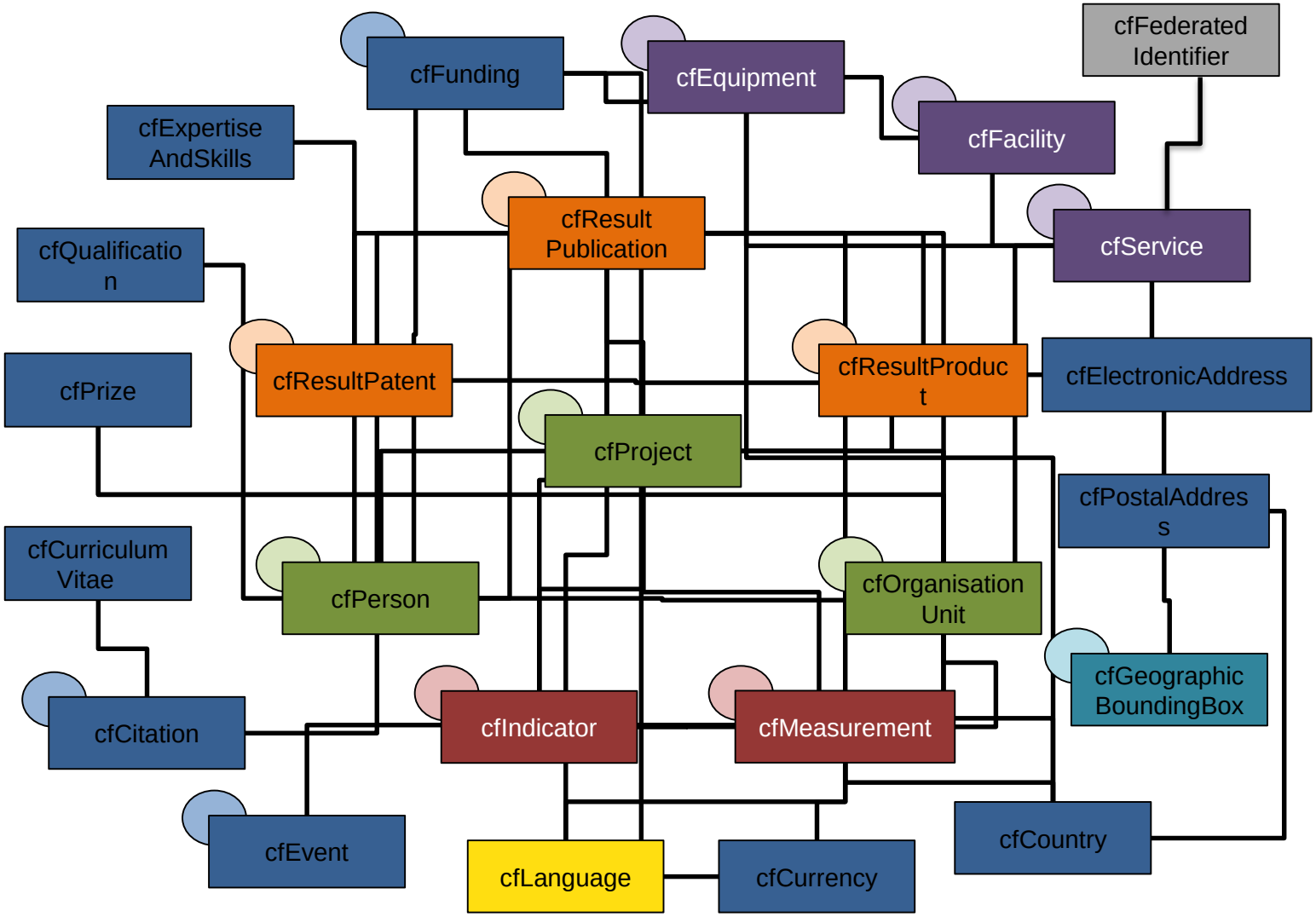
cfName
cfDescription
cfKeywords
cfService
cfID
cfURI
cfAcronym
Type
Status
Subject

CERIF General Pattern

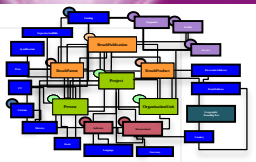
A typical CERIF entity has:

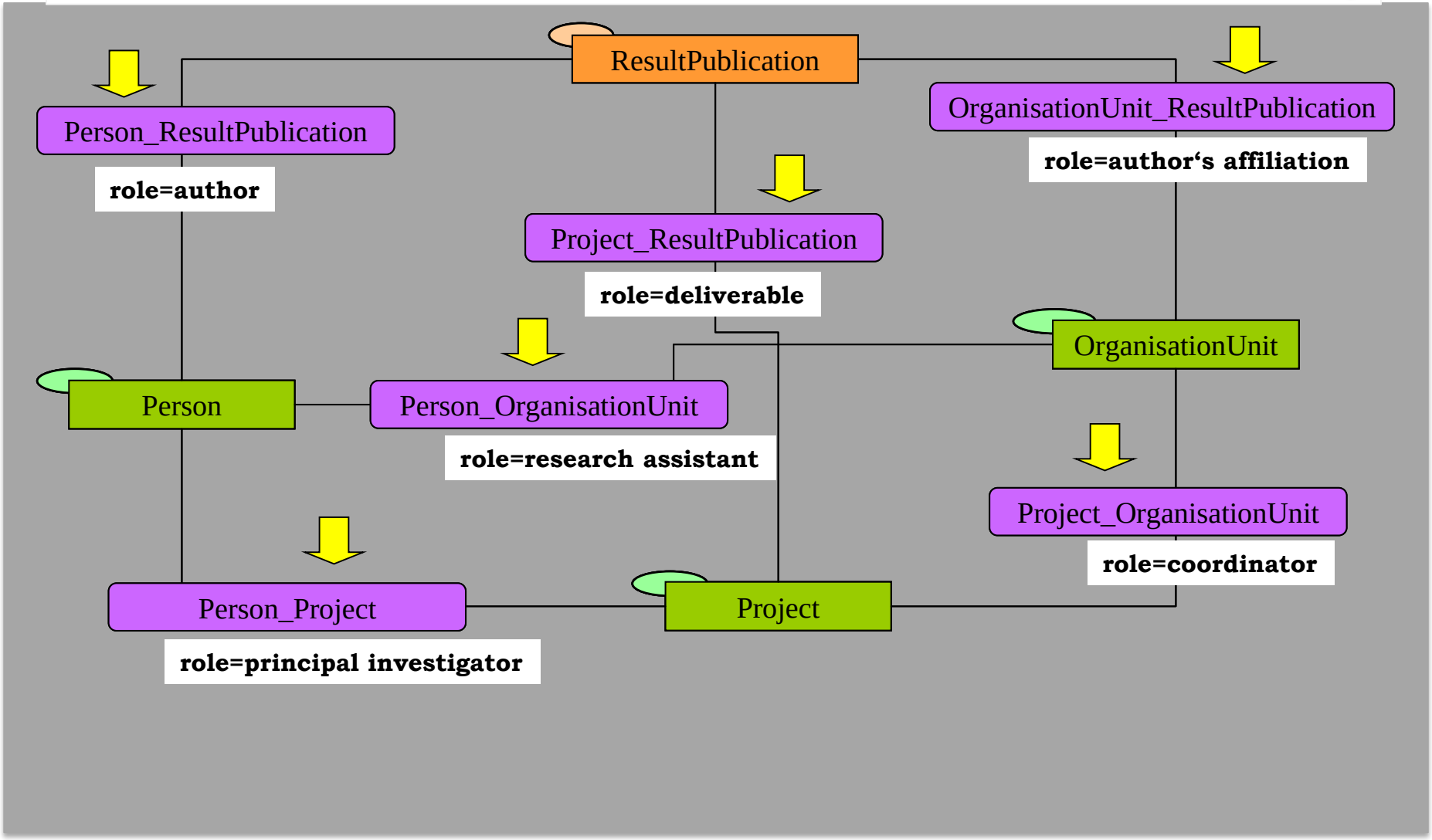
- Identifier (internal)
- Attributes
  - the basic ones
  - the multi-lingual ones
- External Identifiers
- Classifications
  - Type
  - Status
  - Subject area
  - + other
- Links
  - to other entities
  - recursive, a.k.a. self-referential

