



Národný informačný systém podpory výskumu a vývoja v SR
– prístup k elektronickým informačným zdrojom

NISPEZ



Informačné systémy o vede

Integrácia pre otvorený prístup k vedeckým výstupom

Zborník z medzinárodnej konferencie

CVTI SR, Bratislava 2. apríla 2014

CVTI SR, Bratislava 2014

Podporujeme výskumné aktivity na Slovensku / Projekt je financovaný zo zdrojov EÚ

EuroCRIS and CERIF. The importance of an international standard metadata model for research information



Ed Simons

Radboud University Nijmegen, Holandsko

prezident EuroCRIS

e.simons@ru.nl

pracuje ako manažér medzinárodných projektov v oblasti informačných technológií na Radboudovej Univerzite v Nijmegene, Holandsko. V roku 2013 sa po viacročnej práci v Rade stal prezidentom európskej organizácie pre medzinárodné informácie o výskume EuroCRIS.



- Ph.D. in Social Sciences
- International IT-Project Manager at Radboud University, Netherlands.
- Information Manager of IFCU: International Federation of Catholic Universities (www.fiuc.org)
- Leader of the OPUS-College Project: an open source Student Information System for universities and HEI's in developing countries (Mozambique, Zambia) (www.opus-college.org)
- Initiator and Project leader of METIS: CRIS of Dutch universities and Royal Academy of Sciences
- President of euroCRIS since 2013.

- The euroCRIS Organisation
- The CERIF Model
- The Research Information Ecosystem
- An integrated Framework for Research Information Management and Policy Making

- An international not-for-profit organisation of experts in research information and research information systems.
- MISSION: *To advance Interoperability in the Research Community through CERIF*
- Main activity: development of the CERIF metadatamodel and its application (*Common European Research Information Format*): a standard model for storage and interoperability of research information (metadata: data/information about research).
- Other function: international platform for exchange of research information knowledge and expertise.
- Bi-annual (academic) Conferences, annual Seminars and Membership Meetings (twice a year)
- Next activity:



Register at: <http://www.cris2014.org/>

Membership: the euroCRIS Community



Australia – Canada – China – Iran – Israel – Malaysia – Nigeria – Pakistan – South Korea – USA

Current membership is 194: 128 institutional, 42 personal and 24 affiliate
(Coming from 43 countries, mainly Europe)

Strategic Partners of euroCRIS.



In preparation:

