IMPLEMENTATION STORIES

FAIRSFAIR
Fostering Fair Data Practices in Europe

THEME 6

SELECTING DATA, SERVICES, AND REPOSITORIES FOR FAIR

IMPLEMENTATION STORIES

THEME 6 SELECTING DATA, SERVICES, AND REPOSITORIES FOR FAIR

Two links in the research data life cycle: collaboration between a university and long-term repository

Authors
Marjan Grootveld and Maaike Verburg (DANS)

From interviews with
Marion Wittenberg (former repository account manager, DANS, The Netherlands) and the coordinator of Research Information Services, Radboud University (The Netherlands)

Data Archiving and Networked Services
DANS

Introduction

This story describes a collaboration between a university and a national data repository. More specifically, it involves the Research Information Services (RIS) at Radboud University (RU) and the long-term data repository at DANS (hence: the Repository), both in The Netherlands. A substantial part of the digital research data produced at RU are deposited at the Repository, which publishes and preserves them. Originally, the Repository was developed and implemented to provide individual researchers in the Netherlands with a trustworthy digital repository. It was designed for self-archiving, putting the data producer – the researcher – in charge of depositing the files and entering the metadata. In recent years, DANS witnessed a gradual shift towards institutional deposits.The RU has a home-built Current Research Information System (CRIS) called METIS, which is also used as a portal for archiving datasets in the Repository. RIS staff take care that their researchers prepare the data well for deposit and curate each dataset. Using the SWORD protocol the data is forwarded to the Repository, where a Repository data manager inspects and publishes the data. The Repository preserves them in the long run.

FAIRsFAIR recommendation

"Develop and implement guidance and support for selection of appropriate trusted digital repositories (TDRs)"

FAIRsFAIR Recommendations on practice to support FAIR principles

Aims and approach taken

For many years individual RU researchers in the Social Sciences and Humanities have deposited their data at the Repository. However, about five years ago the RU and DANS agreed that both organisations could benefit from a closer, streamlined collaboration. Since the first RU policy on Research Data Management in 2013, the RU commits to allowing researchers and faculties to choose services themselves, while also providing at least one option. In practice, researchers do not ask the RIS team for help in finding or selecting a repository. This could mean that they are familiar with relevant repositories, such as the domain-specific Donders repository for brain, cognition and behaviour [1].

The option selected and implemented by the RIS team is the CoreTrustSeal[2]-certified long-term repository service provided by DANS. The RU was looking for the combination of metadata registration in their CRIS with easier data publication, secure long-term data preservation including persistent identifiers, and more control over the data generated in the university. DANS always aims at publishing as much data as possible in the Repository (or in other trustworthy repositories). At the time, DANS was also developing and promoting the so-called front-office/back-office model[3]. In this model the front-office in the research-performing organisation can effectively support local researchers and encourage them to archive their data. The back-office in turn, possibly at a national level, can efficiently provide services to the front-office and the research-performing organisation, without too many dealings with individual depositors.

The process started with a very engaged contact person in each organisation: the RIS coordinator and the DANS Repository account manager. In the initial phase they discussed both procedural and technical aspects, as well as the respective responsibilities. Where necessary, colleagues and management levels were involved. This phase took about six months, during which also initial versions of the collaboration contract were drafted.

This contract was based on the standard DANS deposit agreement[4], but has been adapted to the organisation level and to describe responsibilities regarding the data quality and implementing the SWORD protocol[5]. For example, it is contractually agreed that the depositor on behalf of the RU describes the actual data producer(s) in the metadata in the Dublin Core Creator field – in contrast, individual depositors would typically enter their own name in this field. The RU is the rightsholder of the data. The RIS team is responsible for data curation (in collaboration with the researchers): for metadata compliant with the Repository's metadata schema, preferred file formats for sustainability, and relevant documentation such as codebooks. Another element in the contract is that DANS no longer accepts individual datasets from RU researchers; should they be deposited, they will be relayed to the RU front-office, which is Research Information Services or RIS.
The next stage, with the collaboration contract in place, was an iterative communication process about the – delivered versus expected – data quality. The RIS team started to promote the so-called RIS-DANS route. It was and is essential for them that researchers don't do double work. Researchers upload their data with metadata to the local CRIS and don't see any Repository interface. When a researcher submits data via the CRIS interface to the Repository, this ends up with the RIS curators. Where needed they contact the researcher, as they don't change data themselves. This is also a good opportunity to explain why data curation is needed: basically, to increase the reusability of the data and to meet for instance research funder requirements (see also below).

1. https://data.donders.ru.nl/?0

2. https://www.coretrustseal.org/

3. P.K. Doorn, I. Dillo, P. Witkamp (2014). Building a Federated Infrastructure for Preservation of and Access to Research Data in the Netherlands: The Front Office-Back Office Model. In: D. Katre, D. Giaretta: APA/C-DAC International Conference on Digital Preservation and Development of Trusted Digital Repositories. New Delhi: EXCEL INDIA PUBLISHERS. pp 72-77.

