

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/200704>

Please be advised that this information was generated on 2021-06-15 and may be subject to change.

The Added Value of a CRIS in Archiving Research Data and Registering Data Management Plans *A Case Study of Radboud University and the National Data Archive DANS, the Netherlands*

Mijke Jetten

Inge Slouwerhof

Abstract

In 2016, Radboud University (Nijmegen, The Netherlands) and national data archive DANS initiated a project to extend Radboud University's CRIS (METIS) to allow researchers to register and archive their research data in the CRIS, at the same time making data available for reuse at the DANS archive. As we speak (spring 2018), a data management plan (DMP) module is being added, allowing researchers to draft and register DMPs via the CRIS. All modules together offer a one-stop-shop for researchers: data registration and archiving combined with registration of DMPs and articles, uploading full text to the publication repository, linking between results and the creation of researcher's profile pages.

The paper will explain the technical adjustments to the CRIS and DANS systems, and the services guaranteeing FAIR data archiving. This latter also includes services on data curation processes, which secure the quality of the metadata and documentation of the archived data. We use Radboud University as a good practice of the use of CRISs in the research data life cycle. It will be demonstrated that both for researchers and research institutes, a CRIS oriented approach to research data management brings added value.

Keywords

CRIS, Metadata, Research data management, FAIR, Data management plans, Archiving data, DANS data archive

Rationale: the benefits of incorporating research data management in the CRIS

The scientific environment nowadays experiences a shift, prompted by the urge for implementing an open science policy. Proper registration and archiving of data sets resulting from research has become an important issue in the last decade, both for the reason of scientific integrity and potential reuse of data. Key aspects of open science are storage, curation and sharing of research data, and, even more importantly, the FAIRness of research data, based on the FAIR Guiding Principles (Findability, Accessibility, Interoperability, and Reusability of data) (Wilkinson et al. 2016)¹.

Metadata is a highly important aspect of making data FAIR, allowing data to be findable ('data are described with rich metadata', F2 of FAIR) and reusable ('meta(data) are richly described with a plurality of accurate and relevant attributes', R1 of FAIR). Potential re-users can only find, interpret and reuse data based on a rich set of metadata, which describe for example the content, nature, format, language, location and authors of the data set. On the contrary, data lacking proper metadata is hard to trace and difficult to value. Thus, rich metadata is a necessary condition for open science (Mons et al. 2017; Prosser 2003)^{2,3}.

For researchers, however, providing metadata for research data may be considered a time-consuming administrative task, especially when this implies introducing 'yet another administrative system or interface'. Thus, at Radboud University, the idea was born to incorporate research data management in the university's CRIS, hence creating a one-stop-shop for researchers to register and upload both publications and data sets. This process is described in detail in the paper by Simons et al. (2017).⁴ The benefit of incorporating research data management in the university's CRIS is that it is placed in a much broader context, integrating the open science aspects of open access publishing and open data, without having to introduce researchers to yet another system; at Radboud University, at least part of the research staff was already familiar with or even using the RIS interface for researchers.

There are other benefits of incorporating research data management in the university's CRIS, in addition to using only one administrative system. This includes the external visibility of research output, making metadata and linked publications and data sets visible on the researcher's profile page. But it also entails linking to

contextual information, such as research projects and programs, funder information and professional employment history (Schöpfel et al. 2017)⁵. Research output, including data sets, can thus easily be included in the reports of the university's research institutes. Furthermore, a one-stop-shop CRIS solution will help Radboud University's recently appointed data stewards with implementing a FAIR data policy, making open science as feasible as possible.

Overview: the one-stop-shop CRIS solution

Radboud University's executive board stimulated to use the institutional CRIS as the central system and resource for registering and archiving research data. This decision was part of the university's central research data management policy.

Radboud University's CRIS, called Metis, has a long history. About 25 years ago, it was developed by Radboud University itself and it has been used for many years by most Dutch universities, functioning as a kind of national CRIS. With the dawning of Elsevier's PURE CRIS system, the number of other universities using Metis has dropped. Despite apparent financial and collaboration-oriented drawbacks, a benefit is the prospect of modifying Metis at a faster pace, as lesser stakeholders reduce the time spend on discussion and fine tuning between those stakeholders. That allowed Radboud University to modernize Metis over the past few years, adjusting it to the requirements of proper research data management, more specifically making research data FAIR, and during the last year, registering data management plans using the CRIS as well.