CERIF 1.6

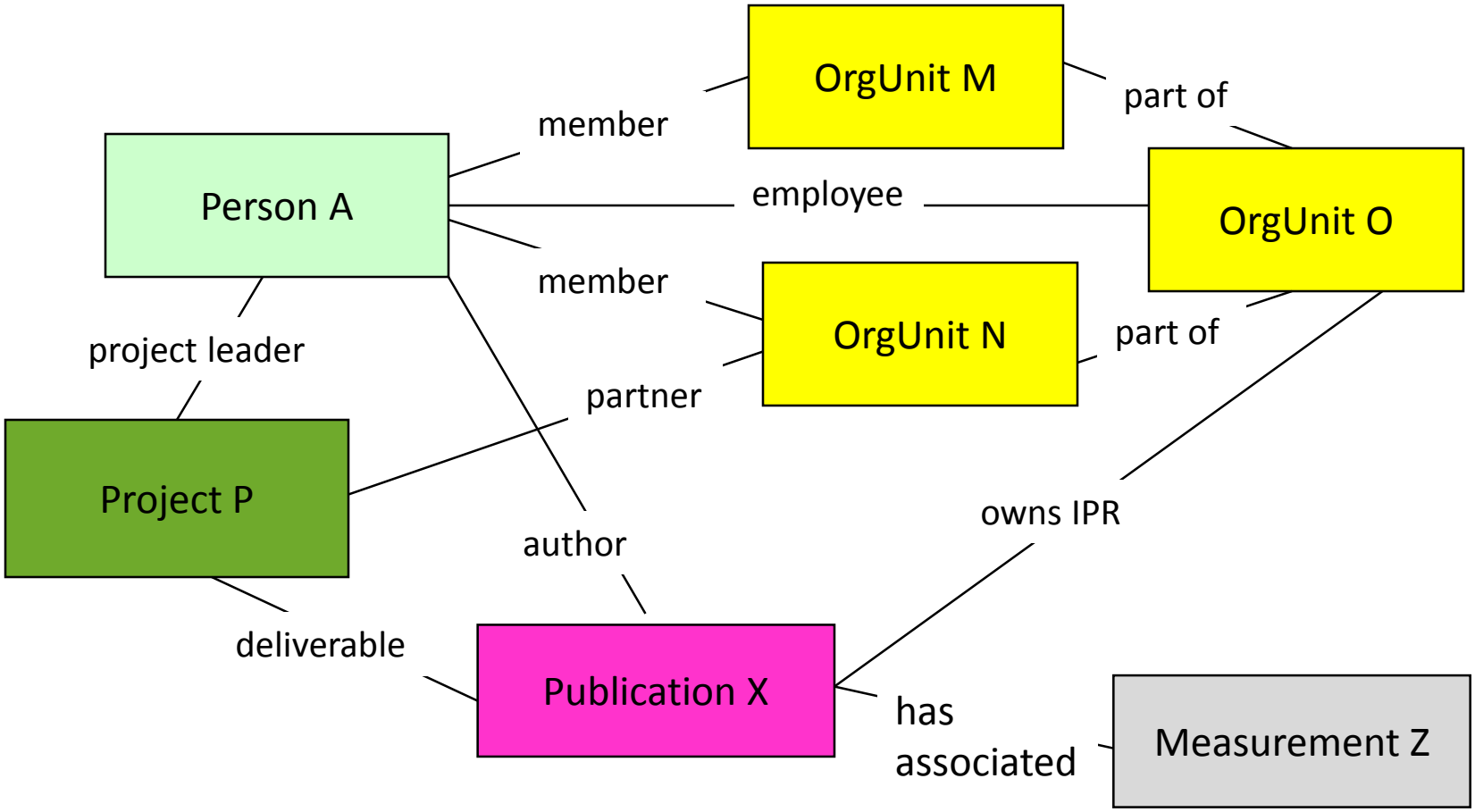

cfFederated Identifier
cfFunding
cfEquipment
cfExpertise AndSkills
cfFacility
cfResult Publication
cfQualificatio n
cfService
cfPrize
cfResultPatent
cfResultProduc t
cfElectronicAddress
cfProject
cfPostalAddres s
cfCurriculum Vitae
cfPerson
cfOrganisation Unit
cfGeographic BoundingBox
cfCitation
cfIndicator
cfMeasurement
cfEvent
cfCountry
cfLanguage
cfCurrency

Some CERIF Link Entities

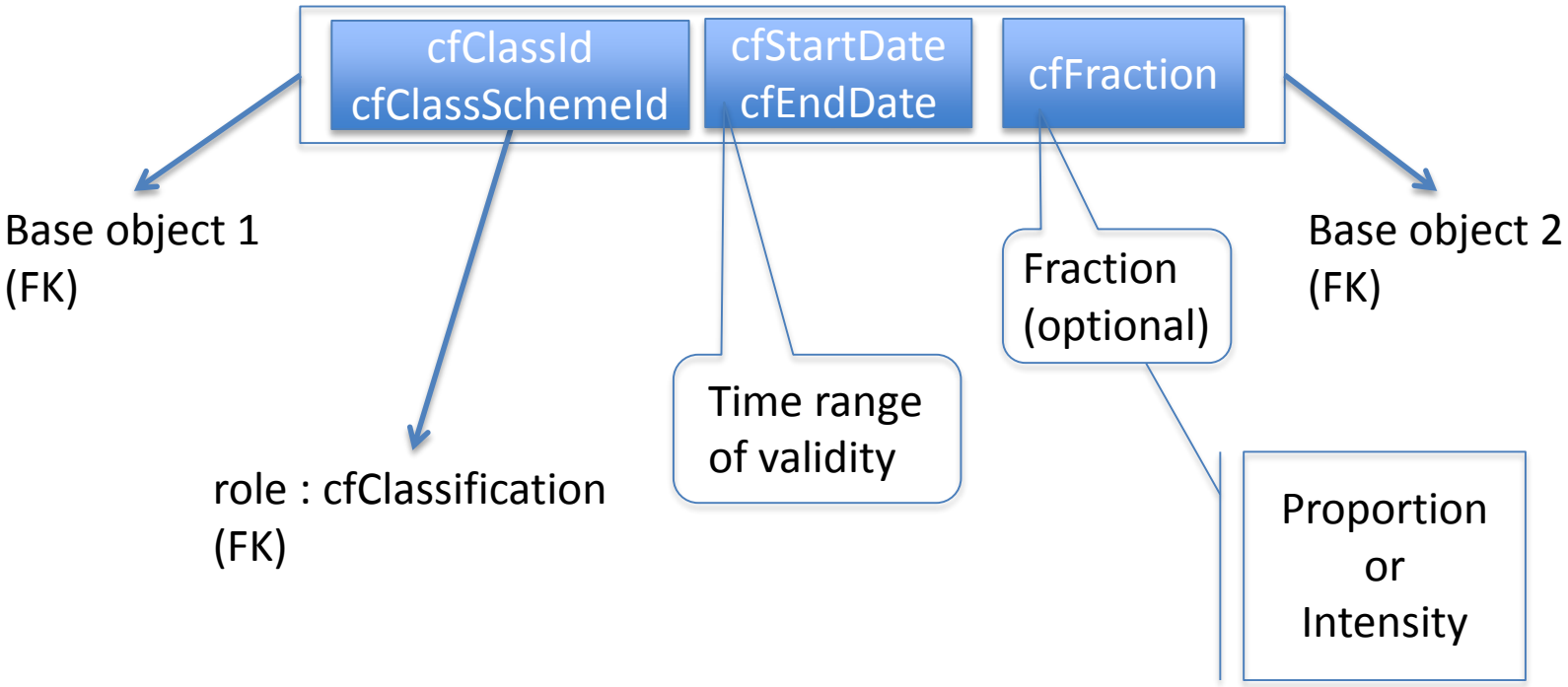




Another example
(slide by Keith Jeffery)

