Open Science and availability of information: The role of CRISs

Ed Simons (ed.simons@ru.nl)
euroCRIS / Radboud University
BE-OPEN Workshop
Leiden, May 18th 2017
Speaker

Currently:
- President of euroCRIS, the *International Organisation for Research Information*.
- Senior Project Manager Research Data Services, *Radboud University*, The Netherlands.

Formerly:
- Head of the Central IT Support Unit at *Radboud University*.
- Information Manager of the *International Federation of Catholic Universities (IFCU)*.
- Project Leader *International Association of University Presidents (IAUP)*
- International Project Leader, IT Development Projects for *NUFFIC, the Dutch Organisation for University Development Cooperation*.
Core idea underlying the presentation

*Open Science requires optimal availability of information about the scientific research, its results (publications, datasets,...) and the people (researchers) and organisations (institutes, universities, networks, funders,...) involved.*

(current) Research Information Systems (CRIS) are key resources in this respect.
What is euroCRIS?

- An international not-for-profit association of experts and users of research information and research information systems (CRIS).

**Main activities:**
- Development and governance of CERIF, a standard data model and exchange format for research information, and promotion of its use.
- Promotion of community building and cooperation between stakeholders (organisations) in the Research Information Domain. To fulfill this function, euroCRIS regularly organises international events: Membership Meetings and Conferences.

**Upcoming events:**
- Strategic Membership Meeting, Bratislava, October 2017.
euroCRIS: Membership

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

200+ members from 45 countries (mainly Europe)
## euroCRIS: Membership

### Members

<table>
<thead>
<tr>
<th>Name of Institution</th>
<th>Type of Membership</th>
<th>Country</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>CyberLeninka</td>
<td>institutional</td>
<td>Russia</td>
<td><a href="http://cyberleninka.ru/">http://cyberleninka.ru/</a></td>
</tr>
<tr>
<td>LabHUB.ru</td>
<td>institutional</td>
<td>Russia</td>
<td>LabHUB.ru</td>
</tr>
<tr>
<td>University of Novi Sad, Faculty of Sciences</td>
<td>institutional</td>
<td>Serbia</td>
<td><a href="http://www.pmf.uns.ac.rs">http://www.pmf.uns.ac.rs</a></td>
</tr>
<tr>
<td>Laboratory for ideas - IDEALAB</td>
<td>institutional</td>
<td>Serbia</td>
<td><a href="http://www.idealab.rs">www.idealab.rs</a></td>
</tr>
<tr>
<td>University of Novi Sad</td>
<td>institutional</td>
<td>Serbia</td>
<td><a href="https://www.uns.ac.rs/index.php/en/">https://www.uns.ac.rs/index.php/en/</a></td>
</tr>
<tr>
<td>Slovak Centre of Scientific and Technical Information (CVTI SR)</td>
<td>institutional</td>
<td>Slovakia</td>
<td><a href="http://www.cvtsr.sk">http://www.cvtsr.sk</a></td>
</tr>
<tr>
<td>Institute of Information Science (IZUM)</td>
<td>institutional</td>
<td>Slovenia</td>
<td><a href="http://www.izum.si">http://www.izum.si</a></td>
</tr>
<tr>
<td>Slovenian Research Agency (ARRS)</td>
<td>institutional</td>
<td>Slovenia</td>
<td><a href="http://www.arrs.gov.si">http://www.arrs.gov.si</a></td>
</tr>
<tr>
<td>Josef Stefan Institute (JSI), Center for Knowledge Transfer</td>
<td>institutional</td>
<td>Slovenia</td>
<td><a href="http://ct3.is.si">http://ct3.is.si</a></td>
</tr>
</tbody>
</table>
euroCRIS: Strategic Partners

In preparation:

DURASPACE

OCLC
Recent News
Videos and presentations Strategic Membership Meeting are online
12-19-2016
In November 2016 euroCRIS organized the Strategic Membership Meeting (SMM) in close collaboration with the National Documentation Centre (EKT/NHRI), the host for this meeting.