In addition to an administration and service desk module, Metis has a separate online interface for researchers. This interface, called RIS (Research Information Services), gives researchers direct access to the registration and management of their research results, among others monographs, articles, annotations, dissertations and lectures. Since 2015, this includes the possibility to register data sets in the CRIS, and, at the same, archiving the accompanying data files at the national DANS archive, without having to enter metadata twice or having to access the separate DANS EASY interface. Relations between data sets and other results, such as articles and dissertations, are part of the CRIS as well. These relations are made visible in the Dutch national publication database Narcis (see <https://www.narcis.nl/>), via the university's publication repository (Radboud Repository, see <http://repository.uhn.ru.nl/>).

For the newest feature, drafting data management plans, basic project registration and a data management plan module were added to the CRIS system, allowing researchers to register data management plans and completing the questions posed by funders and the university's research institutes.

In summary, the already existing online RIS interface for researchers was expanded with several functions to provide services for the whole research data life cycle:

- Registration of metadata of data sets
- Uploading data files for sustainable archiving at the national DANS EASY archive
- Linking data sets to publications, and vice versa
- Registration of projects and metadata of data management plans
- A data management plan module allowing researchers to fill in their research data management plans in the CRIS, including a collaboration and feedback function

As will be explained further below, the CRIS not only includes a RIS researcher module, but was extended with a RIS service desk as well. This way, researchers only have to use one administrative system (RIS) and request help, have data curated and ask questions about publication and data management at one service desk (RIS service desk, see <http://ris.ru.nl>), instead of having to use different systems and different communication channels.

Thus, over the past few years, at Radboud University, the CRIS was developed to act as a central system in the research information landscape. This position of the CRIS as a central pin is depicted in figure 1.

Extending the CRIS: technical and functional aspects

Metis is a Java-based application, including a relational database management system (RDBMS) based on Oracle. There are interfaces with local systems, such as the personnel system, Radboud University's publication repository, the personal profile portals, Web of Science and, for data sets, the DANS EASY data archiving system.

Tools used are Bootstrap, a widely used HTML-, CSS-, and JS-framework for developing responsive web applications/mobile projects; iBATIS, a framework that automates the mapping between SQL databases and objects in Java, .NET and Ruby on Rails; and jQuery, a cross-platform JavaScript library created to simplify the client-side scripting of HTML.

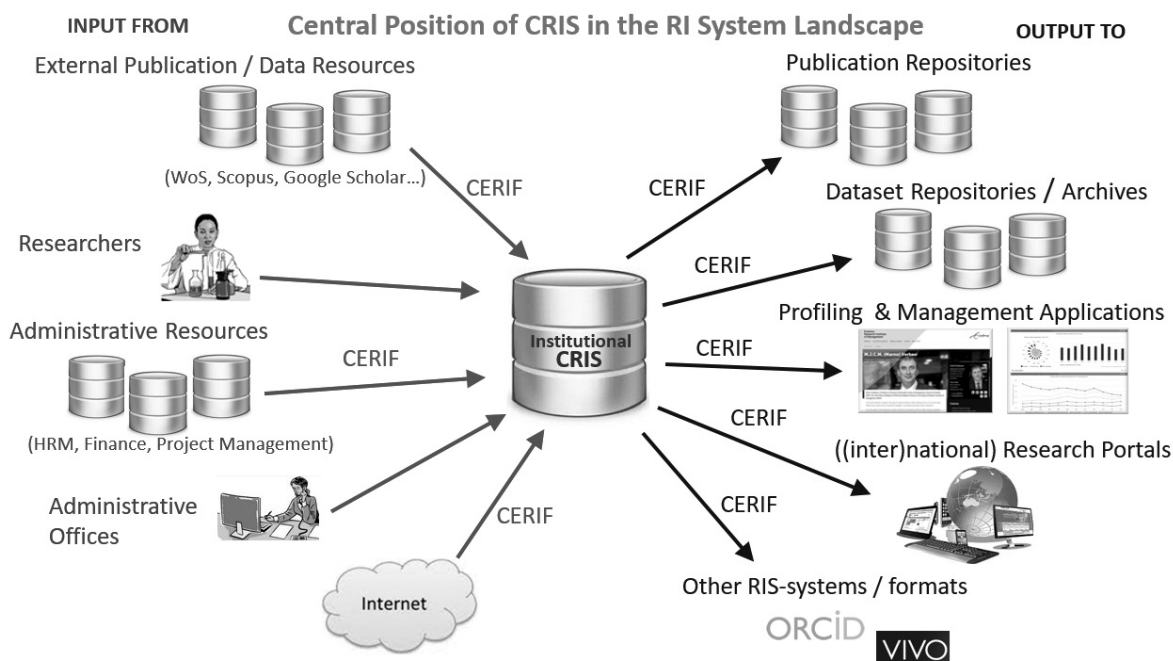


Figure 1 The central position of CRIS in the research information landscape.