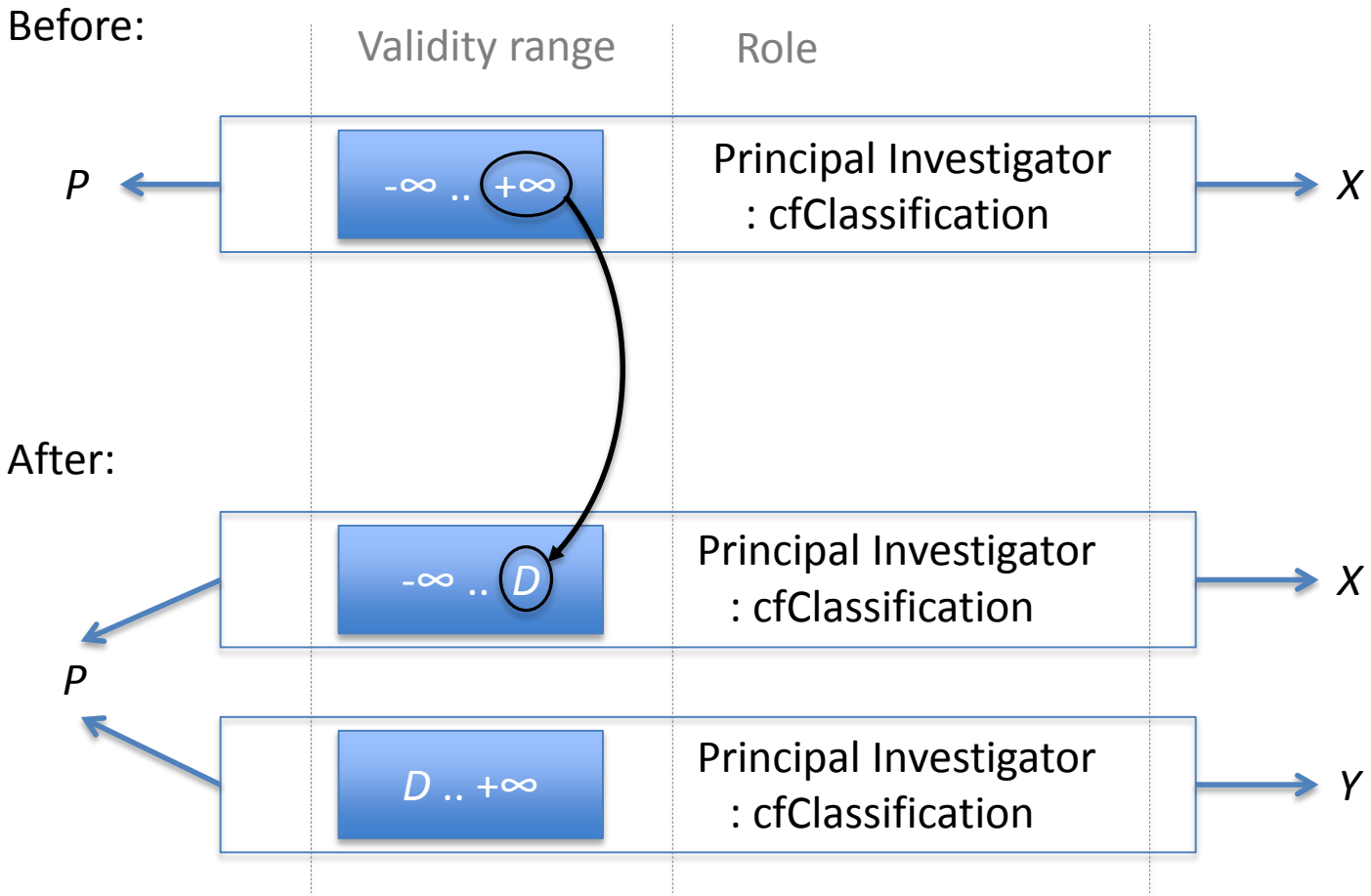


Generic Linking Entity Structure

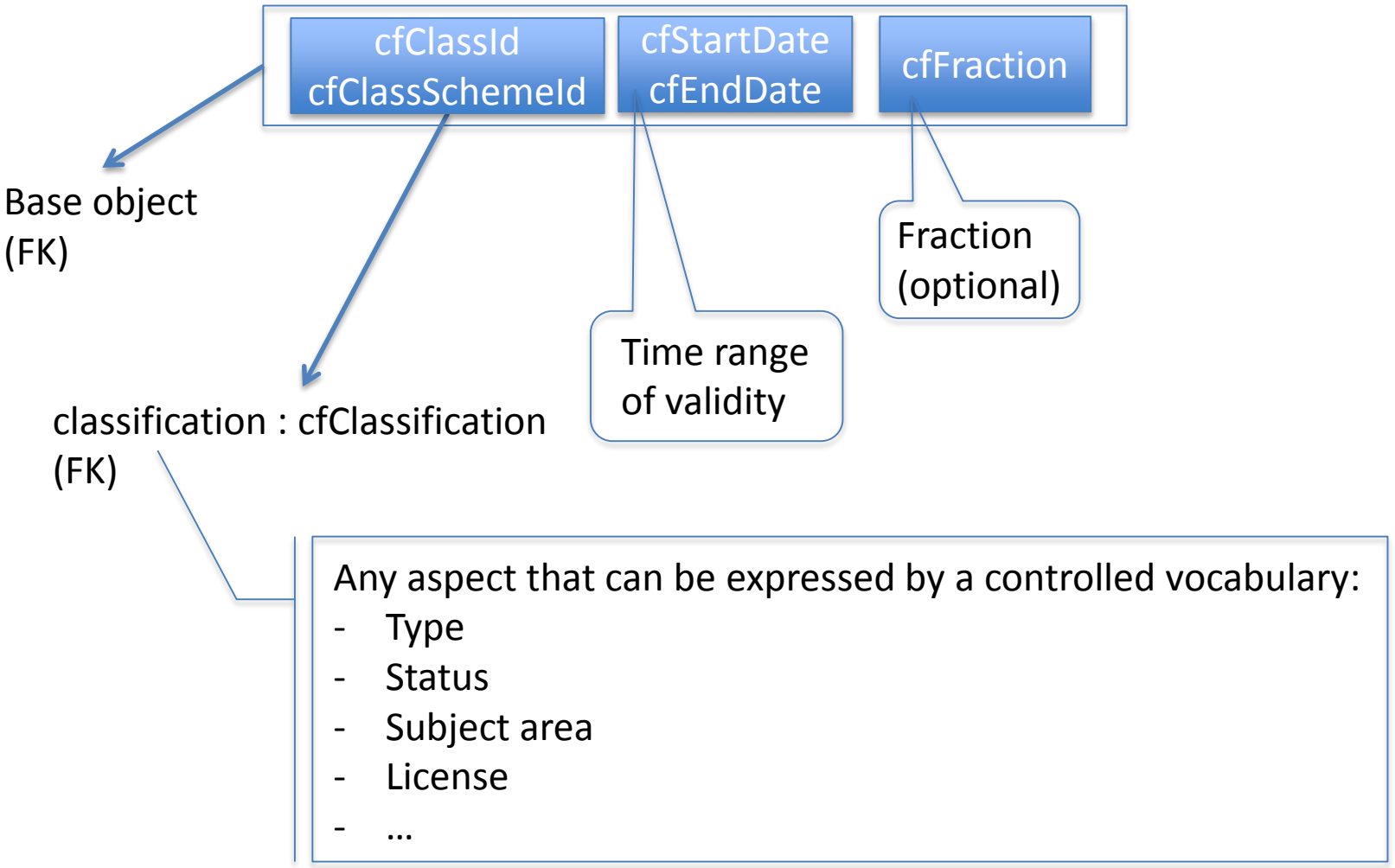


Recording Change in CERIF

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

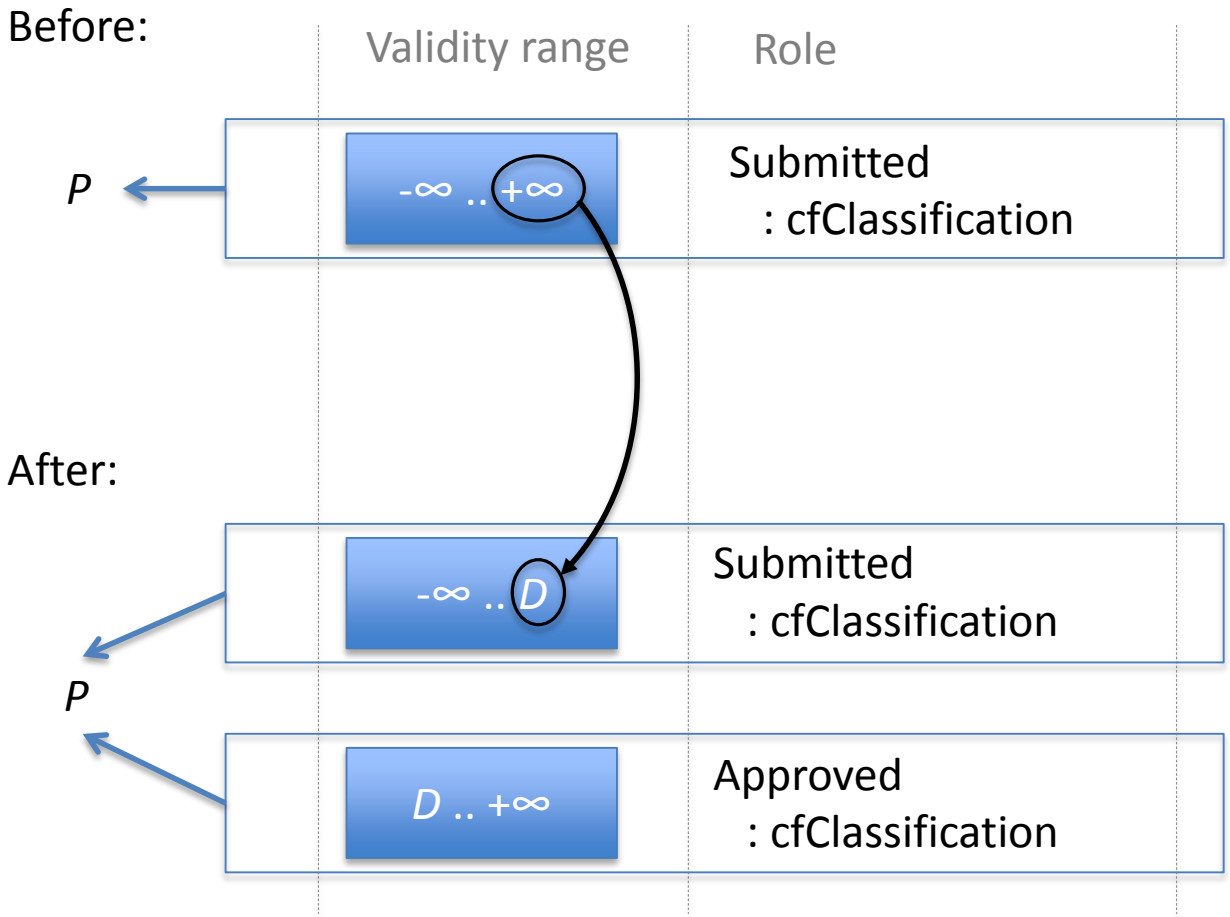


Generic Classification Link Structure



Recording Status Change in CERIF

Example: The project *P* is approved effective date *D*.



