Introduction to euroCRIS and CERIF

Ed Simons,
Radboud Univeristy, The Netherlands
President of euroCRIS
Ankara, Open Access Week, 20th of October 2015

Content of the Presentation:

Part 1: The euroCRIS organisation

Part 2: The CERIF data model

Part 1: The euroCRIS Organisation

What is euroCRIS?

We could start with looking at the name “euroCRIS”, which has two parts:

- “euro”: referring to “Europe”, where the organisation was established and still has its base.
- “CRIS”: short for “*Current Research Information System*” (sometimes also “RIS” is used (“*Research Information System*”)).

So, euroCRIS is a European based organisation dealing with or focusing on CRIS’s.

What Europe is, needs no explanation, but perhaps the word “CRIS” does.
So let’s look a bit more detailed into this concept.

What is a CRIS?

A CRIS is an information system that holds a broad range of *information about* research, in other words: *metadata* on research (metadata = data about data).

What is a CRIS?

The information (metadata) stored in a CRIS, includes the following:

Research projects: title, description, duration, academic field, language(s), level (institutional, national, international), participating institut(e)ions, etc...

Researchers: name, role in the research (PI, researcher, manager, author, reviewer, …), CV-related info (age, field of expertise, educational background, affiliations, positions, etc…)

Organisations involved: name, role or position in the research (e.g. funder, leading unit, etc..), type of organization (university, research institute, network...), contact info, partnerships, etc...

Input for research: amount of money invested in the research, investment in time/personnel (f.t.e.'s) equipment , infrastructure and/or services used, funding sources, etc…

Output of research: publications, datasets, patents, awards, (other) products, (software, media), etc...

Rights information: user/authorization rights, distribution rights, IP (copy) rights, etc...

… and the relations between all these entities/objects.

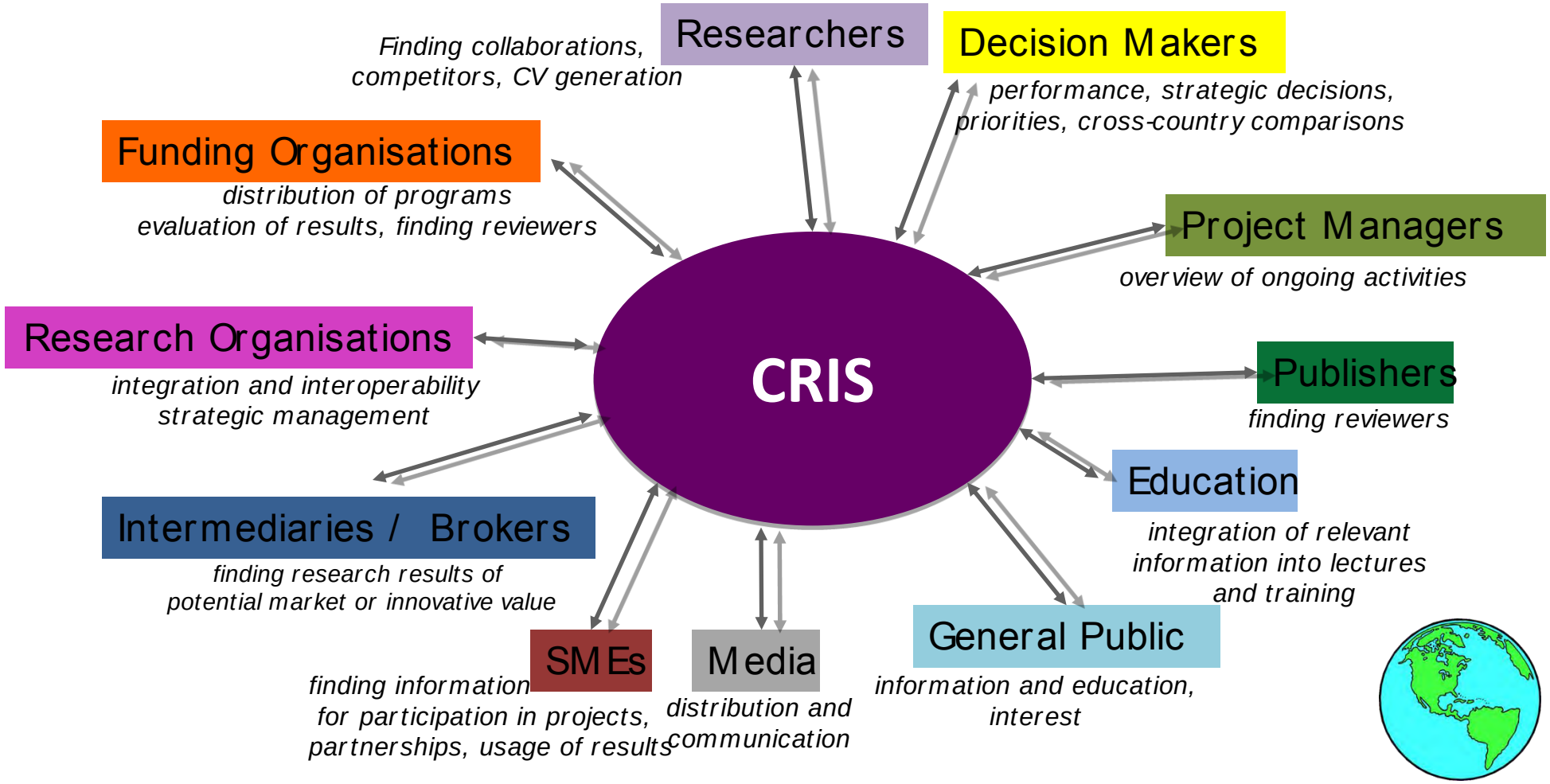
CRIS's: a bit of History

To get a proper idea about CRIS's and their function, a bit of history may be useful.

- At the end of 1980's-beginning 1990's, due to an increasing strain on the financial means for research, the question came up in some European countries whether the taxpayer's money for research was properly (effectively / efficiently) spent.
- To check this research reporting procedures were defined in these countries asking the institutions for information about their research.
- In order to register and supply the necessary information to the government, CRIS were developed and implemented at the institutions.
- In the course of the 1990's more and more countries in Europe followed in this development.