Strategic Membership Meeting, November 8-10, 2016, Athens, Greece

The videos, presentations and selected photos of the SMM are available. They can be found on the EKT webpage of the euroCRIS event:
http://www.ekt.gr/el/events/program/20010 and on http://www.ekt.gr/el/events/photos/20023

New Science Europe Position Paper on Research Information Systems
11-15-2016
At the euroCRIS Strategic Membership Meeting Dr Emily

Mission
"To advance interoperability in the research community through CERIF"

Contact us
Secretariat euroCRIS:
Anna van Saksenlaan 51
2592 HW Den Haag
The Netherlands
Tel. +31 70 340 44 50
Email: eurocris@eurocris.org
CERIF: euroCRIS’s main “product”

An international standard relational data model for storage and interoperability of research information

Standard exchange format (CERIF-XML) for interoperability between systems

Official EU Recommendation to Member States

Reference model for development of Research Information Systems (CRIS)
CERIF: euroCRIS’s main “product”
What is a CRIS?

A CRIS is an information system that holds data about virtually all aspects of research:

• *The researchers* (ID, name, title, affiliation, ...) and their role.
• *The projects* (ID, title, description, key words, start- and enddate, ...)
• *The organisations* (institutes, universities,...) involved and their role (owner, funder, ...)
• *The input* invested in the research both in time (f.t.e.) and money.
• *The results* from the research (publications, datasets, software, patents, images, etc...)
• The *equipment and services* used.
• *Cooperations* with other projects, researchers, groups, organisations.
• *Links to other systems* (HRM, financial, external: ORCID, WoS, Scopus, ...)
• The *domain or subject area* of the research belongs to.
• The *impact of the research*, both in- and outside of academia (metrics, impact indicators,...).
• Semantic *classifications of the research* on various dimensions (typologies, controlled vocabularies ...),
• *Rights metadata*: who is authorised or which conditions may apply to access (some of the) information in the CRIS
• Etc...

And all these aspects related or linked to each other!
What is a CRIS: a bit of history

• CRIS’s first appeared around the end of the 80’s, early 90’s, closely linked to emerging policies of research assessment and control within some European countries (e.g. The Netherlands, Norway, Denmark,…).

• So initially they more or less uniquely had an administrative purpose and were also as such conceived by the research community.

• In the course of time, and especially the last decade, however the image has changed and CRIS nowadays are evolving into multifunctional instruments for a multitude of stakeholders among which research managers and, not the least, the researchers themselves.
Functionality of CRIS

CRIS: Multifunctional Information Resources

Researchers:
- Finding possible collaborations
- CV generation, profiling of research
- Performance, strategic decisions, priorities, cross-country comparisons

Decision Makers:
- Finding possible collaborations
- CV generation, profiling of research
- Performance, strategic decisions, priorities, cross-country comparisons

Project Managers:
- Overview of ongoing activities
- Finding reviewers

Publishers:
- Finding reviewers

Teaching Staff:
- Integration of relevant information into lectures and training

Intermediaries / Brokers:
- Finding research results of potential market or innovative value

Research Organizations:
- Integration and interoperability, strategic management, profiling

Funding Organizations:
- Distribution of programs, evaluation of results, finding reviewers

SMEs:
- Finding information for participation in projects, partnerships, usage of results

Media:
- Distribution and communication

General Public:
- Information and education, interest
Functionality of CRIS

Central Position of CRIS in the Research Information Landscape

**INPUT FROM**
- External Publication / Data Resources
  - (WoS, Scopus, Google Scholar...)
- Administrative Resources
  - (HRM, Finance, Project Man.)
- Inst. Secretariat

**OUTPUT TO**
- Publication Repositories
- Dataset Repositories / Archives
- Profiling & Management Applications
- ((inter)national) Research Portals
- Other RIS-systems / formats
- ORCID
- VIVO

**Internet**
Functionality of CRIS

Awareness is growing that, because of the broad coverage of – and at the same time interlinked – aspects of research information, CRISs are not only of relevance for administrative and management purposes but also, and even more so that

CRISs can be considered as primary resources for (communication of) information on and profiling of research and scientific activities as such.
CRIS: Instrument for Profiling of research(ers)
M.J.C.M. (Marno) Verbeek
Professor of Finance, Scientific Director
ERIM, Dean of Research RSM

Marno Verbeek is a Professor of Finance at Rotterdam School of Management, Erasmus University (RSM). He is also Dean of Research of RSM and Academic Director of the Erasmus Research Institute of Management (ERIM).