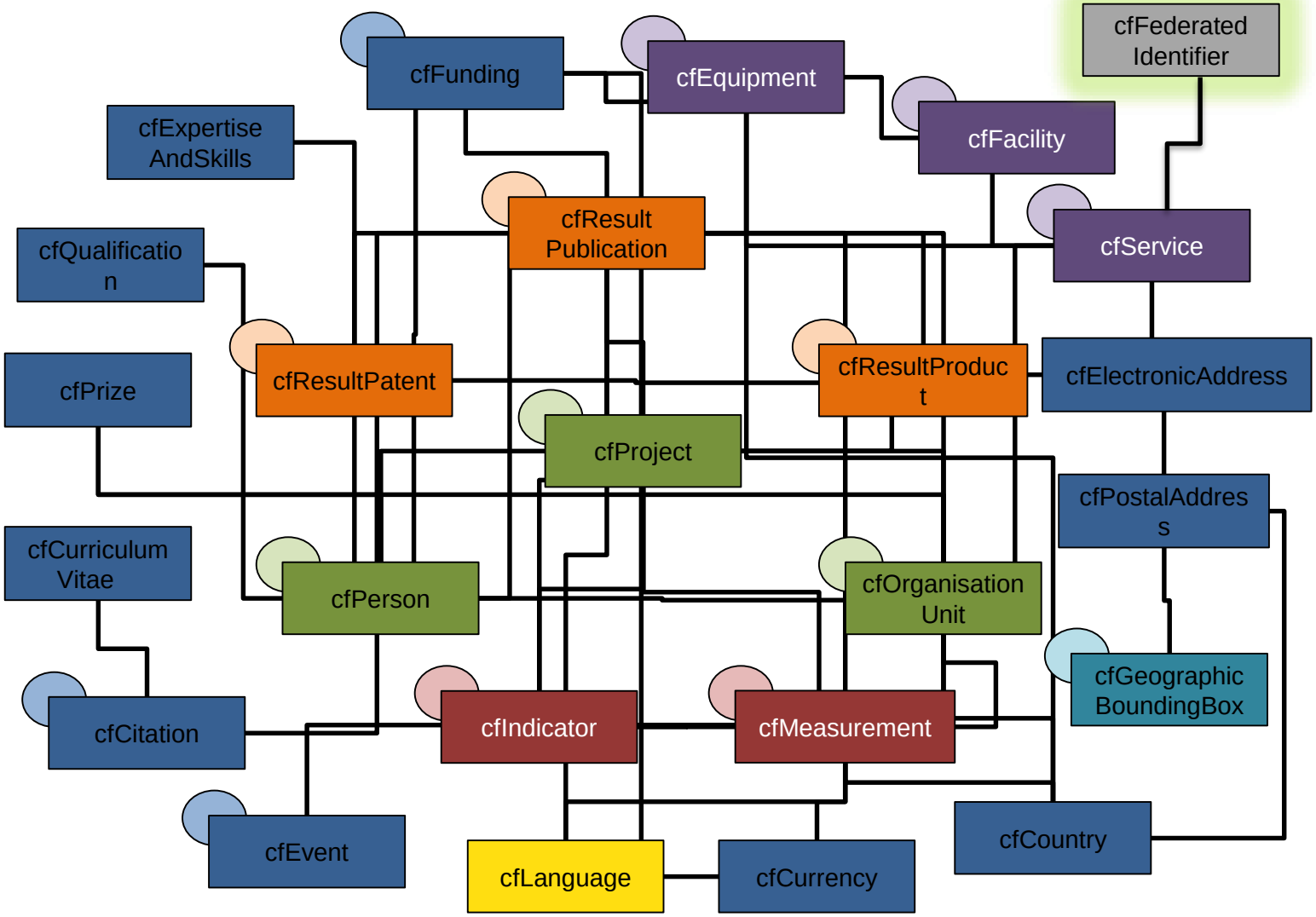
Recording History in CERIF

=> The database can be queried for:

- The present state
- The state at any instant in the past

Indispensable for any serious analytics

CERIF 1.6


cfFederated Identifier
cfFunding
cfEquipment
cfExpertise AndSkills
cfFacility
cfResult Publication
cfQualificatio n
cfService
cfPrize
cfResultPatent
cfResultProduc t
cfElectronicAddress
cfProject
cfPostalAddres s
cfCurriculum Vitae
cfPerson
cfOrganisation Unit
cfGeographic BoundingBox
cfCitation
cfIndicator
cfMeasurement
cfEvent
cfLanguage
cfCurrency
cfCountry

CERIF Federated Identifiers

Publication

- ISBN
- ISSN
- DOI
- WoS Accession Number
- Scopus EID
- PubMed Central ID

Person

- Social Security Number
- ORCID
- Staff Id in HR system
- Author identifier
  - IdRef, DAI, Número Lattes, ResearcherID, Scopus Author ID

Project/Grant

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

Organisation

- VAT Identification Number
- FundRefID
- GridID
- OrgID

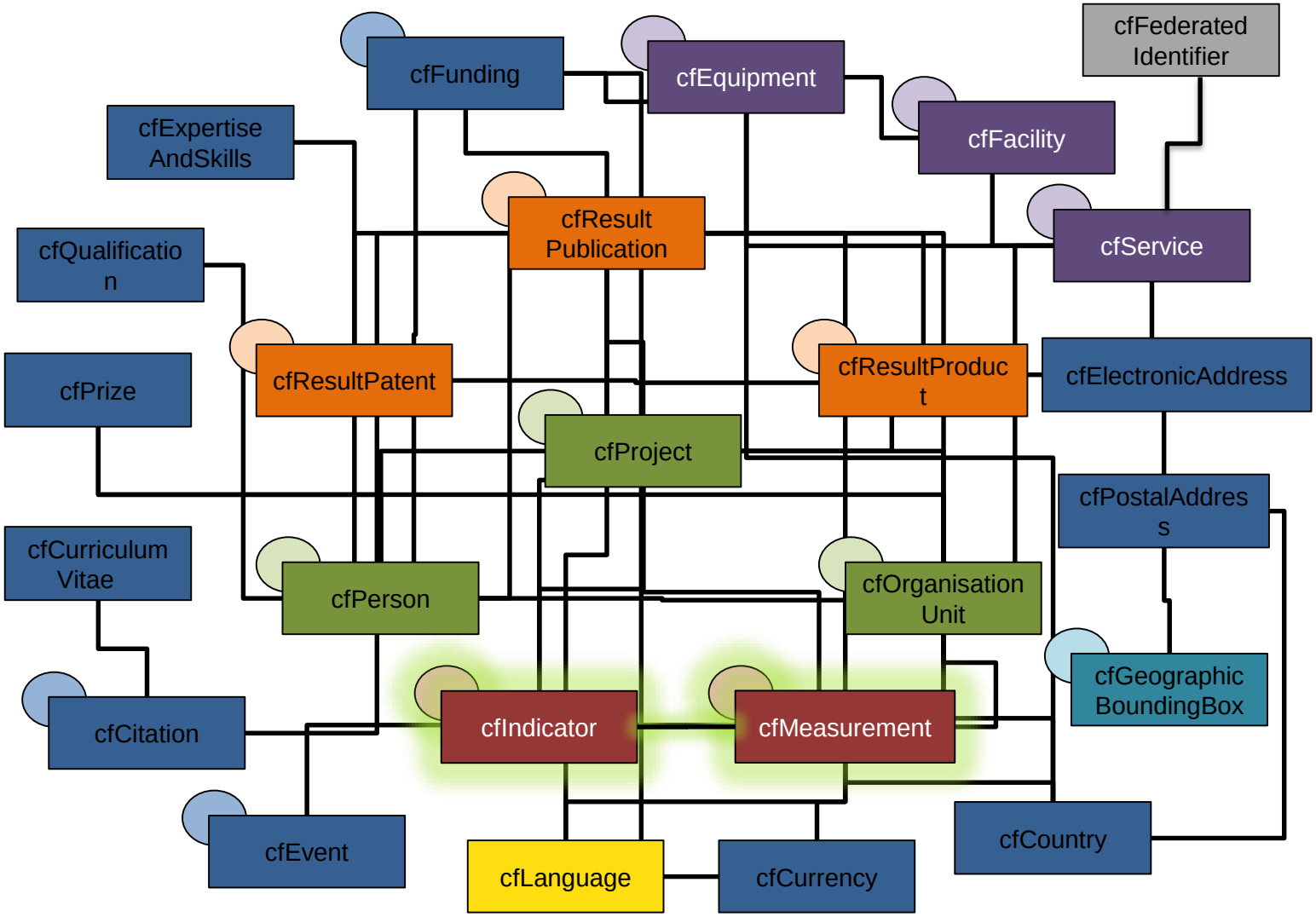
Classification

- External Code

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 1.6


cfFederated Identifier
cfFunding
cfEquipment
cfExpertise AndSkills
cfFacility
cfResult Publication
cfQualificatio n
cfService
cfPrize
cfResultPatent
cfResultProduc t
cfElectronicAddress
cfProject
cfPostalAddres s
cfCurriculum Vitae
cfPerson
cfOrganisation Unit
cfGeographic BoundingBox
cfCitation
cfIndicator
cfMeasurement
cfEvent
cfCountry
cfLanguage
cfCurrency

Indicators & Measurements: examples

- University *U* has published *N* journal articles in discipline *D* in year *Y*