The data files are uploaded from the CRIS, using the so-called SWORD-protocol (Simple Web-service Offering Repository Deposit), a lightweight protocol for depositing content from one location to another. It is a profile of the Atom Publishing Protocol, known as APP or AtomPub (see <http://swordapp.org/about>).

Data is stored in EASY (Electronic Archiving SYstem), the certified long-term preservation archive of DANS (see <https://easy.dans.knaw.nl/>). EASY offers functionalities for manual uploading data sets as well as bulk import of data. EASY is built upon the Fedora platform and maintained and developed by DANS' technical staff. Currently, EASY contains more than 52 thousand data sets, mainly originating from the humanities and social sciences, but increasingly from other scientific disciplines as well. Principally, data files in EASY are being curated by the DANS data managers, although for the SWORD ingested data files by Radboud University another agreement applies, involving curation by the front office of the archiving institution, i.e. Radboud University's RIS service desk.

A researcher registers the metadata of the data set via the RIS interface and uploads the data files. The files are not directly uploaded to EASY, but in a first stage stored on a local server at Radboud University. The front office at Radboud University Library (RIS service desk) checks the metadata and data files and, if necessary, communicates with the researcher for corrections/additions. Subsequently, the front office initiates the upload to DANS through the upload function in the CRIS. Since data curation is done by the front office, all data sets deposited from the Radboud University's CRIS are published almost directly, with only incidental checks by the DANS data managers.

Metis has a standard (generic) metadata schema for describing the various aspects of a data set. As point of departure, the international DataCite metadata schema combined with Dublin Core (used at the DANS EASY archive) were chosen. Because relevant metadata fields vary between the different disciplines, a wide variety of metadata fields were included in the CRIS, while making only basic metadata mandatory for all researchers.

For the newest feature, drafting data management plans, a question and answering module was added to the CRIS system. It works pretty straightforwardly, but the data management module has all the functionalities the DMP online tool (see <https://dmponline.dcc.ac.uk/>) has as well, including a collaboration and feedback function. Its strengths are its extensive functions, the connection to Radboud University's personnel system, the option to request feedback from the RIS service desk or the data steward within the tool, and the fact that data management plans are registered in the CRIS. This latter allows researchers to show the metadata on their profile page (in the future, if requested, publishing the plan itself as well) and link it to data sets. Additionally, it enables research institutes and their data stewards to monitor research data management via the content of the registered data management plans.

Support services and workflows

A centralized service desk

A central service desk is an elementary success factor of a CRIS based online interface as a one-stop-shop solution. At Radboud University, it is an expert team at the university library that operates the central service desk. Tasks of the service desk are service level management, quality management, call handling and reporting. It also includes research data management training of researchers and curating archived data sets. Thus, the original publication management services that the library has been providing for many years already have recently been supplemented with the new data management tasks mentioned before; such as, support in the registration of metadata of data sets, assisting researchers in uploading data files for sustainable archiving at the national DANS EASY archive, linking data sets and publications, and offering feedback on data management plans via the new data management plan module in the CRIS.

The front office back office model

An agreement has been made with the DANS archive, in which data sets are archived for the long term. Radboud University Library functions as front office, while the DANS national data archive serves as back office. The front office gives support to system users and trains the researchers in using the system and proper research data management. This also includes curation of the archived data sets. The back office ensures sustainable archiving of research data, provides persistent identifiers and guarantees the availability of research data, including exposure and the findability of research data.

The role of Radboud University Library as a front office has several benefits. First of all, by centralising research data management support, a stable and future-proof solution is created, building expertise in a fast-growing environment of research data management and open science. Additionally, a clear image is formed about the questions and needs that are present in the current research environment at research universities. Lastly, a central support desk could act as a linking pin, having contact with researchers and other stakeholders, and being involved in (almost) all research data management projects at a research university.

As part of the front office back office model, below, we sketch two workflow examples: data management plan checks and data curation. They demonstrate the variety of tasks of the RIS service desk and the embedded role of the CRIS in the research data life cycle.

Workflow example: data management plan checks

Early in the research data life cycle, the university library assists researchers in writing data management plans. For the researcher, drafting a data management plan is generally the first step in the awareness of proper data management and the FAIRness of data. Reflecting on research data management at the start of a research project helps researchers to comply to funder mandates, to make research reproducible, and to comply to privacy laws. Moreover, thinking about research data management in an early stage of the data life cycle helps to make a research project more efficient, and above all, enhances the FAIRness of data.