His recent research is largely in the area of empirical finance with a particular focus on mutual funds, hedge funds, asset pricing, investment strategies, survival bias and performance evaluation.


He received his PhD from Tilburg University in 1991.

- **Work in Progress**

- **Publications (53)**

  **Key Publications (30)**

**Centres**
- Erasmus Finance Group
- Mutual Funds and Hedge Funds

**Work in progress**

**Latest publication**
CRIS: new field of application: research data sets

- Data collection
- Selection
- Analysis
- Temporary storage
- Data life cycle management
- Articles
- Final dataset
- CRIS: linking publications to dataset + archiving of data
- Publication of results
- CRIS: registration of (metadata) for dataset
CRIS: new field of application: research data sets

The RIS/RDS project at Radboud University in cooperation with DANS (national data hosting organisation in NL)
CRIS: basic resources for an Open Science Infrastructure

Because CRIS systems:

- Hold a detailed set of information (metadata) on basically all objects and attributes related to research, and

- All the information elements in a CRIS are related or interlinked to one another, and therefore

- All these different pieces of information can be used as entries for queries to find and interpret (whether it is of relevance) research and directly also find (all information on) related research.

A network or infrastructure of interlinked CRIS systems would be a powerful if not indispensable part of an Open Science Infrastructure.
CRIS: basic resources for an Open Science Infrastructure

More specifically: a CRIS-based information infrastructure would be of great relevance and added value in order to realise a real FAIR-data infrastructure.

**F**: Findability (given the rich set of interlinked metadata in CRIS)

**A**: Accessibility (given the ID’s and the rights information in CRIS)

**I**: Interoperability (given the ID’s, the standard format and interlinked semantics in a CRIS)

  - Interpretability (given the rich semantics – description, controlled vocabularies - in a CRIS)

**R**: Reusability (given the info on used equipment and tooling – e.g. software in a CRIS)
Interoperability: a basic condition for success

In order to be able to function as a basic information layer for Open Science

**Interoperability**

*based on standards*

between the information resources in the RI Infrastructure is a *conditio sine qua non.*
Aspects hampering interoperability

Lack of standard, shared vocabularies (terms) and semantics (meaning of terms).
Aspects hampering interoperability

Solution: use of standard, shared, concepts and semantics.
Aspects hampering interoperability

Solution: implementation of international standard identifiers for ALL data objects in the Research Information Domain (e.g. ORCiD, ISNI, Researcher ID, DOI, ...)

No unique identification of data elements
Aspects hampering interoperability

Matilda and Bob want their systems to communicate but it isn't happening...

No standard exchange format
Aspects hampering interoperability

No standard exchange Format
Aspects hampering interoperability

Matilda swaps her circle system for a new star system... and they can still interoperate.

CERIF ↔ CERIF

It still looks the same to me!!
Conclusion: the 3 Pillars of Interoperability

- **Standard Vocabularies** *(shared definitions of terms)*.

- **Standard Identifiers** *(unique and persistent)*.

- **Standard Exchange Format** *(enough complete, fine-grained and flexible)*.
The 3 pillars of interoperability: a challenge

Avoiding to uniquely or too much focusing on one aspect – Identifiers – and forgetting or underestimating the other two, because identifiers alone will not bring a complete solution.
Interoperability: positive developments

- Both standards concerning the various aspects of interoperability as well as organisations governing these standards have matured and are now in place.

- **Definitions**: e.g. CASRAI-vocabularies and use cases (they call “profiles”).
- **Identifiers**, e.g.: ORCiD, ISNI, DOI,....
- **Exchange format**, e.g.: CERIF (euroCRIS)

There is a willingness and commitment to collaborate between these standards organisations.
Interoperability: positive developments

- Communities within the research information domain, e.g. CRIS and Repository communities are more and more approaching each other and working together.

- There is a growing awareness on the policy and strategy level both on a national and international level that interoperability and therefore standardisation is necessary.
Communities working together

- The last few years various Memoranda of Understanding have been concluded between organisations representing various stakeholder groups within the research information domain (e.g. Between euroCRIS and COAR, EARMA, VIVO, etc...)