Number of journal articles

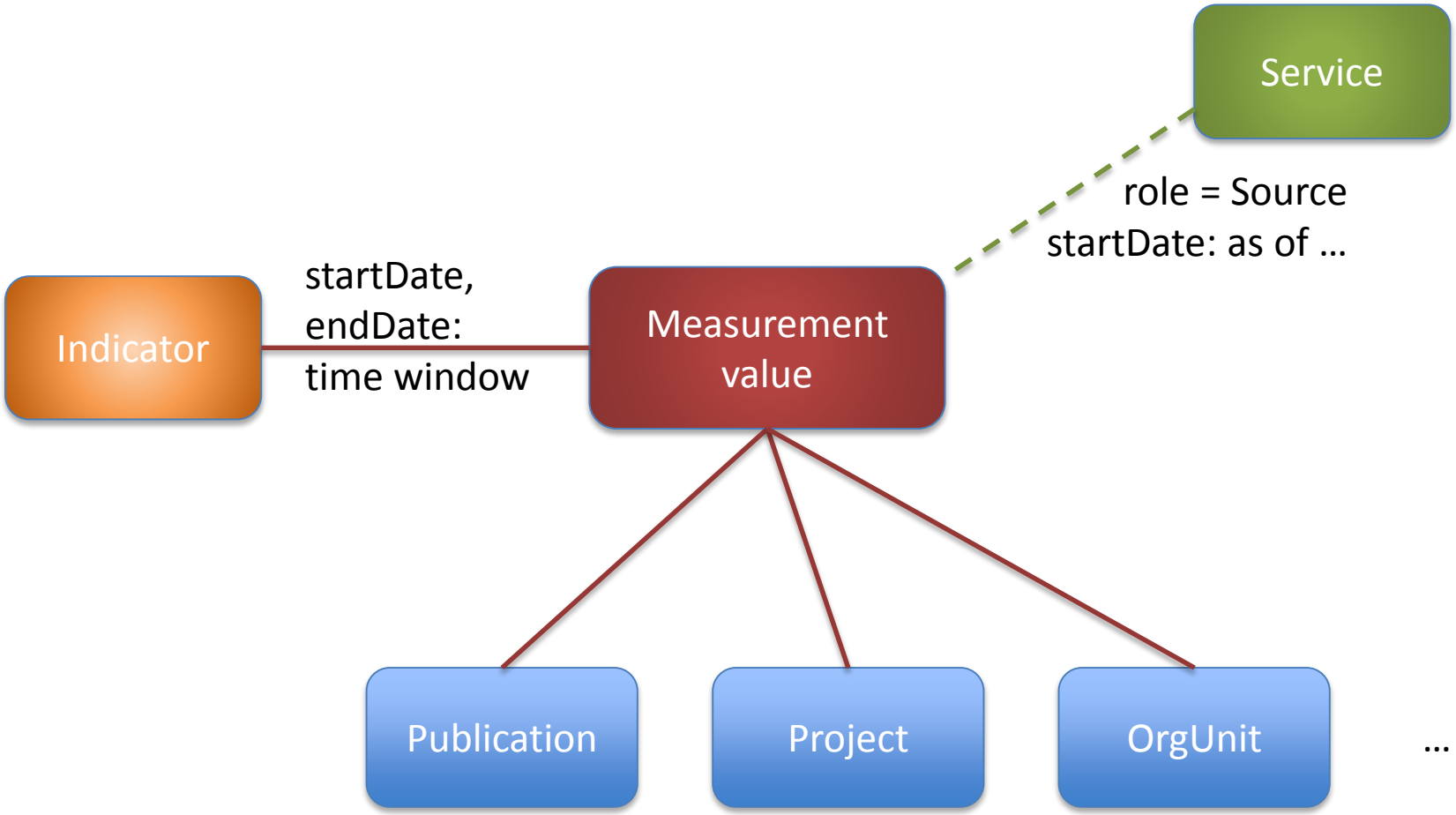
- Department *O* has H-index *H*, measured for years *Y*–9 .. *Y*, based on the Web of Science data as of date *D*

H-index

- School *S* has attracted total funding *F* for years *Y*–2 .. *Y*

Sum of funding

Indicators & Measurements



Indicators & Measurements (other examples)

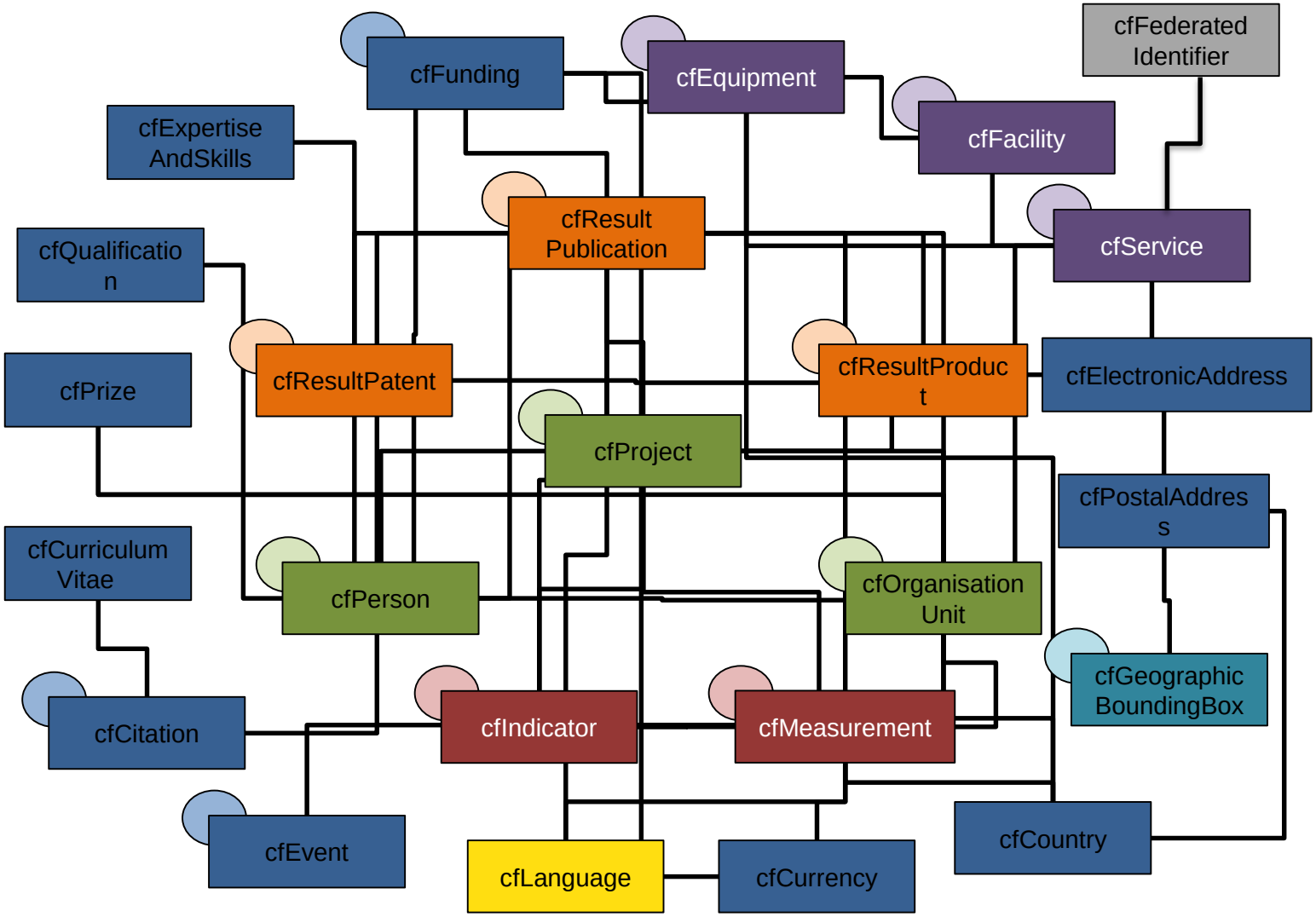
- economic and commercial
  - economic
    - impact on business
      - » improving performance of existing businesses
        - increased turnover *by 1.2M€ in 2012*
        - time savings *of 14.56%*
        - reduced costs *by 42%*
      - » new products/processes
        - creating numbers of new products/services
        - commercialising / other success measures