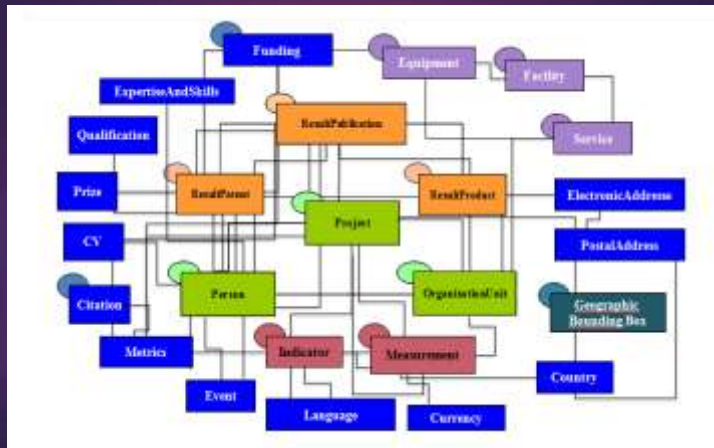
Web site: www.eurocris.org



The CERIF Model

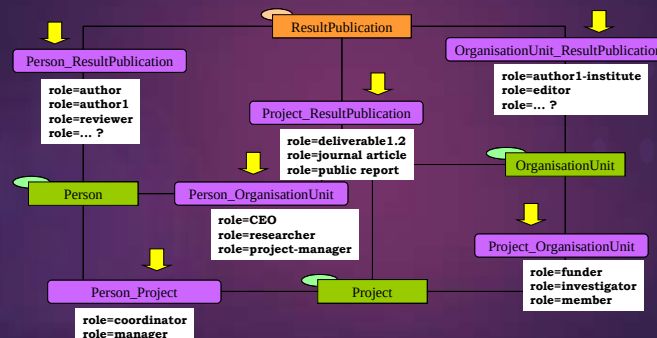
- Standard model for storage and interoperability of research information (metadata on research), and as such the base for the development of Research Information Systems (CRIS).
- Official EU Recommendation to Member States.
- Strong points of CERIF:
 - **Broad Coverage:** includes all aspects of RI (projects, persons, organisations, funding, publications, datasets, other results, impact, equipment, etc...)
 - **Fine-grained data-structure:** allowing import and export (aggregation) of almost any metadata format used in other applications (e.g. DC, Mods/DIDL, Marc21, etc..)
 - **Optimal (relational) architecture:** able to express any kind of relation between entities (by means of *link entities*), with every relation "time stamped" and semantically defined
 - **Separate "semantic layer":** allowing the use of multiple controlled vocabularies (classifications, typologies) as well as their crosswalks.
- Systems (CRIS) based on CERIF: Converis (Thomson-Reuters), PURE (Elsevier) and various locally developed systems.

The CERIF Model



Key feature of CERIF: Link Entities

- The characteristics (attributes) of an entity (object) are not stored with the entity (in the entity table) but expressed through "link entities", allowing multiple roles/characteristics to be expressed for the same aspect.

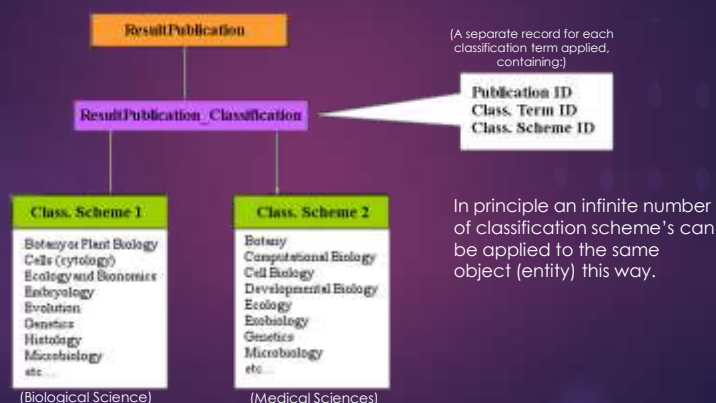


Key feature of CERIF: Semantic Layer

- Classifications of objects/entities (“controlled vocabularies”: keyword lists, typologies, etc..) are dealt with by the Semantic Layer.
- Controlled vocabularies can be both formal (e.g. typology of publications) or content-oriented (e.g. MESH: Medical Subject Header classification).
- The principle of “link entities” also applies to the semantic layer.
- This allows for:
 - Applying multiple classification scheme's to the same object, e.g.: a publication in Biochemistry could be classified both according to a classification scheme used in Medical Sciences and another one used in Chemistry.
 - Mapping (translating) different classification schemes to one another.

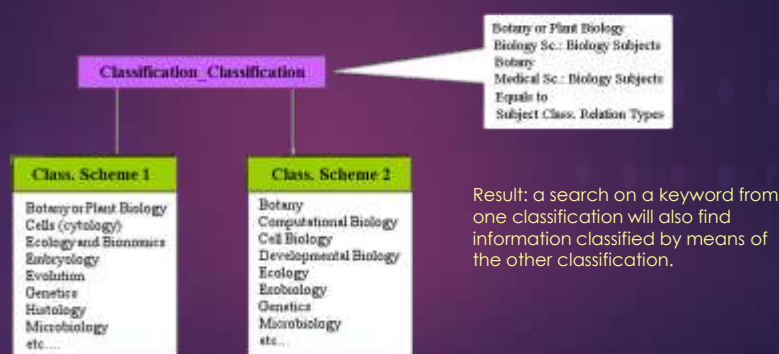
Key feature of CERIF: Semantic Layer

Example: applying multiple content classifications to a publication

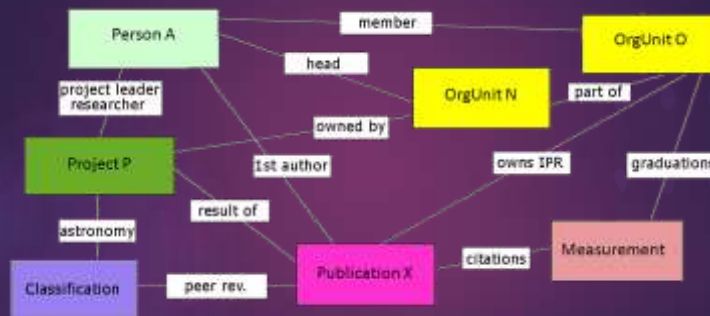


Key feature of CERIF: Semantic Layer

Example: mapping different classification scheme's to one another



CERIF: example of a Use Case



We could of course extend this use case with things like:

Funding

Dataset(s)

Equipment

etc...

