Research Information Management Systems: covering the whole research lifecycle

Pablo de Castro[1,3] and Hanna-Mari Puuska[2,3]
[1] University of Strathclyde, UK
[2] CSC – IT Center for Science Ltd., Finland
[3] Vereniging euroCRIS, Netherlands
pcastromartin@gmail.com, hanna-mari.puuska@csc.fi

Abstract

Since the joint EUNIS/euroCRIS CRIS/IR survey report was published in 2016, Current Research Information Systems (CRISs) have become even more widely implemented at institutions and in countries all over the world. These platforms traditionally serve a double purpose: as tools to showcase the research activity conducted within its walls and as internal databases to enable evidence-based decision-making processes. Increasingly, they also act as a reliable and comprehensive information source for external systems and services. This text provides a snapshot of the current CRIS landscape – with an emphasis on Europe – and summarises the various ways CRISs are currently used as well as introduces new emerging uses and scenarios. Special attention is paid to the ever-growing number of national and regional CRIS platforms, which are increasingly seen as valuable research information collection systems for the analysis of the strengths and weaknesses of research in specific geographies.

Keywords: research information management, CRIS, information systems, interoperability, research lifecycle, responsible research evaluation

1. Background

A joint EUNIS/euroCRIS "CRIS/IR Survey Report" was published by the EUNIS Research and Analysis Initiative (ERAI) in March 2016 (Ribeiro, Mennielli, De Castro, 2016). The report, based on a survey conducted in 2015 that received 84 full responses from 20 different countries, was aimed at answering two main questions, namely whether CRISs were gradually replacing institutional repositories (IRs) and whether both types of systems were overlapping in their functionalities. The results of the exercise showed that the answer to both questions was negative: CRISs and repositories were found to be complementary; while IRs were the preferred choice for managing research outputs,

CRISs were regularly used to manage the institutional research information as a whole including metadata for research papers.

This paper is a follow-up to that exploration of the research information management landscape conducted eight years ago. The role of repositories has now significantly widened especially with regard to research data management and preservation, but these systems essentially maintain the function of managing (institutional) research outputs that was highlighted in the 2015 report and will not be analysed in any detail in this paper. The authors will instead focus on providing an updated snapshot of the much-evolved CRIS infrastructure with some emphasis on the ever-more widespread national CRIS. Same as for repositories, the functions CRIS are fulfilling these days have also greatly expanded. Some examples of this array of functions will be shown below.

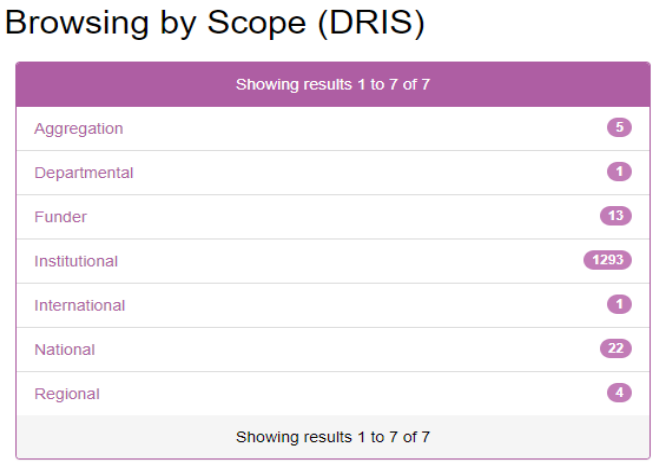
Current Research Information Systems (CRIS), also known as Research Information Management Systems (RIMs) are described in Wikipedia (Wikipedia, 2023) as:

> *Databases aimed at collecting, storing, and exchanging information on all aspects of the research activity conducted at an institution (or a region or country, sometimes on a specific discipline) or funded by a research funder.*

CRISs have been around for decades – one of the first institutional CRISs to be implemented, the METIS system at Radboud University Nijmegen, started operating as early as 1993 (Simons, 2004). However, it is only in the last fifteen or so years that these CRISs have become widely implemented at institutions and in countries all over the world. The euroCRIS Directory of Research Information Systems (DRIS), with over 1,300 entries at the time of writing (euroCRIS, 2023), provides ample evidence for this steep growth in the number of CRIS solutions at an institutional, regional, national, and funder level.