Indicators → increased turnover, time savings, reduced costs

Measurements → *by 1.2M€ in 2012*, *of 14.56%*, *by 42%*

Extract from the MICE Project's List of Indicators

CERIF 1.6

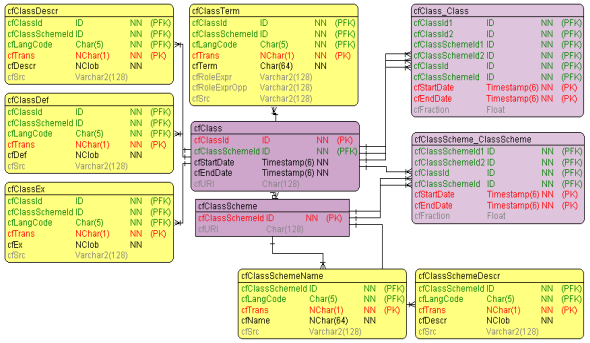

cfFederated Identifier
cfFunding
cfEquipment
cfExpertise AndSkills
cfFacility
cfResult Publication
cfQualificatio n
cfService
cfPrize
cfResultPatent
cfResultProduc t
cfElectronicAddress
cfProject
cfPostalAddres s
cfCurriculum Vitae
cfPerson
cfOrganisation Unit
cfGeographic BoundingBox
cfCitation
cfIndicator
cfMeasurement
cfEvent
cfCountry
cfLanguage
cfCurrency