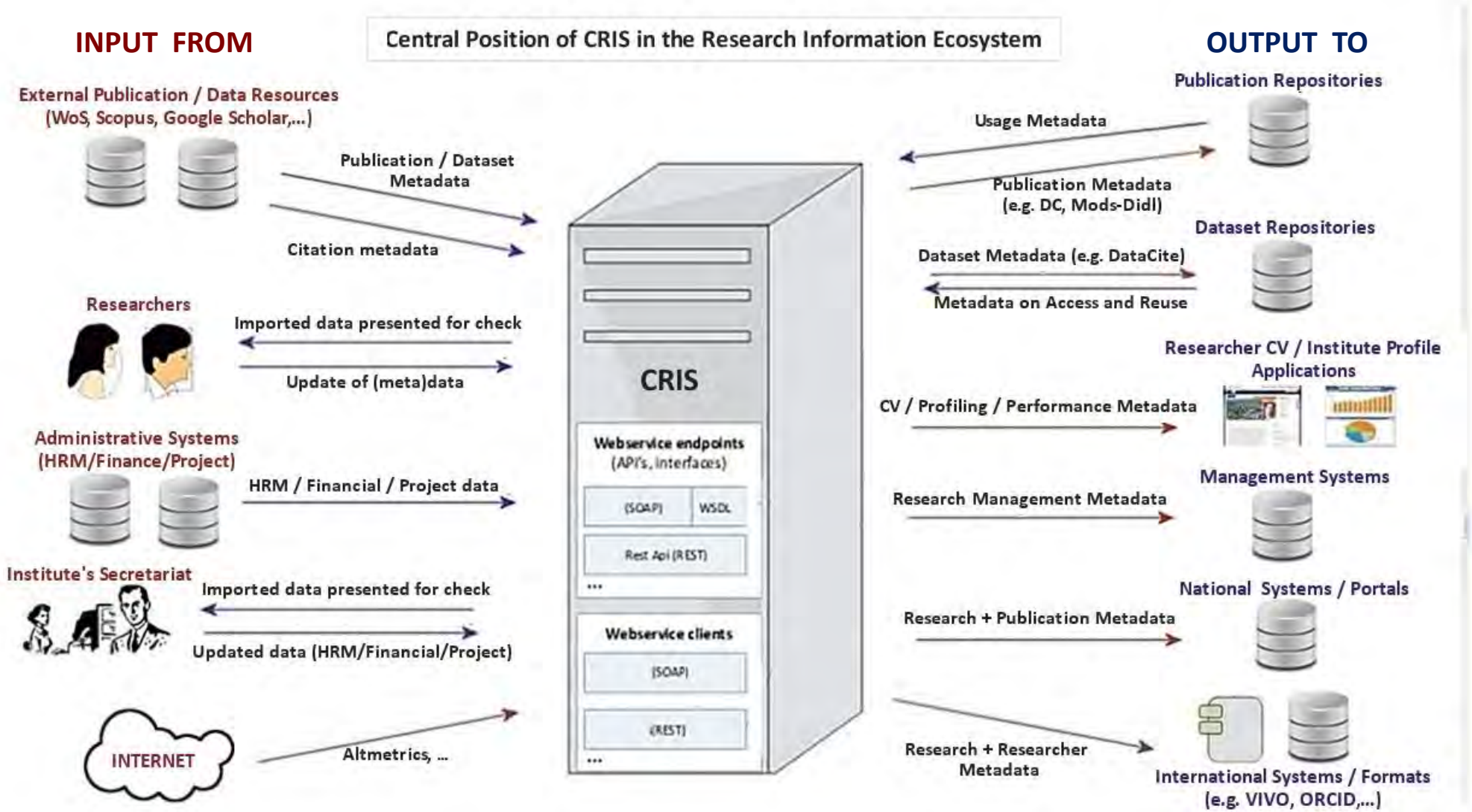
CRIS's: a bit of History

From the years 2000 on CRIS's gradually grew into multi-functional information systems not only useful for reporting of research but also providing researchers and institutes with functions for optimal communication and profiling of research to all kinds of stakeholders.



CRIS: central position in the RI ecosystem

This development of CRIS's into multifunctional sources of research information, is also reflected in the research information system's landscape: CRIS's more and more obtain a central, pivotal position in this landscape.



CRIS: central position in the RI ecosystem

Specifically mentioned can be the fact that CRIS's, given the extensive bibliographical information on publications they store, more and more become *the leading source for the (OA) publication repositories* in their institutions.

More concretely: the metadata on publications is primarily stored in the CRIS and then automatically uploaded to the repository system.

Some CRIS developers even go a step further and integrate the repository functionality in the CRIS (and so the latter replaces the repository altogether).

The euroCRIS organisation

So now that we know more about CRIS, we can get a better understanding and get into more detail on the euroCRIS organisation.

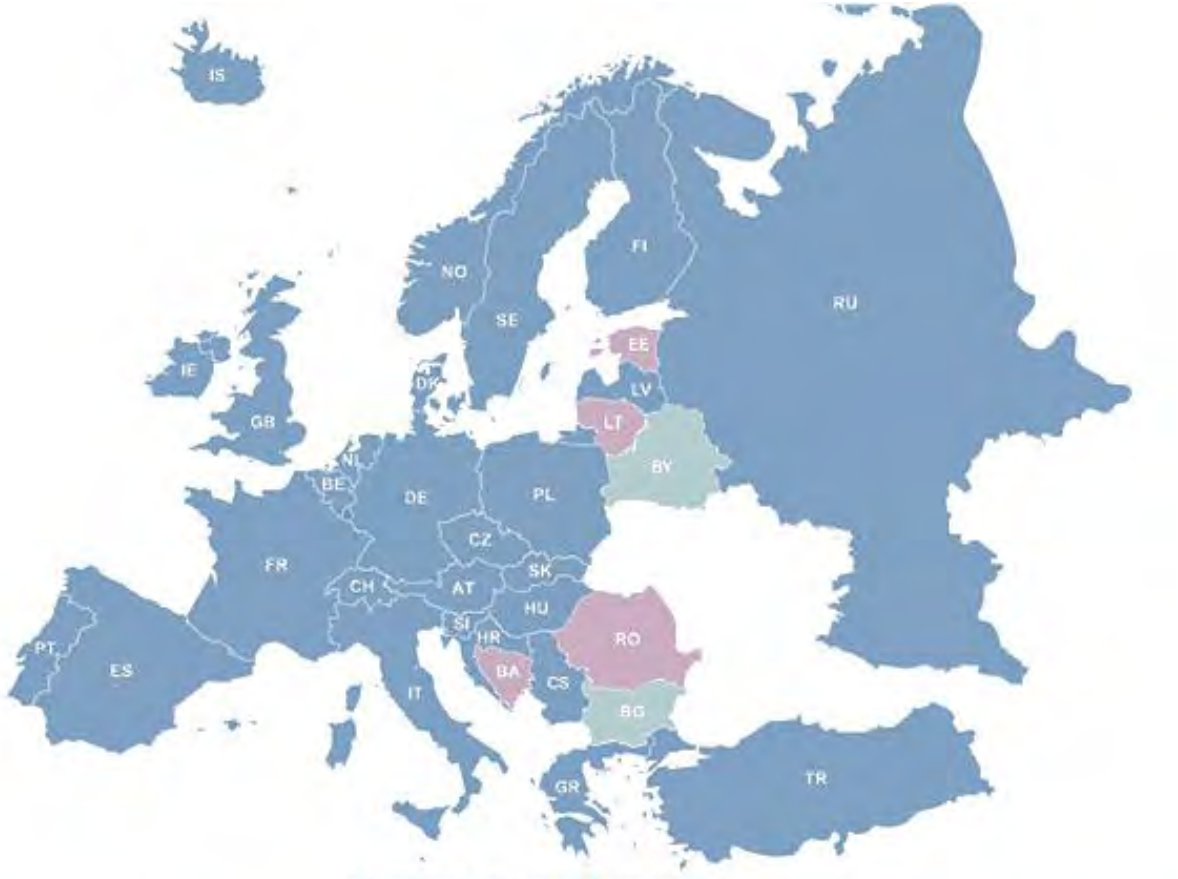
Let's again start with a bit of history:

- Beginning of 90's: research information experts from Norway and The Netherlands started up a regular communication to exchange ideas about each other's CRIS developments.
- This gradually grew into a European network and led to the establishment of euroCRIS as a formal organisation in 2002.

The euroCRIS organisation

- An *international not-for-profit association* of experts and users of research information in general and research information systems (CRIS) in particular, with offices in The Hague, The Netherlands.
- Mission: *To advance Interoperability in the Research Community through CERIF .*
- *Main activity:* development and governance of the CERIF data model and promotion of its use.
- *Other important goal:* promotion of cooperation and exchange of information and expertise between stakeholders in the RI ecosystem. To fulfill this function, euroCRIS regularly organises international events: Membership Meetings (twice a year) and Conferences (every 2 years).
- Another way of pursuing the previous mentioned goal is by *getting into Strategic Partnerships* with neighbouring organisations in the Research Information Ecosystem.
- *Members of euroCRIS are:* Universities, Research Institutes, Funding Agencies, National Research Councils, International Associations of Universities, Vendors of Research Information Systems, Publishers, ...

Membership: the euroCRIS Community



Members outside of Europe:
Australia – Canada – China – Colombia – Iran – Israel – Malaysia – Nigeria – Pakistan – South Korea - USA

Current membership is 236: 132 institutional, 79 personal and 25 affiliate
(Coming from 43 countries, mainly Europe)

Strategic Partners of euroCRIS.

The euroCRIS Board

Executive Positions:

Ed Simons (Netherlands)
President

Arjan Hogenaar (Netherlands)
Secretary

Pascale Blancquaert (Belgium)
Treasurer

Anna Clements (UK)
Executive for Strategy

David Baker (Canada)
Executive for Communication

Greta Christina Lingjærde (Norway)
Executive for Events

Valerie Brasse (France)
Executive for Projects

Michele Mennielli (Italy)
Executive for External Relations

Miguel-Angel Sicilia (Spain)
Executive for Academic Affairs

Task Group Leaders:

Jan Dvorak (Czech Republic)
Task Group Leader CERIF

Pablo de Castro (UK/Netherlands)
Task Group Leader CRIS-IR

Danica Zendulkova (Slovakia)
Task Group Leader Best Practice-DRIS

Mark Cox (UK)
Task Group Leader Indicators

Dimitris Kariaskos (Greece)
Task Group Leader Architecture & Development

The euroCRIS Task Groups

A substantial part of the work of euroCRIS is done in and by the euroCRIS Task Groups (TG's) . More important so is that these Task Groups offer the opportunity to the euroCRIS membership to concretely participate in the activities of the organisation.