The euroCRIS DRIS does not include every single research information system presently available worldwide – the information collection depends on the awareness of the directory by the organisations operating such systems and on their willingness to share the details of their solutions with euroCRIS. The majority of the entries recorded in DRIS come from Europe or Australasia. While not a fully accurate sample geographically, the DRIS provides some general-purpose snapshot of the current CRIS landscape. The vast majority of recorded CRISs (96%) are institutional ones, with 1,293 out of a total of 1,347 DRIS entries. On top of this, 30 systems in the directory are classified as supra-institutional solutions: 22 national, 4 regional, 5 aggregation and 1 international CRIS (Table 1). Commercial or other off-the-shelf products dominate the field, with only 6% of the entries reported are built in-house.

Table 1. The scope of CRISs recorded in euroCRIS Directory of Research Information Systems (DRIS).



The analysis of the software solutions implemented by the research information systems in the DRIS shows a wide range of coexisting approaches. Platforms provided by commercial vendors are very frequent, with Pure being the most widely implemented commercial system. In-house-built solutions are also prominent in this classification, in fact, outweighing commercial systems if we include in this category nationally developed systems widely implemented in specific countries like the Norwegian CRIStin, the Italian IRIS, or the Indian IRINS. Finally, there is also a significant presence of open-source software solutions like DSpace-CRIS and VIVO.

2. Different functions of CRISs

Traditionally, CRISs may serve at least two different – and to some extent contradictory – functions at an organisation or at a national or regional level: first, in what could be called the **research portal function**, they aim to showcase the research activity conducted in it by providing public access to entities such as researchers, research groups, the organisational hierarchy, funded projects, research equipment and facilities and research results (publications, datasets, patents, dissertations, software, etc). Research portals are typically used for finding research results or exploring scholars in a research field interesting to the user.

Secondly, in fulfilling a **data warehouse function**, CRISs collect and store all this information in order to support the internal evidence-based decision-making processes or monitoring research performance in an institution, department, field of science, or country. These processes usually include an economic element related to (among others) investment policies or staff hiring and promotion. Frequent indicators used for these decision-making processes include the income raised via funded projects by research groups, departments, or disciplines, the number of research results (meaning publications but also research data and patents for instance) produced by specific units or persons or the number of collaborations with industry by economic sectors. The research information elements supporting these processes tend to be kept internal and are typically not openly shared.

However, to an increasing extent, the information maintained in CRISs is also used in other systems, either with the consent of the researcher or to fulfill statutory obligations. Meaning, as an emerging third main function, CRISs also act as **an information source** for other systems and services that need high-quality machine-readable information on research and researchers. For example, a researcher can automatically update their own data from their home organisation's CRIS to their own ORCID profile (www.orcid.org), or HEIs can provide their data to, for example, research funders' systems.

As part of their research portal function, CRISs are often key resources for the implementation of Open Science. For this purpose, institutional CRISs tend to be integrated with institutional repositories, with the metadata for the institutional publications recorded in the CRIS and linked to the full text offered from the repository. This complementary role was actually the main finding of the EUNIS/euroCRIS survey on CRIS/IR interoperability conducted in 2015, and the situation around interoperability remains essentially unchanged.

CRISs also provide a comprehensive database for monitoring the progress in Open Access publishing and open research data, allowing data to be displayed on the share of open publications and data for faculties or departments or across disciplines. The recording or harvesting of metadata on research data by many CRISs provides a key support mechanism for the FAIR principles: Findable, Accessible, Interoperable, and Reusable (GO FAIR, 2023). CRISs make research data easier to find and by following a uniform standard for metadata assignment, they make metadata well accessible.

