

# The Front Office–Back Office Model: Supporting Research Data Management in the Netherlands

Ingrid Dillo, Peter Doorn  
Data Archiving and Networked Services

## Abstract

High quality and timely data management and secure storage of data, both during and after completion of research, are an essential prerequisite for sharing that data. It is therefore crucial that universities and research institutions themselves formulate a clear policy on data management within their organization. For the implementation of this data management policy, high quality support for researchers and an adequate technical infrastructure are indispensable.

This practice paper will present an overview of the merging federated data infrastructure in the Netherlands with its front office – back office model, as a use case of an efficient and effective national support infrastructure for researchers.

We will elaborate on the stakeholders involved, on the services they offer each other, and on the benefits of this model not only for the front and back offices themselves, but also for the researchers. We will also pay attention to a number of challenges that we are facing, like the implementation of a technical infrastructure for automatic data ingest and integrating access to research data.

*Received* 16 October 2013 | *Accepted* 26 February 2014

Correspondence should be addressed to Ingrid Dillo, DANS, PO Box 93067, 2509 AB The Hague, The Netherlands.  
Email: [ingrid.dillo@dans.knaw.nl](mailto:ingrid.dillo@dans.knaw.nl)

An earlier version of this paper was presented at 9<sup>th</sup> International Digital Curation Conference.

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: <http://www.ijdc.net/>

Copyright rests with the authors. This work is released under a Creative Commons Attribution (UK) Licence, version 2.0. For details please see <http://creativecommons.org/licenses/by/2.0/uk/>



## The Importance of Storing and Sharing Research Data

Sharing research data is important. Data availability enables replication of research and thus enhances scientific transparency. Reusability of data also contributes to scientific progress and the advancement of knowledge.

Research funders acknowledge the importance of data sharing and require a high degree of open access. The Netherlands Organisation for Scientific Research (NWO) and the Royal Academy of Arts and Sciences (KNAW) both argue that research data obtained with public funding should be publicly accessible as much as possible. The European Commission wants to increase the impact of publicly funded research by making its results available to open access. In the context of the Horizon 2020 program, the European Commission is examining ways in which open access to research data can best be realised.

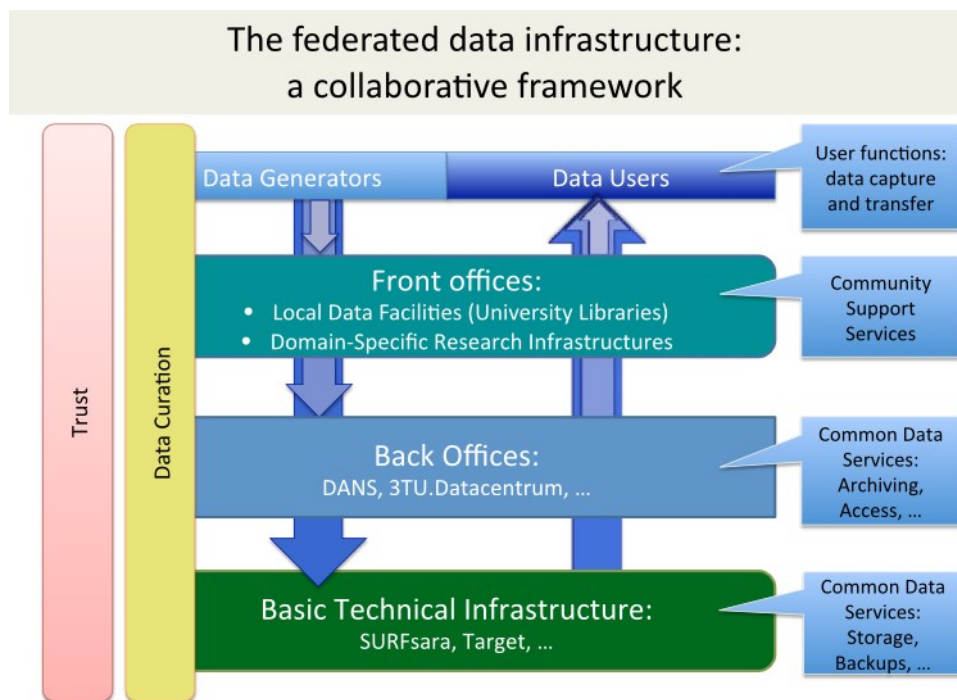
There are all sorts of snags to sharing data. The data should be searchable and accessible to others than those who collected them, and they must be stored in a consistent and sustainable way. Sometimes the privacy of individuals needs protecting. Additionally, conflicts of interest may arise when researchers collaborate with companies that want to exploit the data obtained. Also, many researchers do not want others to publish about the data collected by them before they themselves have had the opportunity to do so.