In the TG's, members of euroCRIS work together on important issues concerning research information and the role of CRIS's therein.

Currently, euroCRIS has the following Task Groups:

- *CRIS Architecture and Development*
- *Best Practice/DRIS*
- *CERIF*
- *CRIS-IR*
- *Indicators*

CRIS Architecture and Development TG

The Task Group CRIS Architecture and Development focuses on those aspects of CRIS software and development processes which are suitable for sharing or reuse throughout the CRIS community.

The main priorities are the development of a reference implementation for CRIS and the definition of a standard API for programmatic access to data in CERIF-CRIS systems. Furthermore, the scope of the TG includes best practice and "soft" topics, like task analysis, software specification, selection of technologies or third party products, software architecture, user interaction patterns, management of the implementation process, and testing.

Concrete work done/in progress: development of CC-REFIM architecture and CERIF Rest API.

TG Leader: Dimitris Karaiskos (Greece) karaiskos@ekt.gr

Best Practice / DRIS Task Group

The Task Group Best Practice / DRIS connects actors in CRIS and research e-infrastructure development for getting mutual benefits.
Aims of the Task Group are:

1. *To record cases of modern practice in developing and using CRISs, including CERIF usage*, symbiosis with open access repositories, approaches to connect CRISs with research e-infrastructure, building of data and information spaces (DIS) over integrated CRISs content, and other CRIS-related innovations;

2. *To propagate recorded best practices and assist CRIS developers and users in implementing advanced concepts, design and tools.*

3. *To setup an international Directory of (existing) Research Information Systems (DRIS).*

Work in progress: (continuously) updating the DRIS and developing a “Best Practice Catalogue”. Planned: the establishment of “National Focus Points”

TG Leader: Danica Zendulková (Slovakia) danica.zendulkova@cvtisr.sk

CERIF Task Group

The CERIF Task Group maintains and further develops CERIF, the Common European Research Information Format. This also includes CERIF-XML, the exchange format, and CERIF canonical vocabularies of semantic terms.

This work keeps up a number of deliverables (specifications, schemas, scripts) and provides the basis for other euroCRIS Task Groups. The group also interacts with the CERIF user base, including vendors of CERIF-based commercial solutions, with the goal of enabling maximum interoperability of Current Research Information Systems.

Work in progress: continuous development/update of the CERIF model.

TG Leader: Jan Dvořák *jan.dvorak@infoscience.cz*

CRIS-IR Task Group

The CRIS-IR Task Group aims at furthering the science and technology of the linkage between CRIS and repositories and specifically open access institutional repositories of publications, although data and software repositories are also of interest.

This involves working on the architecture for linkage, the metadata and the mechanisms. There are outstanding issues in these areas concerned with syntax, semantics and software processes. Technical work must be preceded by communication and approximation of views between two communities: CRIS and Repository managers and specialists.

Work done/in progress: inventory of CRIS-IR relation (integration/exchange) models and practices in European countries.

TG Leader: Pablo de Castro (UK) pcastromartin@gmail.com

Indicators Task Group

Aim of the Indicators TG is to develop an active programme of research and generation of best practice (linked with TG Best Practice / DRIS) in the use of indicators (scientometrics, bibliometrics) for evaluating research.

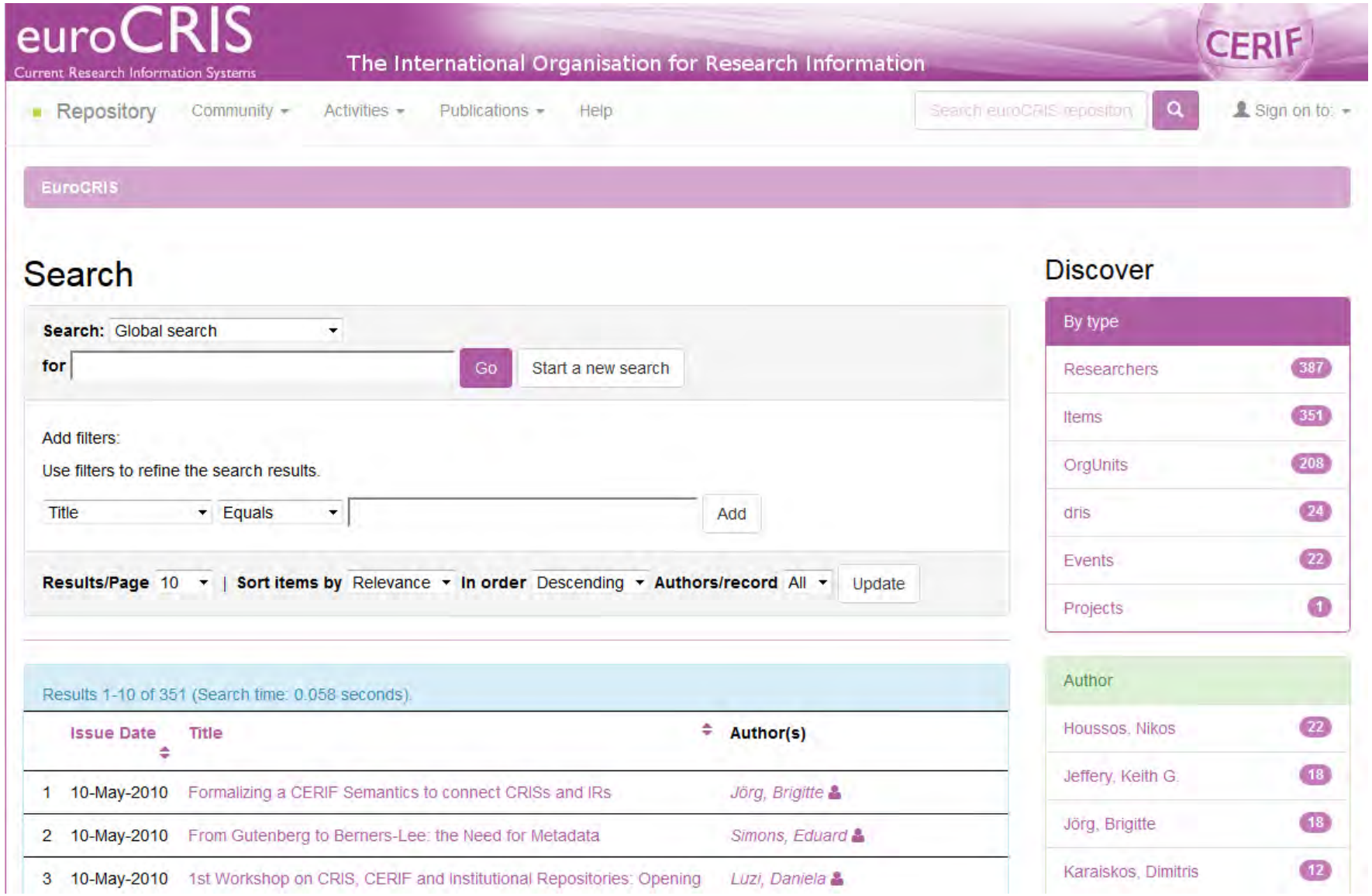
The TG will develop a catalogue of known methods with appropriate analysis of their effectiveness and efficiency. The expected output is CERIF-compliant software services (jointly with the CERIF and Architecture TG) to perform evaluation of research including for commonly used national or international methods.