A study on publication databases in social sciences and humanities (SSH) identified 21 national databases in Europe (Sîle et al, 2017). These are operated either as centralised systems into which data is entered directly by different organisations, or as data aggregations from institutional CRISs, funders' research information systems, and other local systems. Originally, publication output was used in many countries as a criterion in performance-based funding models (Sivertsen, 2016). In the past years, national systems have been developed to cover a wider range of research outputs and activities, and in addition to statistics, monitoring, and funding allocation, they have started to fulfill new uses, such as providing a single access point to information on researchers and research results in the country and acting as a so-called hub that gathers information from several sources and provides it for multiple uses. In addition to SSH fields, the majority of these national databases cover all other scientific disciplines and, besides publications, many of them include information on also other research outputs and researchers.

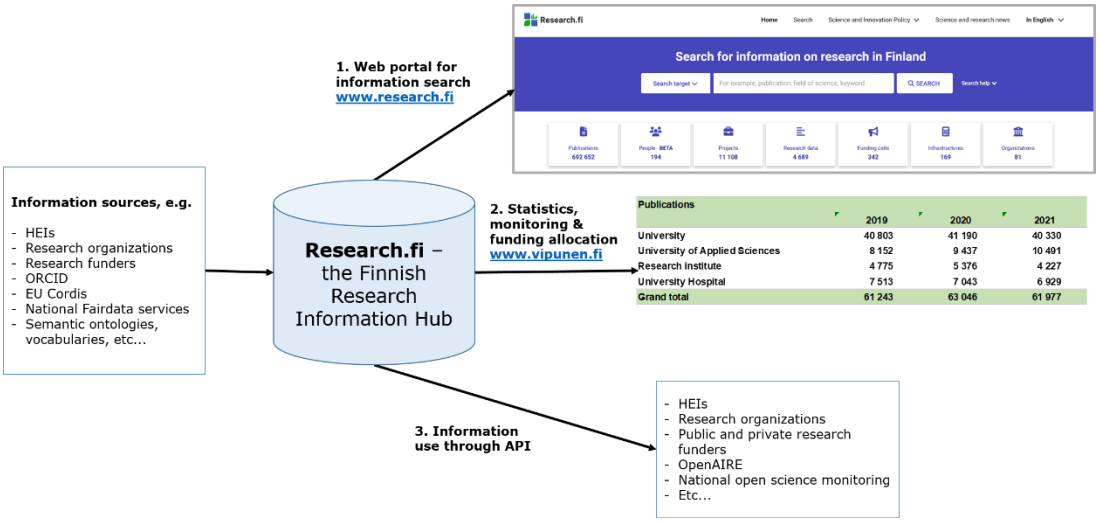


Figure 1: Research.fi serves as the Finnish research information hub by aggregating information from various sources to national level and providing them to various purposes of use.

The Finnish Research.fi is an example of a national CRIS based on an aggregating service that combines information sources and provides the ensuing data collection for various uses (Figure 1). For example, researchers get their publications and other activities automatically reported to funders' systems and the funder does not need to integrate their system with all the institutional systems available at universities but gets information from all universities via the API for the national Research.fi CRIS (Puuska, 2019).

As shown in the analysis of the euroCRIS DRIS above, institutional CRIS solutions are often based on large, international products such as Pure, Esploro, Converis and others, while national CRISs are often in-house-built. On top of this, the CRIS landscape is becoming increasingly country-specific, with numerous software solutions available on a national- or occasionally language-based approach. Examples of such specific solutions in Europe are among others Worktribe and Haplo in the UK, HISInOne-RES in Germany, and SIGMA Research and DialnetCRIS in Spain.

The key concept when trying to capture a national-level research information snapshot is **system interoperability**. It does not really matter what software solution a specific institution uses as long as the (meta)data can be exchanged and aggregated on a national-level platform such as Research.fi. The Common European Research Information Format (CERIF) maintained by euroCRIS and endorsed by the European Commission is the standard approach used by CRISs to harmonise the research information they collect so that it may be exchanged across systems.