Radboud University Library, together with the research institutes' data stewards, functions as a central hub in supporting researchers from all disciplines (and more and more often, students as well, as they are the future generation of researchers) in writing data management plans, reading along and giving feedback on the plans written.

At the moment of writing this paper, a data management plan module is being added to the CRIS. Consequently, from this moment onwards, feedback on the plans written will be given using the CRIS, and it also allows for monitoring research data management via the data management plans by the institutes' data stewards

Templates included in the CRIS are the funders' formats and Radboud University's format. All research institutes at Radboud University have a disciplinary research data management policy, supplemental to the university's central policy. To align the data management plan formats to the demands of the various research institutes, we are currently implementing a cross-fertilization of the standard university format with the various research institutes' policy criteria. We assume this will make the drafting of data management plans easier and less time-consuming, and, at the same time, we expect it to enhance awareness of the data management criteria set by the research institutes.

Workflow example: data curation

Researchers at Radboud University use the CRIS based RIS system to archive research data at the national DANS archive. This includes uploading the data files and documentation and entering the accompanying metadata. Before the data set is uploaded to the DANS archive, the metadata, documentation and data files are checked by the RIS service desk. Criteria for that are based upon the before mentioned FAIR Guiding Principles.

As making research data findable is a first step towards FAIRness of data, the metadata are always checked by the RIS service desk. The metadata should be rich, understandable and complete in order for other researchers to find it, understand it and to be able to use it. Furthermore, the data set itself is curated by the RIS service desk. This includes checks and adjustments on the aspects of privacy-protection of the respondents, participants or subjects involved in the data set. Files that are open access accessible to others may not by any means contain privacy-sensitive information which can lead to identification of an individual.

Additionally, it entails rich documentation, as data should be well documented in order for potential other researchers to reuse the data. Variables should be explained in a codebook, syntax files have to be added, questionnaires should be included, and the context, content and structure of the data should be clear to a potential re-user.

FAIR data also entails reusability, and thus data curation by the RIS service also includes checks on the used software and preferred and accepted formats to ensure the preservation and future access of research data.

Feedback by the front office (RIS service desk) to the researcher is communicated and processed, before the dataset is sent to the back office (DANS archive). When the RIS service desk and the researcher are satisfied about the data set to be archived, the metadata and data files are transferred to the DANS EASY archive. A persistent identifier (DOI) is automatically assigned by the back office, including a DOI landing via the DANS EASY interface. The DANS data manager checks data sets incidentally and publishes data sets. The metadata, including the DOI and the access level, is made findable via the DANS EASY archive as well as via the university's publication repository (Radboud Repository). This process is visualized in figure 2. A potential re-user can access the dataset via the DOI landing page at DANS EASY.