Concrete work done/in progress: “CERIF-ying” of the Snowball Metrics and a “Catalogue of Indicators”

TG Leader: Mark Cox (UK) mark.cox@kcl.ac.uk

euroCRIS Web site: www.eurocris.org

euroCRIS Repository: dspacecris.eurocris.org



Joining euroCRIS

If you want to become a member of euroCRIS, click on the *"Join euroCRIS"* button on the web site:

institutional membership costs € 300,-- a year and a personal (individual) membership € 60,--

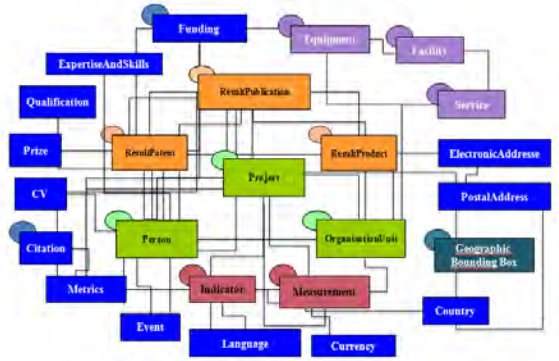
For more info on euroCRIS, please consult the web site or contact the euroCRIS Secretariat:

Secretariat euroCRIS:
Anna van Saksenlaan 51
2593 HW The Hague
The Netherlands
Tel. +31 70 349 44 50
Email: eurocris@eurocris.org

Part 2: The CERIF data model

CERIF: *"Common European Research Information Format"*

- *An international standard relational data model* for storage and interoperability of research information.
- *Official EU Recommendation* to Member States.
- *Reference model for the development of Research Information Systems (CRIS)*
- *Standard exchange format* (CERIF-XML) for interoperability between systems.



- *Strong points of CERIF*:
  - *Broad coverage:* includes all aspects of RI *(projects, persons, organisations, funding, publications, datasets, patents, products, bibliometrics, impact indicators, equipment, etc…).*
  - *Fine-grained structure and flexible architecture,* allowing:
    - In- and output of virtually any (meta)dataformat used in the RI Domain
    - The expression ("translation") of virtually any formalized use case.
    - The ingestion of an unlimited number of controlled vocabularies (taxonomies, thesauri,...) – "semantic layer".

- CERIF = comprehensive "legobox" of research information metadata.

CERIF: Basic Principles

- Only the absolute unique (unchangeable) characteristics (attributes) of an object (entity) are stored in the database with the object itself (the entity table) .

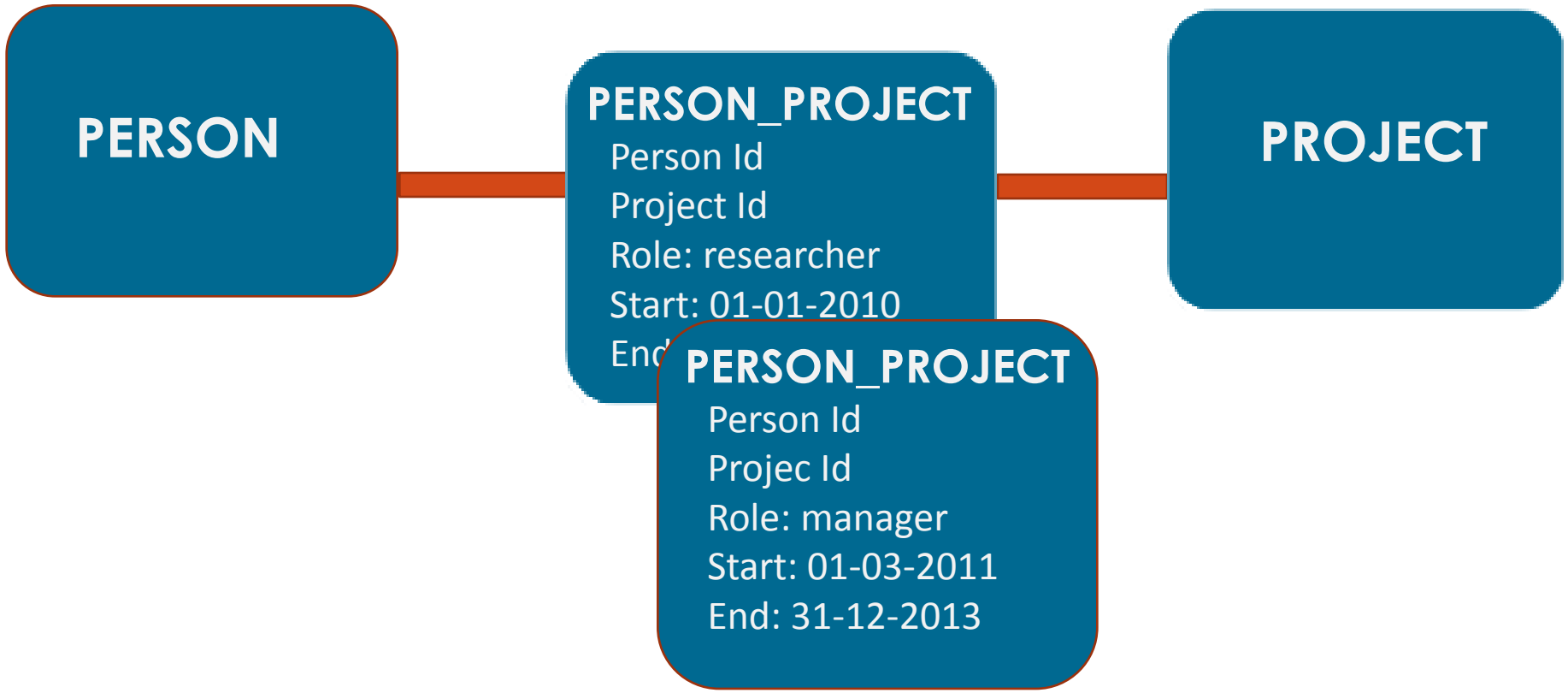
  (E.g. for the entity "Person" only a unique identifier, gender and birth date are stored with the entity itself)

- All other aspects are expressed through linking entities (link tables).
- All links have a start- and enddate and are semantically defined (meaning expressed).
- This way multiple roles/characteristics can be expressed for the same aspect (e.g. multiple roles of a researcher in a project).

Key feature of CERIF: Linking Entities

- Linking entities in CERIF are used to express semantics. They have the following functions:
  - Express the meaning of a relation between two objects, *e.g.: the role of a researcher in a project, or in a publication*.
  - Classify objects according to a given classification scheme (controlled vocabulary, thesaurus), *e.g.: classify a project or a publication according to a keyword list used in a discipline.*
  - Map various classification schemes to each other, *e.g. map the Biochemistry keyword list used by the Medical discipline to the one used by Biologists; or another example: mapping researcher classifications used in one country to the ones used in another country.*
- Linking entities have a start- and enddate, so that the exact time frame or period of each relation or classification is always known.
- An in principle endless number of linking entities (roles, typologies...) can exist between objects, *e.g: a person can be a researcher but at the same time manager of a project*.

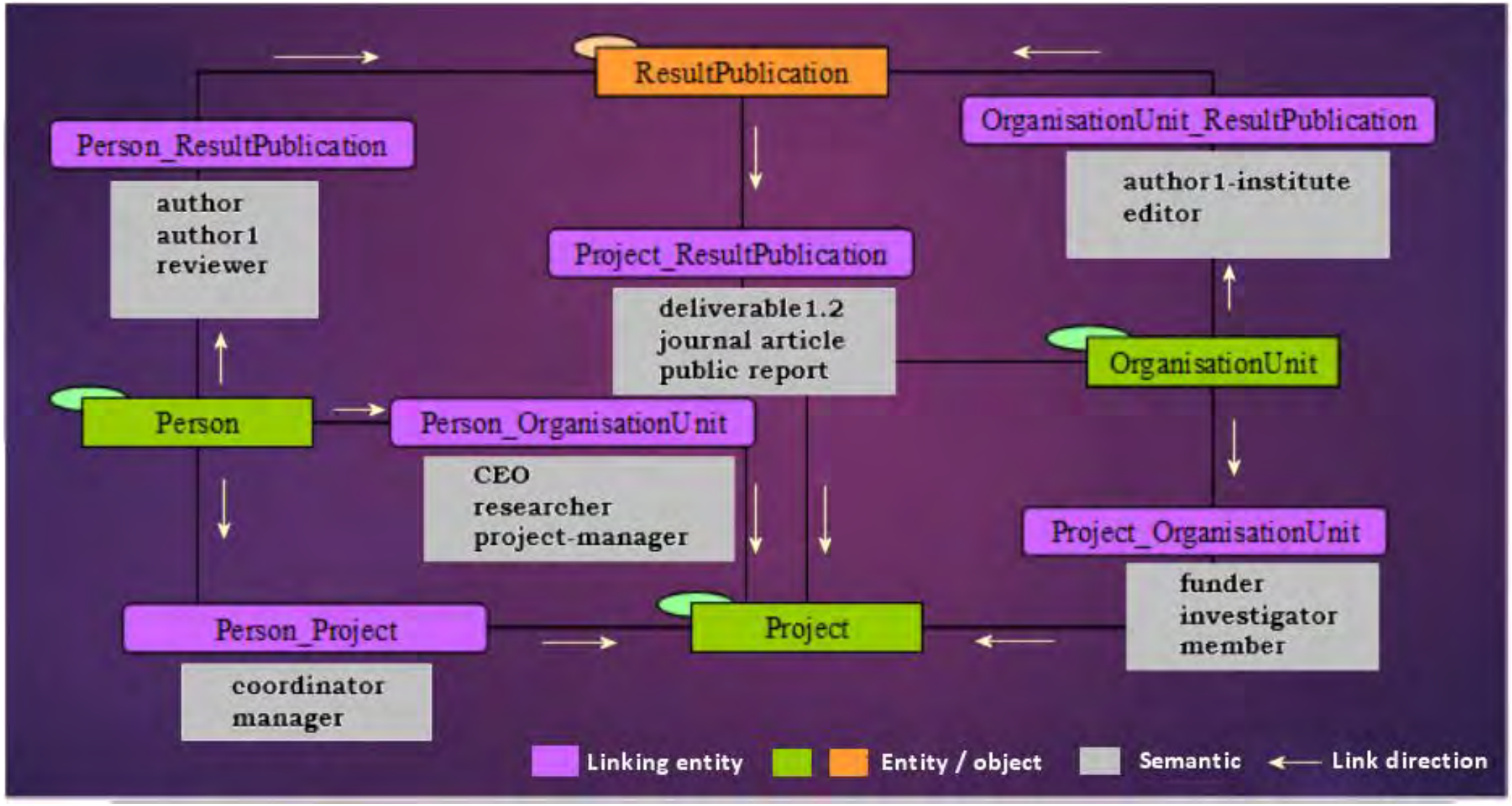
Example 1: a researcher with various roles in a project