3. Role of CRISs in a developing research information management landscape

On the one hand, there is a need for as broad as possible an information base that allows a researcher or research group to easily transfer information on their research activities from one system to another. In this sense, there is rapid development in harvesting and crawling systems, such as OpenAIRE and Dimensions. On the other hand, there is a need for reliable and standardised (meta)data for information management, research evaluation and monitoring where it is critical that the information is collected in a uniform and precisely defined manner and it is necessary to be able to verify the criteria by which the data has been collected and how representative it is.

3.1. Interoperability and comprehensiveness characterise CRISs as unique information bases

It is typical for CRISs both at an institutional and a national level that the metadata and classifications are curated so that the research information is described in a uniform way. Standard data models like CERIF aim to precisely define what can be recorded in the systems and what kind of relationships may be established across entities, for example, the kind of relationship an author has with an institution. This makes CRISs a unique high-quality data source for institutional or national level analysis, statistical reports, and monitoring but also as a source for other harvesting systems.

CRISs also represent a key information source for research knowledge graphs which compile information on different research outputs and activities, such as researchers, organisations and their subunits, publications, projects, data, and infrastructures. All these entities should ideally be interlinked with each other through persistent identifiers (PIDs). An ongoing project, FAIRCORE4EOSC, develops core components for the European Open Science Cloud (EOSC) with the aim of better interoperability between different research-related services. The components being developed in the project aggregate research outputs through APIs and data dumps, support mapping different data models and enable

conversions of metadata between them. CRISs play a central role in the mutual exchange of information across institutional, national and European levels. CRISs not just deliver their information but can also use the core components to access information available in the European Open Science Cloud (EOSC) and thereby improve the contents of the research graphs maintained in different systems. The CERIF data model maintained by euroCRIS is in a crucial role as a shared model for exchanging this information. (Suominen, 2022)

3.2. New uses of CRISs – supporting the whole research project lifecycle

There is a wide range of software solutions at institutions aimed to cater to the whole research project lifecycle. Current Research Information Systems (CRIS) as defined by euroCRIS is the standard name that these systems are given but some of these solutions actually extend beyond the – usually outward-looking – the concept of CRIS and expand their range of functionalities.

Relatively recent additional modules devoted to for instance Ethics, Pre-award, Post-award, or Costing tend to be categorised as part of the Research Management and Administration (RMA) domain rather than within the area of Research Information Management (RIM), but all these systems actually provide a continuum along the whole project lifecycle from the call for project proposals by a specific research funder to the assessment of the project impact once it has been completed and all its outputs have been produced, identified and linked to the project. This whole project lifecycle support is achieved by means of a collection of modules exchanging research information metadata with each other from very early on.

A joint session held by euroCRIS and the European Association of Research Managers and Administrators (EARMA) in February 2021 (euroCRIS, 2021) devoted to exploring the possible RMA vs RIM dichotomy concluded that there is no such thing. The session discussions showed that it is much more practical to examine RMA and RIM from the perspective of the project lifecycle, which is a single concept covering both internal and outward-oriented information.