The Research Information Ecosystem

3 main Components:

- **Research Management and Description Information** (contextual and descriptive metadata): stored and managed in CRIS, used for:
 - Evaluation, assessment, benchmarking of research
 - Policy making and steering of research
 - General description of research, the actors, organisations and tools involved, time and money spent for the research and the results (publications, patents, datasets) and impact obtained.
- **Electronic Publications**: stored and managed in (OA) Publication Repositories
 - Communication of the actual content or findings of the research
 - Exposure of metadata for discovery and description of the publications.
- **Datasets**: stored and managed in Data Repositories.
 - For control of the scientific integrity of the research
 - Re-use (re-analysis) of the data to find/generate additional knowledge.
 - Exposure of metadata for discovery, description and conditions of (re)use of the dataset.

Substantial overlap in metadata between the 3 components and in the systems used for these components.

This calls for an integrated approach in order to avoid the "Nightmare Scenario".

The "Nightmare Scenario"

The Nightmare Scenario is caused by a **Silo-ed existence** of the 3 components and can occur at 3 levels:

Researcher level, i.e. the **Producer** of RI: having to store substantially the same information in different systems.

- At 9:00 a.m. storing information (metadata) in the CRIS
- At 10:00 a.m. having to go to another system to store partly the same metadata in a publication repository
- At 11:00 a.m. when depositing the dataset of her/his research, again having to use a different system.

Consumer or user level: having to search or combine 3 different systems, with the risk of excluding one or two and so missing vital pieces of information.

(meta) Data Management level: risk of (de-)synchronisation of information, both in time and concerning uniformity of information.

Adding Complexity: External Systems

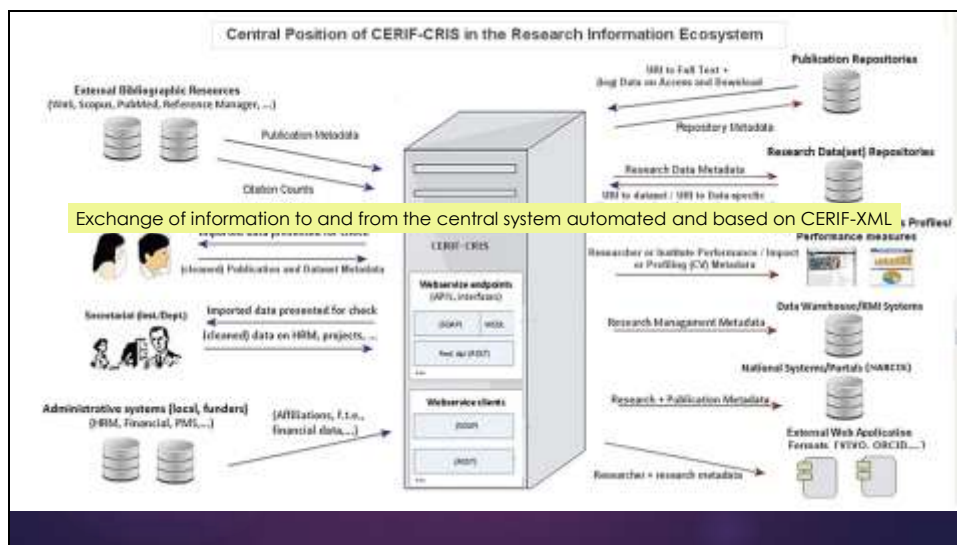
- Information on publications often already available in international (commercial) systems, e.g. Web of Science, Scopus.
- Information on persons, projects, etc... often already available in systems outside the research information domain, i.e. HRM-systems, Project management systems.
- This can add to the “Nightmare Scenario” if not properly taken into account.