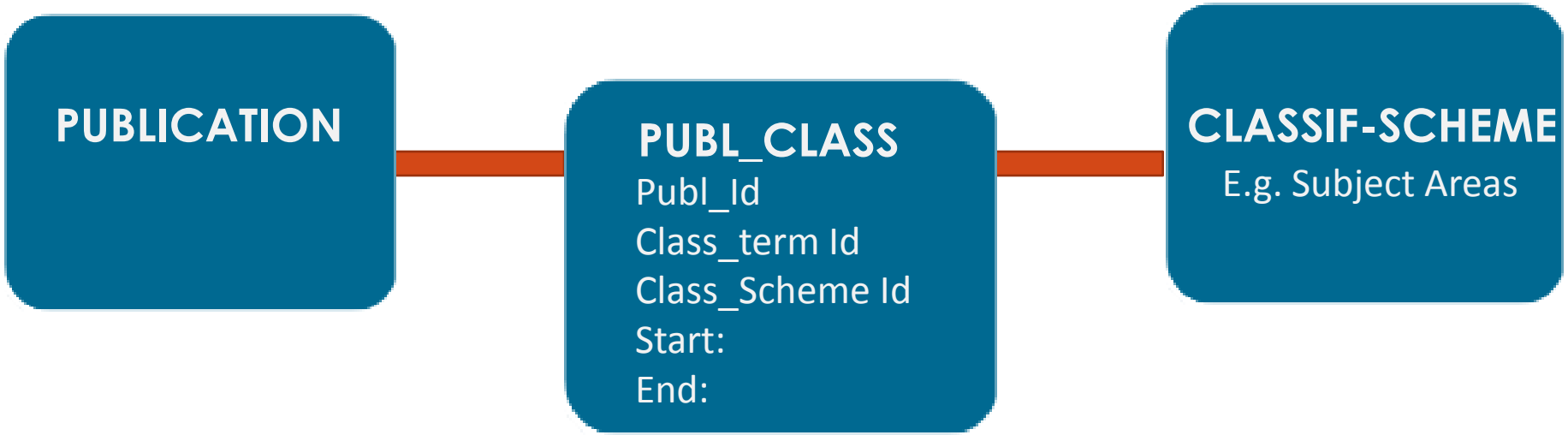
The person is researcher in the project from 2010 on and from March 2011 on he is also the manager of the project. To express this in CERIF there are two records in the "Person-Project" linking entities table, one for each role, with start- and enddate.

An extended example:

Links between Persons, Projects, Organisations and Results and their meaning (semantics).



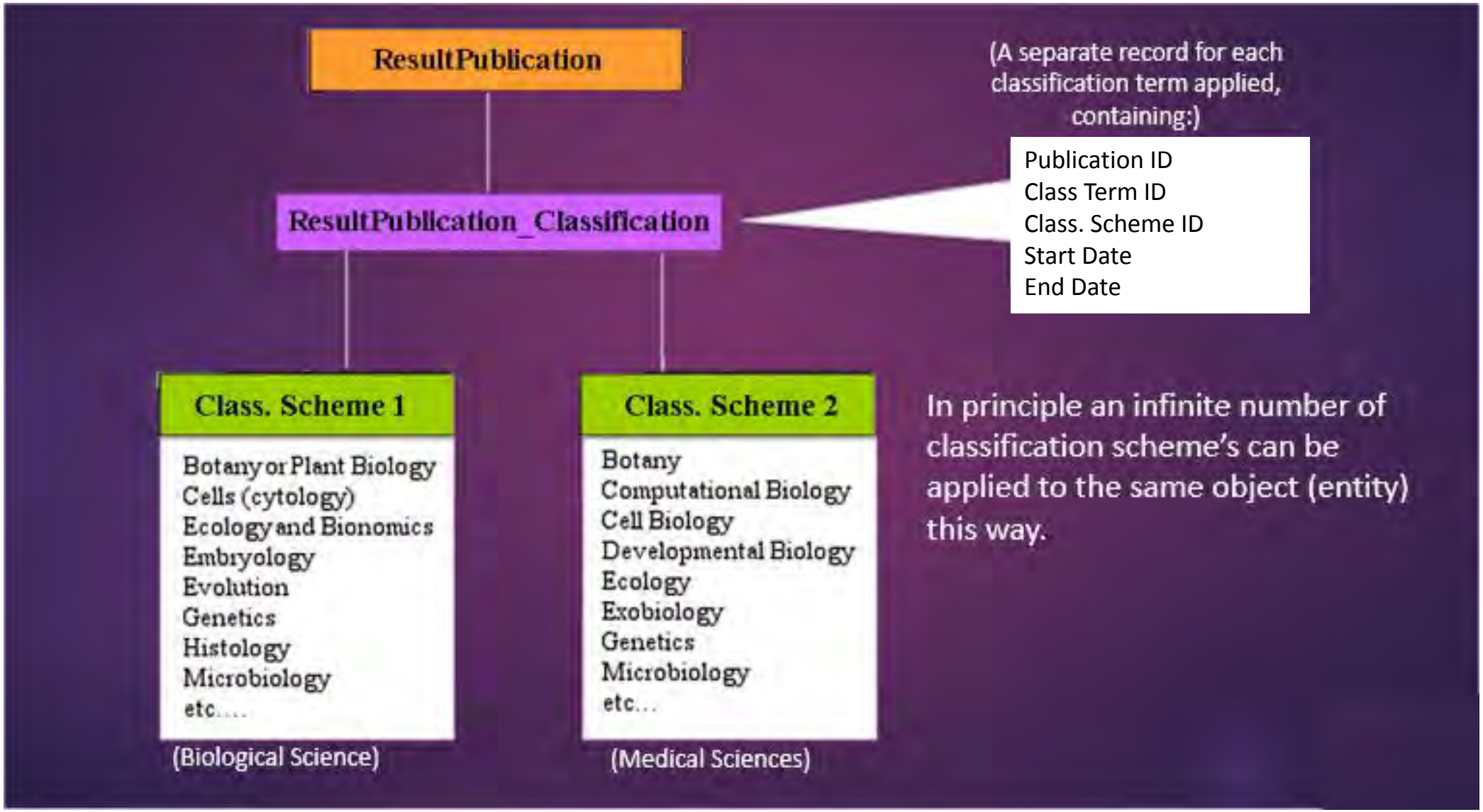
Example 2: Classifying a given publication according to the subject field or area it belongs to.