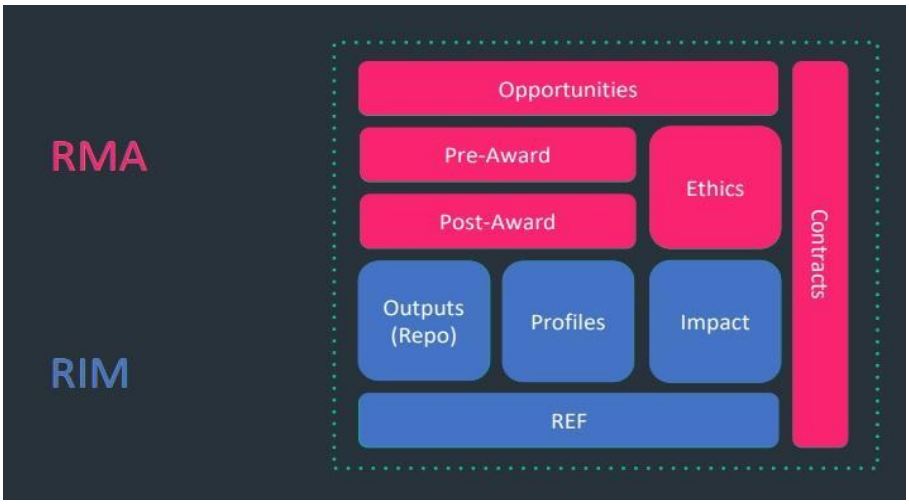


Figure 2: Modular structure for a system serving the whole project lifecycle, from the call for project proposals ("Opportunities") to the research assessment (Research Excellence Framework or "REF" in the UK). Modules in pink colour are closed (i.e. internal to institutions) while those in blue provide openly available information

Software solutions in the research information management domain offered by vendors do in fact increasingly cover both the research portal and the internal information storage functions – see an example of such an approach in Figure 2 above. They tend to do this via a modular architecture where some modules are openly available to the outside world whereas other ones – such as pre-award, post-award, costing, or ethics – are kept internal.

3.3. Responsible research evaluation and the diversity of research output

Researchers are evaluated at all stages in their careers, for example, when they apply for a position or when their personal performance is assessed in salary negotiations, crediting, or other rewards. Researchers or research groups are also evaluated when applying for funding and sometimes as part of comprehensive research assessment exercises of universities or research organisations (TSV, 2020).

Special attention is currently being paid in the research information management domain to how it supports responsible research evaluation. As of January 2023, the Agreement on Reforming Research Assessment issued by the Coalition for Advancing Research Assessment (CoARA, 2022) had been already signed by 455 European universities, research funders, learned societies, and national agencies that have committed to keep the focus on qualitative evaluation with the support of responsible quantitative indicators, to respect the variety of disciplines, and to avoid inappropriate metrics and rankings. The agreement also highlights the recognition of the diversity of research output, for example, in addition to scientific publications also data, software, models, methods, theories, algorithms, protocols, workflows, exhibitions, strategies, policy contributions, underpinning open science practices etc.

CRISs are already playing a part in compiling this information, but will play an even bigger role in this domain in the future. The recommendations and agreements have emphasised the responsible use of information. However, not that much attention has been paid to how this information should be collected in a responsible and appropriate way nor to assess the risks that may be involved so that the information about the diverse output is consistent and valid. These will be big questions in the further development of CRISs, i.e. how the responsible research evaluation policies can be considered, how to describe and demonstrate the versatility of research outputs, how to create an information base that instead of quantitative indicators better describes the true impact and quality of research, and how to reduce the misuse of information contained in CRISs, for example in the evaluation of individual researchers.

4. Towards a European CRIS infrastructure?

Besides exploring how to best use CRISs for the purpose of responsible research assessment, the next challenge in CRIS development – at least in Europe – is the implementation of international CRIS networks relying heavily on system interoperability for research information exchange and aggregation. In 2018, a pilot of a European publication infrastructure was carried out in the framework of ENRESSH network (Puuska, Guns, Pölönen, Sivertsen, Mañana-Rodríguez, & Engels, 2020). The idea of the ENRESSH-VIRTA-PoC was to explore the possibilities of benchmarking, comparing, and monitoring research outputs across institutional and national boundaries. It was built on the concept of the Finnish VIRTA system and compiled data from four countries, Finland, Belgium (Flanders), Norway, and Spain. The PoC demonstrated that it is possible to aggregate publication information across countries and the integration of national databases would create a comprehensive and well-structured information

base at the European level. Inconsistent data models were, however, identified as the main challenge and the conclusion was that further development of a European CRIS infrastructure would require automatised restructuring and reclassifying of data in a uniform manner and enriching metadata also from external sources.