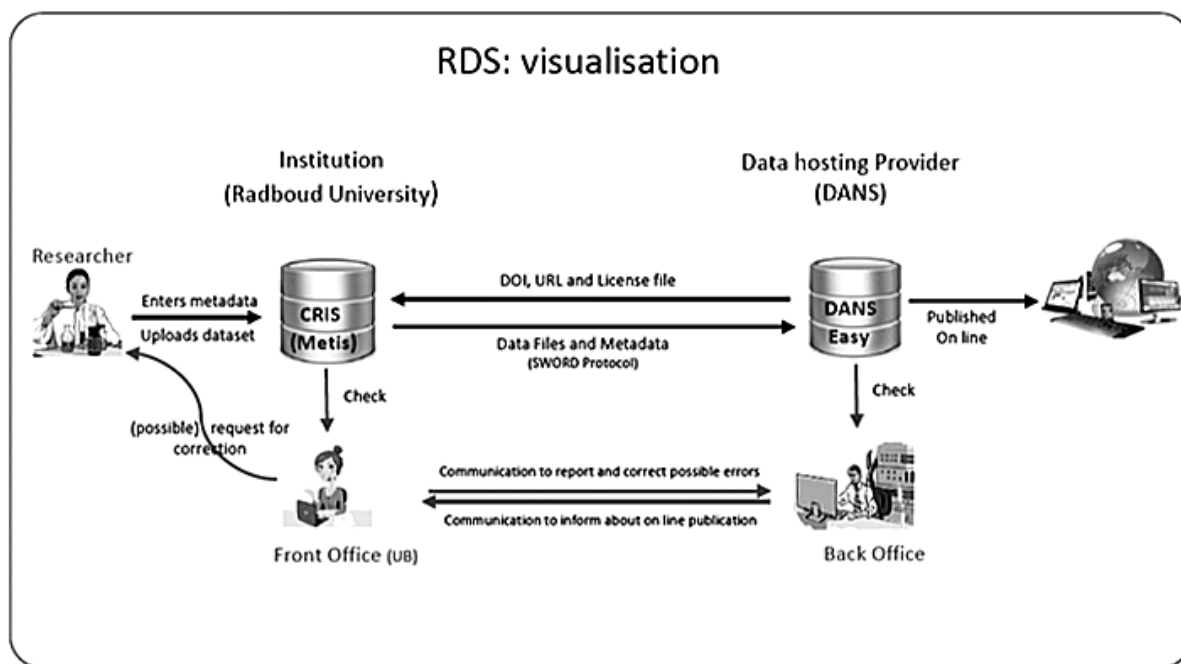


Figure 2 Front office back office model in data archiving at Radboud University

Future plans

At Radboud University, over the past years, steps have been taken to integrate research data management into the CRIS. In figure 1 we sketched the central position of the CRIS in the research information landscape. Adding a data archiving possibility in collaboration with the national DANS archive and the extra data management plan module have been major steps. In the context of the research data life cycle, further steps involving the CRIS can and will be taken over the coming years.

The data management plan module emphasizes the importance of extending the registration of project and funder metadata in the CRIS, allowing the data stewards and research support staff to monitor research data management using the CRIS. Concerning metadata, the main principle is register-once-use-many, and thus efforts are made to connect the local CRIS to the database of one of the major Dutch funders, KNAW-NWO, as well as connect it to the project databases of Radboud University's various research institutes.

For this, we will introduce CERIF (Common European Research Information Format) as XML format to send metadata from and to the CRIS. Ideally, in the near future, all communication between the CRIS and other stakeholders/systems (such as the DANS archive and the KNAW-NWO database) will be built upon the CERIF exchange model of research object and relations.

As part of the research data life cycle, Radboud Faculty of Medicine and the Radboud University Medical Centre have developed a Digital Research Environment (DRE), a cloud-based ICT platform which helps researchers to handle their data and create dedicated and secure workspaces for any kind of data project. Furthermore, the Donders Institute for Brain, Cognition and Behaviour (Radboud University) has created a data repository for internally archiving and publicly sharing research data for the long term. Both systems will surely benefit from exchanging metadata with the CRIS, as it guarantees that data is registered as research output as well. They will benefit from the connection between the local CRIS (Metis) and the national CRIS (Narcis) as well, as data stored in these two systems is or can be made visible and findable through the wider internet as well via the CRIS.

Although making data reusable is a main aspect of the FAIRness of data and a CRIS certainly assists in doing this, internally archiving data for scientific integrity could benefit from the connection to the CRIS as well. This for instance concerns data that can't be made public for privacy or legal reasons, or student thesis data. At Radboud University, in the near future, we will be registering this type of data in the CRIS as well, without making it available for reuse. Using the CRIS enables us to monitor this data over the ten years that it has to be stored, according to the university's central research data management policy.

In all these processes, Radboud University Library plays a central role, as it organizes the service desk for CRIS users as well as helps to further develop the CRIS system to the needs of researchers and policy staff. Researchers as well as data stewards are actively involved in the future plans just presented. As stated before, we benefit from the fact that Radboud University's CRIS (Metis) is self-developed; adjustments can be made relatively fast. Building a one-stop-shop for researchers in the evolving field of research data management requires an adjustable CRIS and iterative processes of discovering flaws and finding solutions.

Bibliography

1. Wilkinson, M.D. et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3(1). doi: 10.1038/sdata.2016.18
2. Mons, B., Neylon, C., Velterop, J., Dumontier, M., da Silva Santos, L. & Wilkinson, M. (2017). Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Information Services & Use*, 37, 49-56.
3. Prosser D. (2003). Institutional repositories and Open Access: The future of scholarly communication. *Information Services & Use*, 23(2), 167-170.
4. Simons, E., Jetten, M., Messelink, M., Berchum, M. van, Schoonbrood, H. & Wittenberg, M. (2017). The important role of CRIS's for registering and archiving research data. The RDS-project at Radboud University (the Netherlands) in cooperation with data-archive DANS. *Procedia Computer Science*, 106, 321-328. doi: 10.1016/j.procs.2017.03.031.
5. Schöpfel, J., Prost, H. & Rebouillat, V. (2017). Research Data in Current Research Information Systems. *Procedia Computer Science*, 106, 305-320. doi: 10.1016/j.procs.2017.03.030.

Mijke Jetten, Inge Slouwerhof

(Radboud University, Scotland) ■