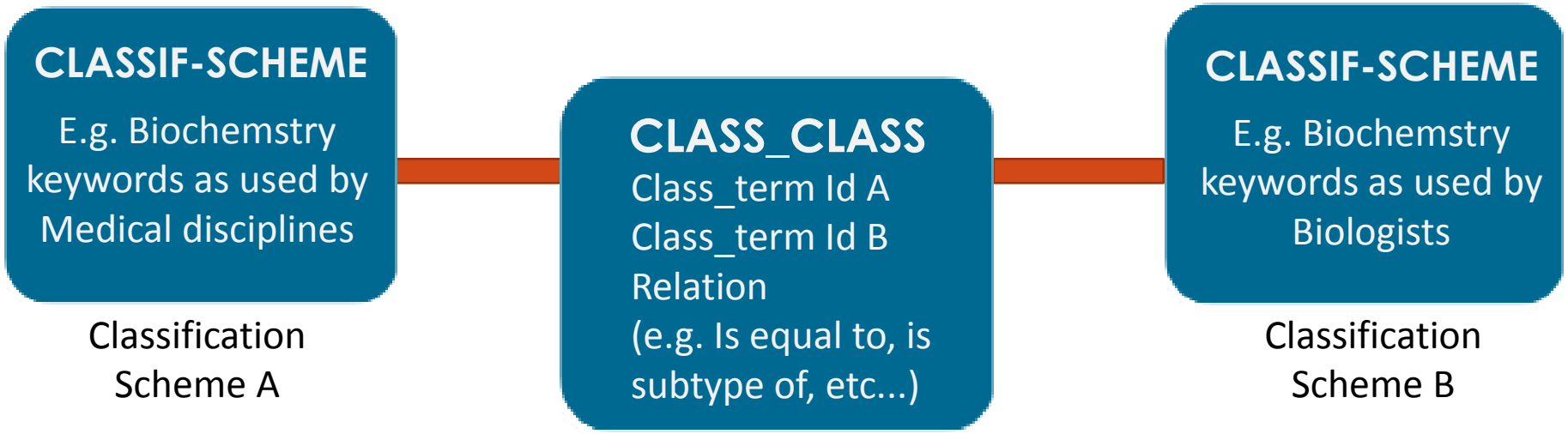
Linking entity connecting the publication to the correct subject area. Notice: also here more linking entities (records) are possible since a publication may be relevant for various disciplines or subject areas.

More concrete:

Applying multiple content classifications (from different fields) to the same publication



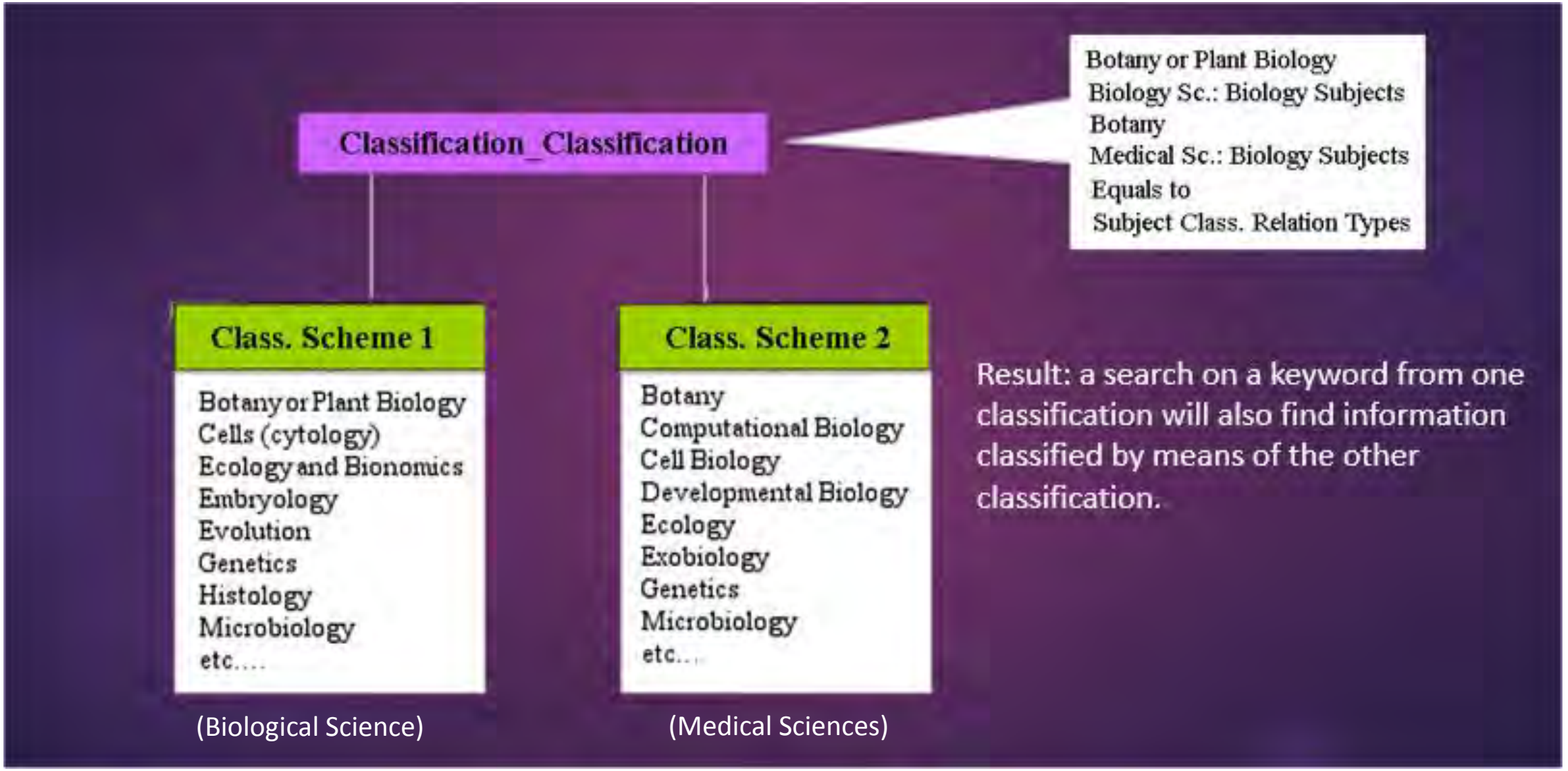
Example 3: Map various classification schemes (e.g. keyword lists)