4. See https://dans.knaw.nl/en/legal-information/

5. SWORD is a protocol that supports automated deposit in a digital repository. https://sword.cottagelabs.com/

Initially it took the RIS curators time to get acquainted with the Repository's demands; the current RIS coordinator describes it as a phase of "double curation", viz. both by RIS staff and subsequently by Repository staff. However, the RIS curators learned a lot in the process, which they also use to support and indirectly train researchers. For several years now both organisations are satisfied about the division of work and the quality of RU datasets that the Repository receives.

Each dataset is checked and curated by two RIS staff members, one after another. If questions about privacy-related data remain, a third team member is involved. On average, curation of one dataset takes between 2.5 and 3 hours, with huge variations. It is evident that much less curation is needed if a researcher has been involved in the curation process before.

In addition to the actual curation activities, the RIS-DANS route contains an instruction[6] and a checklist[7] for registering and archiving data. Providing such guidance is part of the team's RDM support role. Over the years a step-by-step manual (e.g. "click here") evolved into a description that also includes what is expected from the researchers (e.g. "how you should deal with personal data"). A next step is to incorporate rationales for curation aspects in the CRIS itself, to support those who don't read instructions.

The Repository data managers no longer check all RU datasets. Likewise, since those early stages the role of the Repository account manager is limited. The frequency of contact strongly depends now on personal or organisational changes.

Challenges encountered and addressed

Especially at the beginning it was challenging for the RIS team to identify when a dataset is ready for publication. The "double curation" phase described above supported their learning process.

It is essential to explain to researchers why they sometimes have to put more work into their dataset before it can be accepted for deposit and publication, especially when they are pressed for time. There are examples of RU researchers who have deposited in other repositories without any assurance of the data curation quality. On the other hand, the RIS team experiences that researchers grow more confident about archiving and about their own data, when they know that RIS staff provides "extra eyes" on the quality and can answer questions like "can I share this publically?". When they follow the RIS-DANS route they comply with the demands imposed by research funders and journals. Furthermore, immediately when the Repository receives a dataset, the Repository returns the DOI for that dataset, which the researcher can for instance include in a publication or share with the research funder.

The RIS coordinator recognises that research support staff may want to push researchers in a certain direction ("trust me, this solution is good for you"). However, this can lead to frustration on both sides. Instead, one should rather focus on more "willing" researchers and their datasets, in the spirit of the RU vision to provide a service but allow the use of alternative services.

A challenge at the DANS side was that the organisation had little experience with providing services to organisations, because the typical Repository user was an individual researcher. For instance, there was no Service Level Agreement to explicitly address topics such as the maximal time to respond to questions or to solve technical issues. In hindsight DANS may not have been mature enough in this aspect at the time, but it was a valuable, joint, learning process, without leading to problems in the collaboration.

6. https://www.ru.nl/research-information-services/manuals/step-archiving-dataset/

7. https://www.ru.nl/research-information-services/manuals/checklist-dataset-deposit/

Impacts from the collaboration

The number of RU datasets deposited in the Repository has increased from about 10 to about 100 per year (and growing), coming from all RU domains including the Academic Medical Centre. The Repository data managers are very pleased with the curation quality of datasets that are deposited via the RIS. Although this may also be influenced by other developments in recent years, like the Open Science movement and the requirement to write Data Management Plans, it is very likely due to the active role of the RIS team in promoting and implementing the RIS-DANS route.

The former Repository account manager is convinced of the value of the front-office back-office model, when it comes to data curation before publishing the data. Decentralised data management with a central back-office is a strong model. Thanks to the fact that the RU CRIS allows the use of the SWORD protocol - not all CRIS providers support this - it was relatively easy to automate the RIS-DANS route.

The collaboration is clearly an instrument that enables FAIR data: the datasets in the Repository are better findable, thanks to the persistent identifier. Furthermore, the metadata of RU datasets are more consistent than before. Also, on a more abstract level, implementing the front-office back-office model equips the front-office - the RIS team - with more knowledge and services to support researchers in managing their data, which both partners expect to enhance the data FAIRness.

About FAIRsFAIR Implementation Stories

FAIRsFAIR Implementation stories illustrate good practices in research communities and organisations to support the implementation of the FAIR principles. These practices encompass 'FAIR-enabling' actions as recommended in the EC Expert Group on FAIR report Turning FAIR into Reality and the FAIRsFAIR Recommendations on practice to support FAIR principles. FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 Grant agreement 831558. The content of this document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

FAIRsFAIR - Fostering FAIR Data Practices in Europe

European Commission Grant Agreement No. 831558
H2020-INFRAEOSC-2018-4
www.fairsfair.eu - support@fairsfair.eu

Acknowledgements

This report has been produced by the FAIRsFAIR (GA No. 831558) project, which received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-05-2018-2019. FAIRsFAIR gratefully acknowledges everyone who we could interview for this series.

Disclaimer

The content of this document does not represent the opinion of the European Commission, and the European Commission is not responsible for any use that might be made of such content. **January 2022**

Copyright rests with the authors

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 **Grant agreement 831558.**
The content of this document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.