Summarizing the problem

So we have 3 components of Research Information, managed by different systems but to a large extent using the same metadata.
Adding to this: a part of the metadata needed is already present in “external systems”.

In order to avoid extra (double) work for researchers and data managers as well as risks concerning synchronisation, uniformity and completeness of information we need an integrated approach or framework for RI metadata handling.

The CERIF Model and CRIS based upon this model, constitute the core element, or central piece, of such a framework.



Why can CERIF perform this pivotal function?

- **Broad Coverage:** includes all aspects of RI (projects, persons, organisations, funding, publications, datasets, other results, impact, equipment, etc...)
- **Fine-grained data-structure:** allowing import and export (aggregation) of almost any metadata format used in other applications (e.g. DC, Mods/DIDL, Marc21, etc...)
- **Optimal (relational) architecture:** able to express any kind of relation between entities with every relation "time stamped" and semantically defined (the core concept of *link entities*)
- **Separate "semantic layer":** allowing the use of multiple controlled vocabularies (classifications, typologies) as well as their crosswalks.

More in detail: the model applied to Datasets



(© Keith G. Jeffery)

Concrete application of the 3-layer Model

- Suppose: Marine Biology research into the water quality in a certain region of the Indian Ocean and its effects on life/quality of the Coral Reefs in that region.
- Water samples are taken regularly in various parts of this region.
- In **layer 2**, the central contextual/descriptive layer, metadata are stored such as:
 - unique identifier of the dataset,
 - Name/title of the dataset,
 - Language
 - Dataset locator (URI),
 - Conditions for access and re-use
 - Possible restrictions for public access
 - Institution which conducted the research
 - (names, titles, roles, etc... of) Researchers involved
 - Responsible person and contact person for the dataset
 - Project as part of which the research was carried out
 - Publications based upon/linked to the dataset
 - Geographical coordinates of the of the Indian Ocean region the research applied to
 - Pointer (URI) to the data-specific metadata in layer 3, etc...
- From layer 2 the discovery metadata (e.g. DC) for layer 1 are automatically generated and exposed on the Internet.
- In layer 3 the discipline- or data-specific metadata are stored, e.g. the geographical coordinates of the specific parts of the region, specific classification schemes of the coral flora and fauna, chemical procedures applied for analysis etc...
- For each layer specific services may be implemented, e.g.: harvesting services on layer 1, dataset profile creation or linked data services on layer 2, integrity control services on layer 3, etc...

Other Key Elements of the Framework

- **Unique and persistent, standardized identifiers**
(DOI, ORCID, ISNI, ...): euroCRIS cooperates with ORCID
- **Standardized business vocabularies:**
for this euroCRIS cooperates with CASRAI

Both aspects, together with the CERIF model, are indispensable in order to guarantee **interoperability** in research information.

Conclusions: Summarizing the Framework

- The work of euroCRIS, in connection with its Strategic Partners, supplies the building blocks for an optimal, integrated Framework for **Research Information management and interoperability** on an institutional, national or international level.
- The framework consists of the following key elements:
 - The CERIF standard metadata model as the central core or layer of the framework.
 - Standardized unique identifiers.
 - Standardized business vocabularies.

Conclusions: Summarizing the Framework

- The framework forms the base for:
 - Formulation of an **integrated Research Information Policy** and Strategy (**policy level**)
 - Realisation of **optimal technical solutions** for storage, interoperability and exposure of RI (**system level**)
 - Implementation of **appropriate organisational structures** for RI management, both functional and technical (**organisation level**)
 - Creation of **optimal services** for the various target groups (consumers/users) of research information. (**service level**)

To summarize the goal of the Framework in one sentence:

Minimizing the overhead and workload needed for input and maintenance of Research Information while at the same time *maximizing services and functionality* for the users of Research Information.

- euroCRIS, together with its strategical partners, e.g. CASRAI, ORCID, can lend advice and support on how to optimally implement the integrated framework for research information on an institutional (universities, institutes), national (governmental organizations, e.g. SCSTI) and international (e.g. EU) level, or even provide “turn key” solutions in the future.

Thank you very much for your attention and I hope to see you next month at the international euroCRIS Conference in Rome!



You can still register at: <http://www.cris2014.org/>