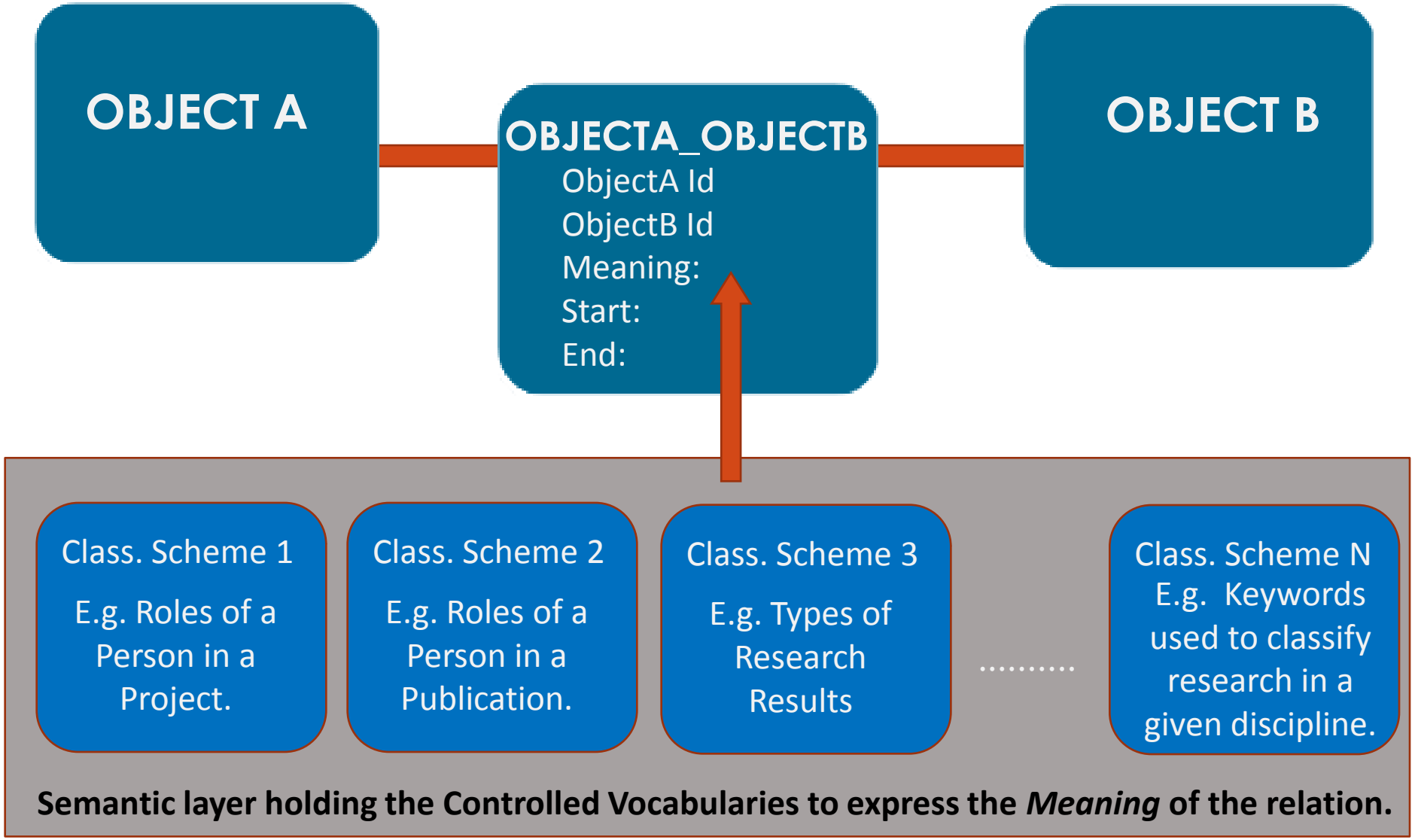
Linking table mapping a term from Scheme A to a term from Scheme B, including expression of the nature or meaning of the relation between the two terms.

More concrete:

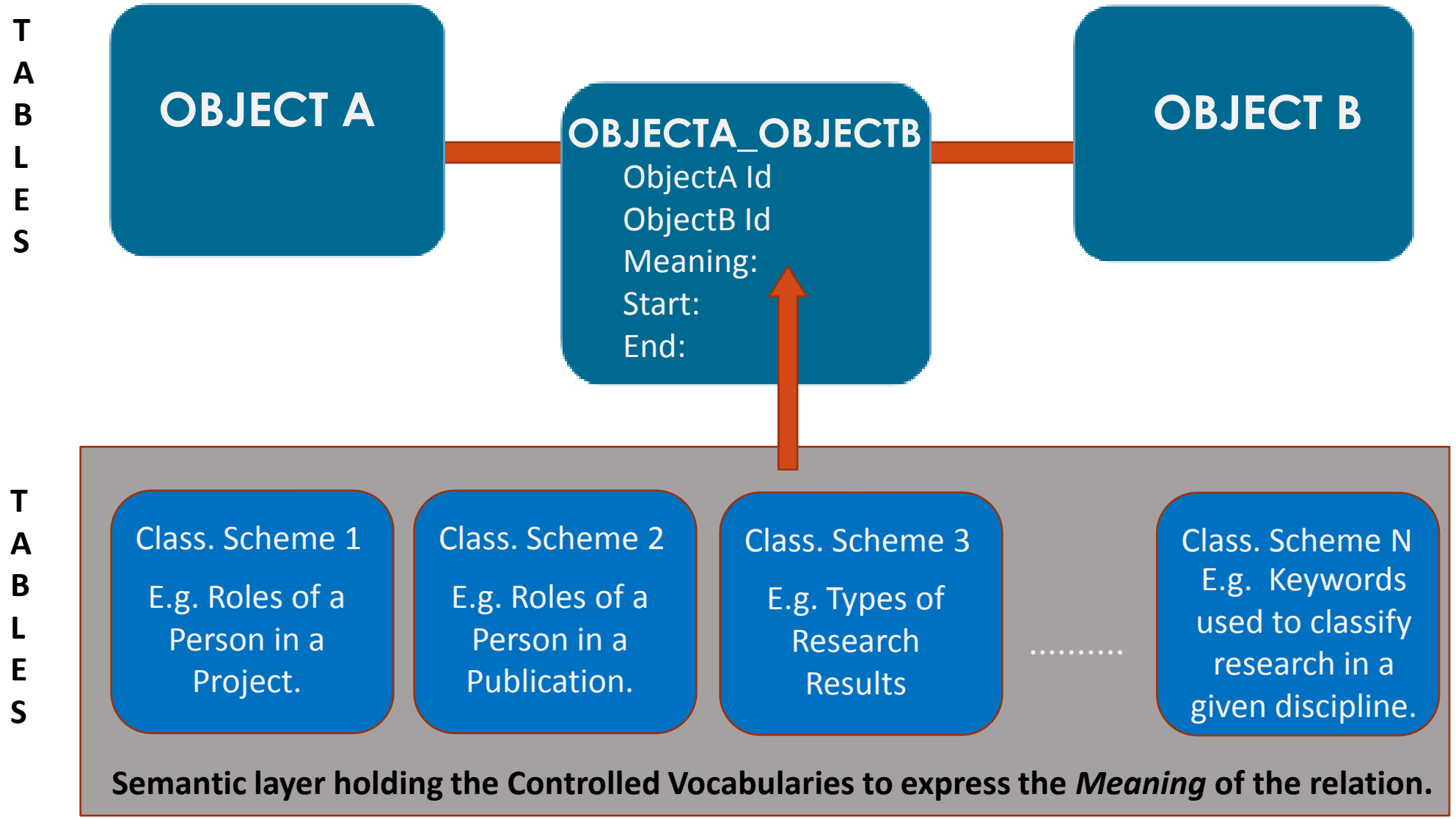
Mapping two classification schemes (controlled vocabularies) to each other.



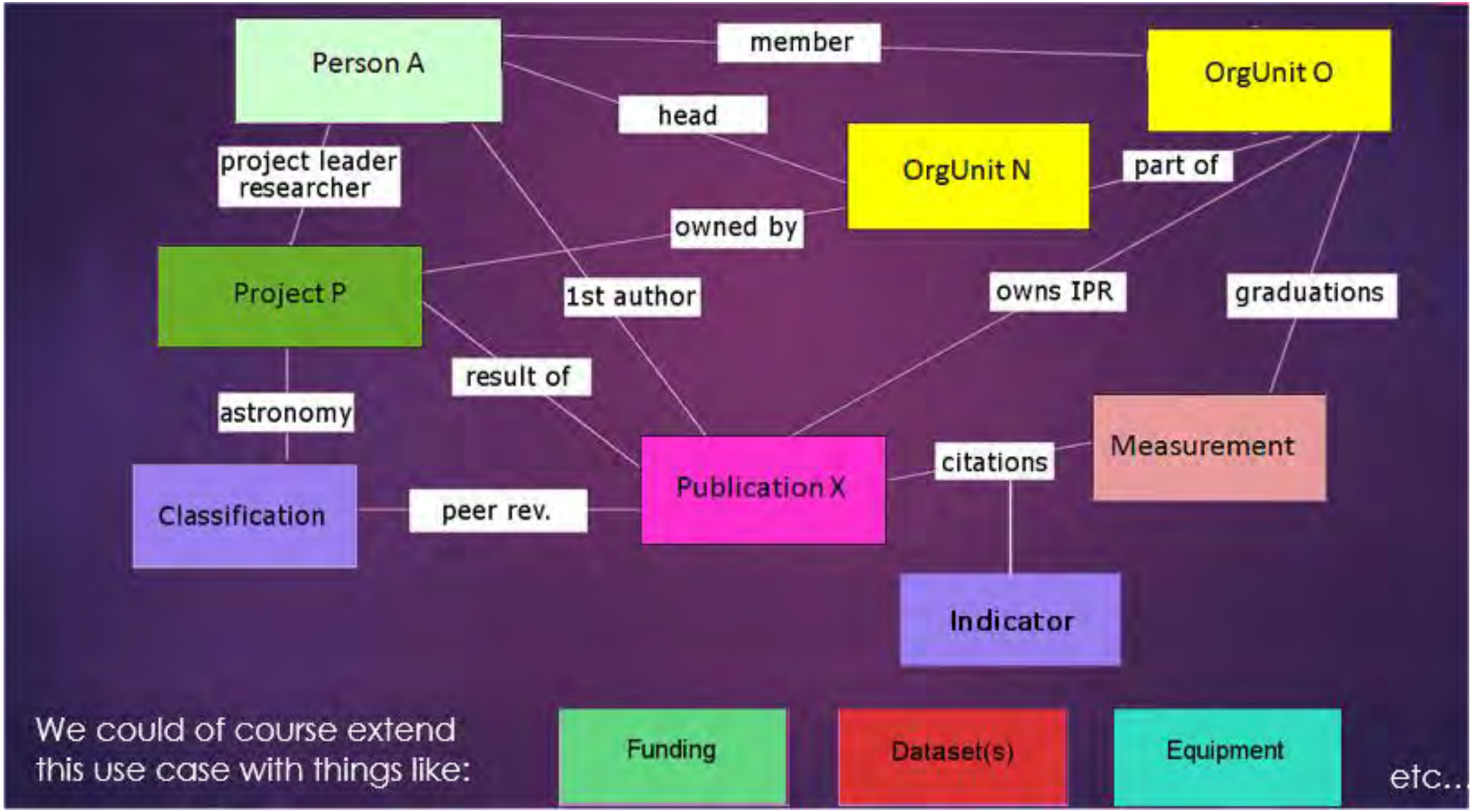
All the classifications and typologies are stored in a separate part of the CERIF model, the Semantic layer.



