

Bridging the Gap: Research Data in Research Information Systems

Introduction

In recent years, the number of implementations of research data as a data type in current research information systems (CRIS) has increased significantly. Firstly, this paper examines the ongoing discourse surrounding the relationship between research data and CRIS in the relevant scientific literature. Secondly, it analyses possible use cases for integrating the “research data” data type into an institutional CRIS, focusing on the research information system of the University of Vienna u:cris.

In addition to the CRIS, which only documents metadata, the University of Vienna has two repositories: an institutional repository, PHAIDRA, and a discipline-specific repository, AUSSDA, catering to the social sciences. Complementing these repositories, the university supports researchers in managing their research data with a data steward team consisting of six people. In 2021, the University of Vienna adopted a Research Data Management (RDM) Policy that describes the responsibilities of the University of Vienna and the researchers in terms of managing research data adequately.¹

Several small-scale studies have sought to assess and monitor the RDM practices of the university's researchers², but so far no substantial data – quantitative or qualitative – has been collected for the entire institution, nor for individual organisational units. With this in mind, this paper examines the possibility of leveraging the research information system u:cris to support and further develop the university's research data management services. The potential applications for implementing a dataset module in u:cris are illustrated by case studies from other CRIS implementations at both institutional and national levels.

Research Data as an Entity in Research Information Systems

CERIF, the Common European Research Information Model, developed and maintained by euroCRIS, depicts Research Datasets as an output type of the publication entity.³ This is also the case for the data exchange format OpenAIRE Guidelines for CRIS managers:⁴ The guidelines specify the use of the COAR vocabulary for resource types, which specifically contains the dataset type. This type also facilitates the differentiation between various types of data, such as clinical trial data, genomic data or geospatial data.⁵ The adherence to widely recognised standards for the depiction and exchange of research information typically enables the integration of research data as an output type in CRISs.

While not directly in the context of CRIS, the study of metadata schemas for research data plays a major role, in the adjacent field of research data management and long-term preservation of research data in repositories. This is especially crucial with regard to the FAIR publication of research data. However, the variety and application of metadata standards suitable for the depiction of research data is diverse, as evidenced by a recent study by Mosha and Ngulube. They argue that amongst the

¹ University of Vienna (2021). RDM Policy, p. 4.

https://rdm.univie.ac.at/fileadmin/user_upload/p_forschungsdatenmanagement/Dokumente/RDM_Policy_UNIVIE_v1_en.pdf. Accessed 29.02.2023.

² <https://rdm.univie.ac.at/research-data-management/rdm-at-the-university-of-vienna/>. Accessed 29.02.2023.

³ Dvořák, J. (2023). CERIF Tutorial: Spring 2023 Membership Meeting Brussels, p. 3.

<http://hdl.handle.net/11366/2452>.

⁴ OpenAIRE Guidelines for CRIS Managers: https://github.com/openaire/guidelines-cris-managers/blob/cb96b925159655adfd97fb11c4a93f3d20c8cbef/docs/cerif_xml_publication_entity.rst.

Accessed 29.02.2023.

⁵ COAR Controlled Vocabularies for Repositories (2022). Resource Types 3.1.

http://purl.org/coar/resource_type/.

most common metadata standards used for research data are DCAT, CKAN, METS, TEI, EAD, EML, ISO 19115, Dublin Core, DataStarR and DataCite.⁶ The University of Erlangen-Nuremberg, for example, advocates for the adoption of the DataCite metadata schema in their CRIS, FAU CRIS, citing its wider usage compared to the data extension in CERIF.⁷

Regardless of the technical feasibility, all recent studies examining the relationship between research data and CRIS explicitly argue in favour of implementing the data type: “CRIS are usually considered as an option to improve library services for research data management, especially by linking and storage. CRIS can enhance workflow and provenance control, improve data quality and efficiency and facilitate the discovery of data underpinning research publications.”⁸

This applies to the bigger picture of universities' efforts in the field of Open Science. Beyond this, the collection of research data in CRIS enables targeted analyses of discipline-specific data practices. Collecting information on various data types, archiving practices, or data sharing can help address questions raised by university management and research support staff.⁹

Use Cases and Case Studies

In an initial examination of the feasibility of implementing a research data module in u:cris, we identified the following use cases:

Showcasing Research Data

This use case examines the option of researchers entering metadata of their published research data into the CRIS, enabling them to showcase their work in a research portal, personal websites, or researcher profiles. While repositories such as Zenodo make it comparatively easy to publish research data, the integration of these publications in personal websites or project reports can be challenging. For this reason, the University of Erlangen-Nuremberg extended their CRIS data model to allow researchers to document published datasets via automatic data import.¹⁰

Needs Assessment for Research Data Management Services

Research support staff, namely data stewards, aim to carry out analyses and needs assessments for organisations or organisational units to promote the further development or adaptation of (discipline-specific) research data management services accordingly. At the Vrije Universiteit Brussels, three discipline-specific data stewards support the institution's researchers throughout the entire research cycle.¹¹ They oversee the manual and automatic import of datasets into the university's research information system (Pure by Elsevier).¹² Radboud University follows the same approach, where its

⁶ Moshá, N.F.; Ngulube, P. (2023). Metadata Standard for Continuous Preservation, Discovery, and Reuse of Research Data in Repositories by Higher Education Institutions: A Systematic Review. *Information* 14, 427, p. 11. <https://doi.org/10.3390/info14080427>.

⁷ Walther, M.; Wagner, M. (2022). FAIR research data integration in CRIS at FAU Erlangen-Nürnberg, *Procedia Computer Science* 211, p. 247. <https://doi.org/10.1016/j.procs.2022.10.198>.

⁸ Schöpfel, J.; Prost, H.; Rebouillat, V. (2017). Research Data in Current Research Information Systems. *Procedia Computer Science* 106, p. 317. <https://doi.org/10.1016/j.procs.2017.03.030>.

⁹ Rousi, A. M. (2023). Using current research information systems to investigate data acquisition and data sharing practices of computer scientists. *Journal of Librarianship and Information Science*, 55(3), p. 606. <https://doi.org/10.1177/09610006221093049>.

¹⁰ Walther, M.; Wagner, M. (2022), p. 247.

¹¹ <https://www.vub.be/en/contact/organisation/research-information-data-management-office>. Accessed 29.02.2023.

¹² Pauwels, R.; Van den Brull, E. (2023). Introducing the PURE User Community in Belgium, p. 15. <http://hdl.handle.net/11366/2457>.

CRIS serves as a one-stop-shop for research data management – from data management plans and the collection of metadata for research data to the long-term archiving of datasets.¹³

Monitoring of Research Data

University management, funders, or other stakeholders seek to monitor and/or evaluate research data as a specific type of research output from organisations, projects or individual researchers. This use case is perfectly illustrated by the Open Science policy initiative in Flanders: The Flanders Research Information Space FRIS is the primary source to monitor progress in Open Science and is therefore obligated to collect and evaluate research datasets from all institutional CRISs in Flanders.¹⁴

Challenges

Possible challenges in implementing a research data module in CRIS relate, on the one hand, to the choice of an appropriate metadata schema and, on the other hand, to the data input itself.

The metadata standard as well as the decision for or against specific mandatory fields should – depending on the institution – be able to represent the widest possible range of data types. The difficulty in choosing and adapting the metadata schema also lies in enabling compatibility with (discipline-specific) repositories, which in most cases are the original source of the metadata.

With regard to data entry, similar to other data in a CRIS such as publications or projects, it is important to mitigate the risk of errors and unnecessary administrative burdens for researchers and research support staff. One solution to this is to enable (automated) data import using interfaces, which in turn influences the choice of metadata schema or any necessary deviations from the standard.

Outlook

This paper aims to examine the possible use cases for showcasing research datasets in research information systems. The analysis indicates that three primary use cases play a key role in the implementation of a dataset module in an institutional or in a regional/national CRIS.

The critical aspect of implementing a dataset module in a CRIS involves selecting a specific metadata schema that ensures data quality across various use cases while allowing user-friendly data entry.

As a result of the description of use cases and the study of international case studies, this paper can serve as a guideline for universities when implementing a module for research data in an institutional CRIS.

¹³ Jetten, M.; Simons, E.; Rijnders, J. (2019). The role of CRIS's in the research life cycle. A case study on implementing a FAIR RDM policy at Radboud University, the Netherlands. *Procedia Computer Science* 146, p. 157. <https://doi.org/10.1016/j.procs.2019.01.090>.

¹⁴ Dengis, P. B.; De Bal, I. (2022). Monitoring Open Science policy using a regional CRIS – the Flanders case with FRIS. *Procedia Computer Science* 211, p. 102. <https://doi.org/10.1016/j.procs.2022.10.181>.