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

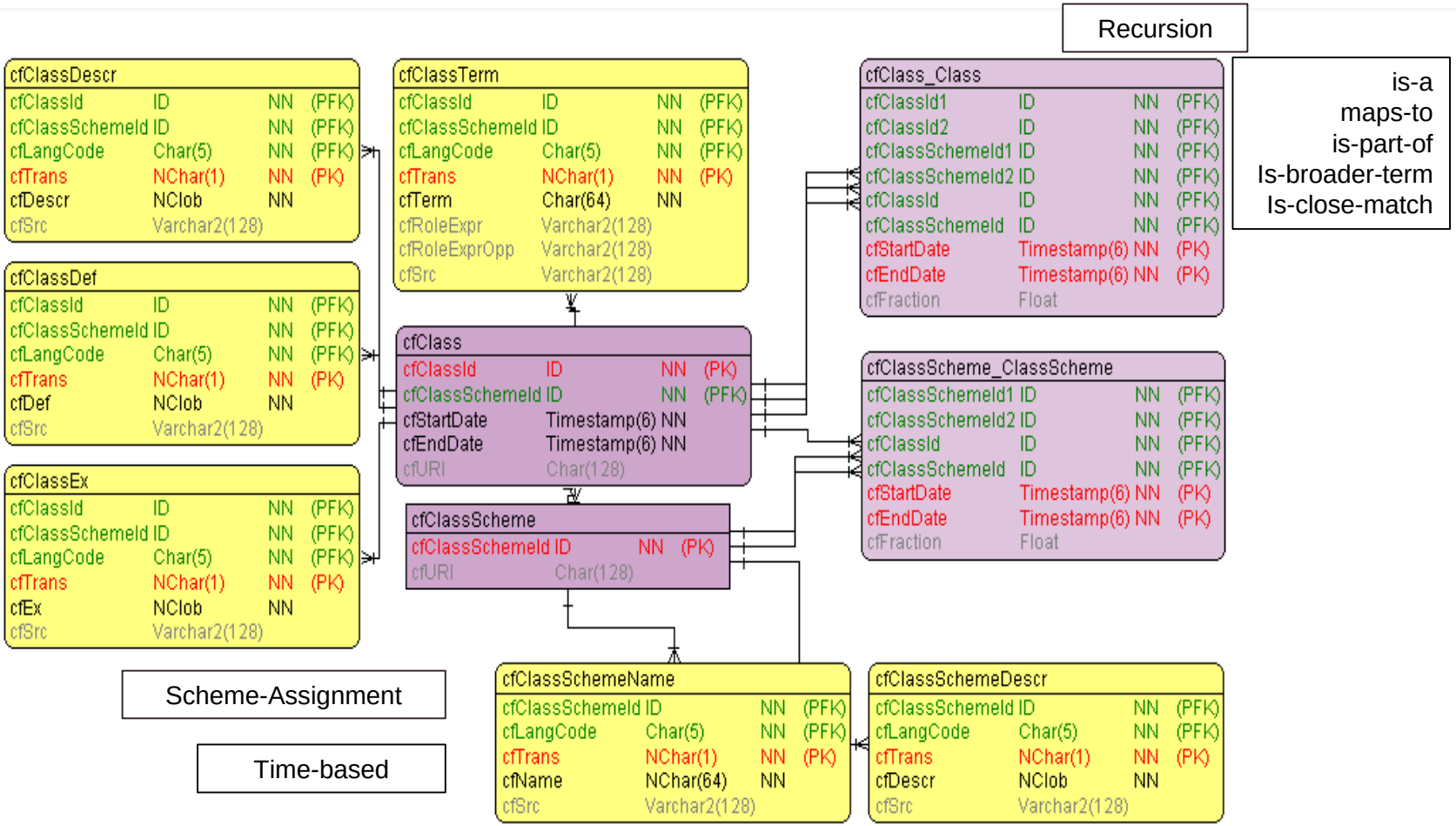
Enables to manage

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

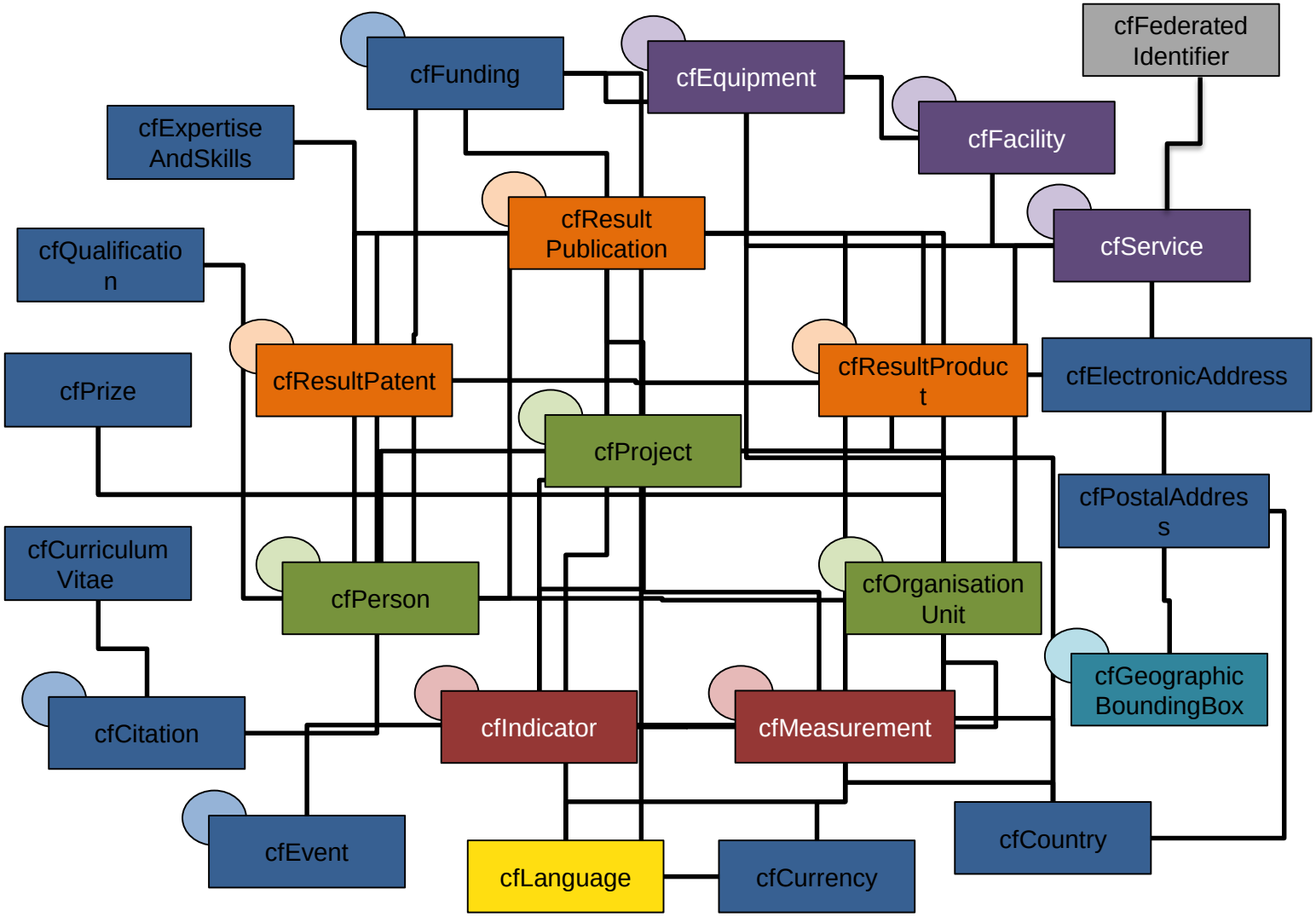
Allows for Mappings between Schemes

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

CERIF Semantic Layer (Declared Semantics)



CERIF 1.6


cfFederated Identifier
cfFunding
cfEquipment
cfExpertise AndSkills
cfFacility
cfResult Publication
cfQualificatio n
cfService
cfPrize
cfResultPatent
cfResultProduc t
cfElectronicAddress
cfProject
cfPostalAddres s
cfCurriculum Vitae
cfPerson
cfOrganisation Unit
cfGeographic BoundingBox
cfCitation
cfIndicator
cfMeasurement
cfEvent
cfCountry
cfLanguage
cfCurrency

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

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

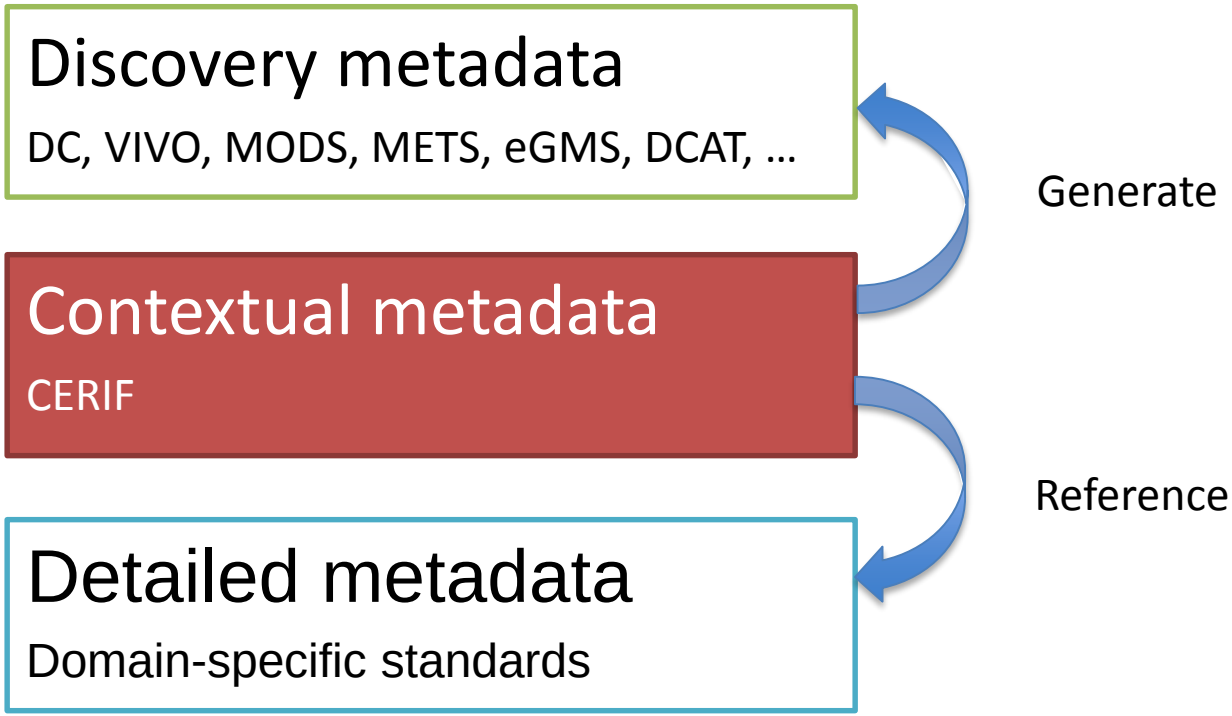
Basic Information