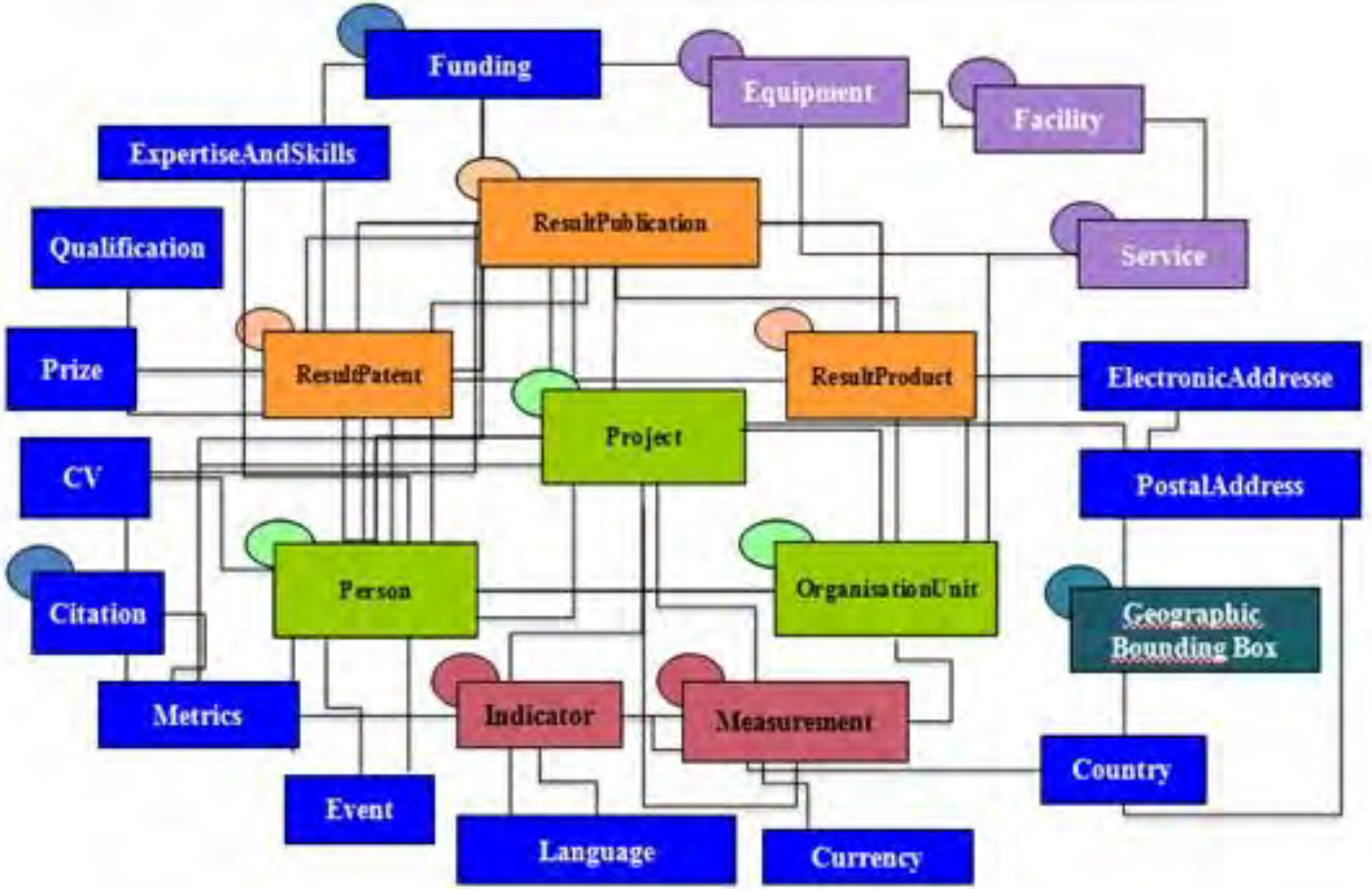
Implemented in a CRIS these are all Database Tables.



Another extended example (linking entities in white)



Summary of the CERIF Model (linking entities left out)



CRIS-implementations of CERIF

- CERIF is the inspiration or base for both commercial and non-commercial Research Information Systems (CRIS).

- Two major commercial ones being:
  - *PURE* from Elsevier
  - *CONVERIS* from Thomson-Reuters

  Together, they have a few hundred implementations at institutions mainly in Europe. Development team leaders of both systems are actively involved in the CERIF Task Group.

- But also non-commercial systems exist, e.g:.
  - *IRIS* from Cineca, the Italian university consortium, implemented at 65 Italian universities and research institutes.
  - Also from Cineca: *DSPACE-CRIS*: that extends the repository functionalities to the CRIS ones based on the CERIF data model.
  - In Greece: a national network of CERIF-based CRIS has been established.
  - The national Slovakian CRIS is fully CERIF-compatible.

Implementations of CERIF(xml) as exchange format.

- CERIF(xml) is used in various EU Projects as a standard format for exchange of information between the systems of the partners / countries involved, e.g.:
    - ENGAGE: http://www.engagedata.eu/about/
    - PASTEUR4OA: http://www.pasteur4oa.eu/
    - HOLA CLOUD: http://www.holacloud.eu/
    - OpenAire: https://www.openaire.eu/

- Examples on a national level:
    - Greece: national CRIS network based on CERIF
    - Germany: CERIF-XML being used to supply benchmark information of research institutes for the “Kerndatensatz”.
    - UK: various applications in preparation
    - The Netherlands: suppy of information from institutional systems to the national CRIS (project in preparation)

euroCRIS

Recently published

03 July 2015

OpenAIRE Guidelines for CRIS Managers 1.0

Houssos, Nikos ; Joerg, Brigitte ; Dvořák, Jan

The Guidelines specify the interoperability layer between Current Research Information Systems (CRIS) and the OpenAIRE infrastructure. The information interchange is based on the Common European Research Information Format (CERIF) data model, the CERIF XML exchange format, and the OAI-PMH protocol. The Guidelines are intended mainly for implementers and administrators of CRIS who plan to communicate research information to OpenAIRE. OpenAIRE (openaire.eu) is the European infrastructure enabling researchers to comply with the European Union requirements for Open Access to research results. OpenAIRE collects metadata from a variety of data sources: publication repositories, data archives and CRIS across Europe and beyond.

CERIF: base for a European Research Information Infrastructure

Conclusion of a report of a working group of the European Parliament (April 2014):

The feasibility of a European integrated research information infrastructure

We conclude that a European Integrated Research Information Infrastructure is technically feasible, thanks to recent technological developments and especially the maturity of the European CERIF standard, which allows seamless interlinking of datasets and/or research information systems, in different formats and including non-CERIF systems.

From the Report (page 14):

euroCRIS

Thank you very much for your attention!