- Last November a basic agreement has been reached between CASRAI, EARMA, euroCRIS, JISC, OpenAIRE, ORCID and ALLEA to work together in the promotion and realisation of a European Research Information Infrastructure.
Communities working together

- Also on a more practical level, e.g.:
  - The joint definition by OpenAIRE and euroCRIS of a subset of the CERIF-format to harvest information for OpenAIRE from CRIS-systems.
  - The collaboration between VIVO and euroCRIS to create a mapping between the CERIF-model and the VIVO-ontology.
  - The Indicators and Metrics community working together with euroCRIS in the “Snowball Metrics” initiative
Growing awareness on the policy level

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

“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”

(page 14 of the Report)
Growing awareness on the policy level

Science Europe Position Statement
On Research Information Systems
November 2016

Promote the adoption of research information systems with the following characteristics:

- Economic value-creation for service providers that does not depend on siloing and closing data;
- No pay-walling of research activity data, given that dataset combination is a basic requirement to understand the research landscape; and
- Interoperability based on a common exchange format and open standards for the definitions of the entities (such as researchers, organisations, grants, activities, and outputs) and attributes in the research information domain, and for the identification of these entities with unique and persistent identifiers.
A concrete example: relevance for the EOSC

• An optimal European Research Information Infrastructure, based on standards, is a necessary, complementary, part of the European Open Science Cloud (EOSC), indispensable to optimally realise the FAIR-data aspect.

• Without such an information infrastructure, providing the needed metadata about the research data in the Cloud, the EOSC will function suboptimally.
The key player: the researcher

• A Research Information Infrastructure is worthless without sufficient (complete, detailed enough) information being registered in the resources that constitute the infrastructure.

• For this “we” are dependent on the researcher: they hold the information and should (be willing to) register it completely, correctly and in time in the information systems.

• This is not always obvious, since there are some “obstacles” in the way here......
Some obstacles from the researcher’s point of view

- Registration of the same information is requested multiple times by different parties (funders, publishers, assessment bodies, ....).

- With different levels of detail.

- With different definitions of concepts and classification schema’s (semantics).

- By means of totally different (looking) interfaces.
A “killer” for the Researcher’s motivation

To:

• At 9 o’clock have to go to the interface of a funder to fill in information.

• At 10 o’clock going to a different interface of an assessment body to fill in (partly) the same information.

• At 11 o’clock again having to fill in this information in the interface of the local information system of the institute.

A Key principle: never force the researcher to again register information that already has been registered previously.
A “cultural” obstacle: the researcher’s mindset

Researcher: the key player

“Research activities

“My World”

Research information activities

“Their (administration) World”
Changing the researcher’s mindset

Research activities

Research information activities

“My World”
The DRE-project at Radboud University (DRE: Digital Research Environment)

Initially the DRE-project basically focused on tools for the research process

(1) Artificial examples, not an existing DRE.
The DRE-project at Radboud University

But it changed into a broader perspective, including research information aspects

(1) Artificial examples, not an existing DRE.
Take home messages

• Open Science assumes the availability of optimal (meaning enough complete and detailed) information on research.

• Therefore, a Research Information Infrastructure (RII) is to be considered a vital, underlying and inextricable part of an Open Science Infrastructure, a.o.t. to guarantee the FAIR-aspect.

• In order for such a RII to function optimally, interoperability between the resources in the infrastructure is necessary, this means: the use of standard and persistent identifiers, standard vocabularies, and a standard exchange format.

• CRISs are to be considered (the) primary resources in the information infrastructure, due to their broad and detailed coverage of aspects and the available, “pre-defined” interlinkage between the information elements in a CRIS.

• A big advantage of CRISs is the accountability of the information: data registered in a CRIS are being controlled and checked by the institution that manages the CRIS. As such a CRIS-based Research Information Infrastructure has a kind of built-in trustworthiness that is difficult if not impossible to achieve in the concept of a self-regulating scientific community.

• The researcher (motivation, commitment) is the key.
Thank you very much for your attention!

Questions?