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				

Metadata Layers



The CERIF Evolution

Similar Ideas
UN/UNESCO
OECD
CODATA

EU Working Group on Research Databases Workshop

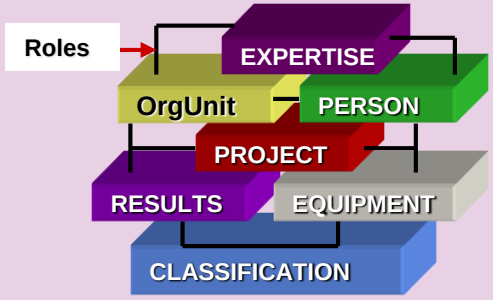
CERIF 91

Acronym: ERGO
Participant:
Keith Jeffery, Anne Asserson, many more
Organisations:
Rutherford Appleton, University of Bergen, …

- Networking of DBs
- Exchange of Records

- EC Recommendation to Member States

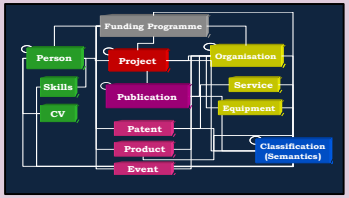
CERIF 2000 Model



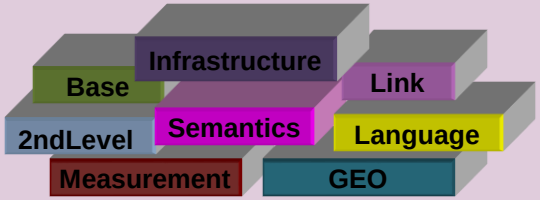
- Data Model
- Multilinguality
- Controlled Vocabulary
- Roles / Types
- User-driven

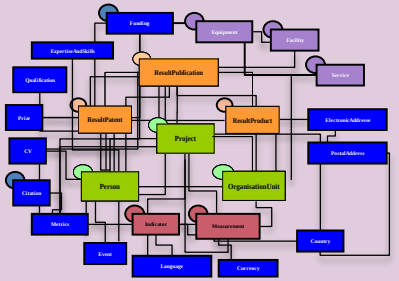
- EC Recommendation to Member States

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 Dataset-ready

Profiles

+ Linked Data

1987 1991 2000 2006 2008 2012 2013 2015

euroCRIS
Current Research Information Systems