At present, there are two particularly inspiring examples of forthcoming development. The first one of these is the recent launch last December of the national/regional research portals working group called CRISCROS (De Castro, 2022). This group, coordinated by the RIS Synergy project team in Vienna and euroCRIS, aims to bring together a large number of national and regional CRIS initiatives across the world to discuss best practices and foster a culture of mutual collaboration. The second one is the currently ongoing early feasibility analysis for putting together a CRIS for specific networks of European universities (of which there is an increasing number these days). Because several members of the university network already have an institutional CRIS – albeit based on a case-specific software solution – it is reasonable to think that a "network CRIS" may eventually emerge that collects and aggregates the research information for all institutions in the network. This is a significant challenge to the way things stand right now, with not all platforms out there having yet implemented the appropriate interoperability standards, but it could mean a big push for exploring the cluster-like opportunities offered by well-implemented system interoperability.

References

CoARA (2022). Agreement on Reforming Research Assessment. Retrieved from https://coara.eu/app/uploads/2022/09/2022_07_19_rra_agreement_final.pdf

De Castro P (2022). Introducing the National/Regional Research Portals Working Group. euroCRIS blog, https://eurocris.org/blog/introducing-nationalregional-research-portals-working-group

euroCRIS (2021). Joint euroCRIS/EARMA webinar: "RIM v RMA: Mutual areas of interest between Research Information Management (RIM) and Research Management and Administration (RMA) professionals" https://eurocris.org/joint-eurocrisearma-webinar-rim-v-rma-mutual-areas-interest-between-research-information-management

euroCRIS (2023). euroCRIS Directory of Research Information Systems (DRIS). Retrieved from https://dspacecris.eurocris.org/cris/explore/dris

GO FAIR (2023). FAIR Principles. https://www.go-fair.org/fair-principles/

Ribeiro L, Mennielli M, De Castro P (2016). Final report: EUNIS-euroCRIS joint survey on CRISs and IRs. EUNIS Research and Analysis Initiative (ERAI). https://www.eunis.org/wp-content/uploads/2016/03/cris-report-ED.pdf

Puuska, H-M (2019). The Research Information Hub as an access point to Finnish research. In *Spring 2019 euroCRIS Membership Meeting*, Espoo, Finland. http://hdl.handle.net/11366/986

Puuska, H. M., Guns, R., Pölönen, J., Sivertsen, G., Mañana-Rodríguez, J., & Engels, T. (2018). Proof of concept of a European database for social sciences and humanities publications: Description of the VIRTA-ENRESSH pilot. https://repository.uantwerpen.be/docman/irua/44ceb4/150108.pdf

Sîle L, Pölönen J, Sivertsen G, Guns R., Engels T (2017). European Databases and Repositories for Social Sciences and Humanities Research Output: Exploring Comprehensiveness. *Science, Technology, and Innovation Indicators.* Paris, France. https://enressh.eu/wp-content/uploads/2017/09/2017_ENRESSH_European_Databases.pdf

Simons, E (2004). Onderzoeksinformatiesysteem Metis nu in gebruik bij alle Nederlandse universiteiten. Radboud Universiteit Nijmegen Concernsystemen.

https://www.yumpu.com/nl/document/view/19395745/onderzoeksinformatiesysteem-metis-nu-in-gebruik-bij-alle- .

Sivertsen, G (2016). Publication-Based Funding: The Norwegian Model. In: Ochsner, M., Hug, S., Daniel, HD. (eds) Research Assessment in the Humanities. Springer, Cham. https://doi.org/10.1007/978-3-319-29016-4_7

Suominen T (2022). FAIRCORE4EOSC: Developing EOSC Core components to enable a FAIR EOSC ecosystem. euroCRIS Autumn 2022 Strategic Membership Meeting (Nijmegen, Netherlands, Dec 1-2, 2022). http://hdl.handle.net/11366/2267

TSV (2020). Tutkijanarvioinnin hyvät käytännöt. Vastuullisen tutkijanarvioinnin kansallinen suositus. / Good practices of researcher evaluation. National recommendation for responsible research evaluation. Vastuullisen tieteen julkaisusarja https://doi.org/10.23847/isbn.9789525995268

Wikipedia (2023). Current research information system. Retrieved from https://en.wikipedia.org/wiki/Current_research_information_system