## Towards a Federated Data Infrastructure

Good and timely data management and secure data storage, both during and after completion of a research project, are essential prerequisites for sharing those data. It is therefore very important for universities and research institutions to formulate a clear data management policy for their organization. In order to implement this data policy, good support and an adequate technical infrastructure are indispensable.

In the influential report *Riding the Wave* (High Level Expert Group on Scientific Data, 2010), which has been embraced enthusiastically by European Commission Vice President and European Commissioner for the Digital Agenda, Neelie Kroes, such a ‘Collaborative Data Infrastructure’ is touted as a framework for the future.

In the Netherlands, a federated data infrastructure is developing from this reference model, with three layers of roles and responsibilities for the various stakeholders (see Figure 1). The foundation is a basic technical infrastructure, which facilitates data storage and back up. Above that is a layer of back office data services, providing facilities and support for long term archiving and accessibility. The highest level includes the front office services. They provide for the first-line contacts, supporting, advising and training researchers and students in responsible data management. The front office can rely on the expertise of the back office.



**Figure 1.** The federated data infrastructure: a collaborative framework.

## The Stakeholders Involved

In this federated and layered data infrastructure, the various stakeholders have specific responsibilities stemming from their respective positions and competencies.

The basic technical infrastructure is provided by data centres, an area where parties like SURFsara<sup>1</sup> and Target<sup>2</sup> have a coordinating role on a national or regional level.

The back office functions are carried out by organizations with a national role to play in the field of long term accessibility of data in trusted digital repositories, such as DANS<sup>3</sup> and 3TU.Datacentrum<sup>4</sup>, collaborating in Research Data Netherlands. Together they have expertise on data from the humanities, sciences and social studies.

The front offices are located at universities (libraries, local data centres), research/knowledge institutes, colleges of applied science, national and international research infrastructures (ESFRI/National Roadmap); for some features they can also be found with the funders (NWO, ZonMw, Ministries). What all these organizations have in common is that they are primarily responsible for the quality assurance of the data produced and processed by them or for them.

<sup>1</sup> SURFsara: <http://www.surf.nl/en/about-surf/subsidiaries/surfsara>

<sup>2</sup> Target: <http://www.rug.nl/science-and-society/target/>

<sup>3</sup> DANS: <http://www.dans.knaw.nl>

<sup>4</sup> 3TU.Datacentrum: <http://datacentrum.3tu.nl/>

## Services in the FO-BO Model

In the federated data infrastructure roles can be divided according to the front office–back office model (FO-BO model). The services provided in this model are all related to data management and data storage.

They fall roughly into three groups:

1. Awareness raising and information provision;
2. Training (focusing both on data librarians/experts and on researchers);
3. Data curation, management and storage during and after research projects.

## Roles and Responsibilities in the FO–BO Model

The focus of the front office is on supporting its own research organization. In the area of data management the front office takes care of awareness raising, providing information and training its researchers.

In addition, the front office features so-called Virtual Research Environments or Data Labs, which offer research tools and secure temporary storage facilities (Sharepoint, Dataverse, etc.) for the organization’s researchers. In consultation with the back office, the front office also facilitates the transfer of data to a trusted back office digital repository after the research has been completed. Facilities that are shared by several universities, including Dataverse, can be hosted and supported by the back office.

Data acquisition within its research community is another front office duty. In all its tasks the front office will, if necessary, maintain contact with the back office.

The focus of the back office is on the expertise surrounding data governance and data stewardship<sup>5</sup>, including long-term storage and accessibility of the research data. The back office is responsible for training the data librarians/experts employed by the front office and providing the front office with substantive support via dedicated contacts. Back office employees may act as experts and contribute to the front office training activities for researchers.

Where required, the back office also provides consultancy services to the front office. In other words, the back office acts as a centre of expertise and innovation. Furthermore, the back office ensures the sustainable and secure storage and retrieval of data upon completion of the research project. For this purpose the data are transferred through the front office to the back office.

In the acquisition, support, consultancy and training services, the duties of front and back offices may overlap. Coordination and definition of responsibilities will therefore be necessary. Specific responsibilities will vary from organization to organization, but it is important to have clear agreements on e.g. data acquisition, NWO data contracts<sup>6</sup> and the use of data management plans.

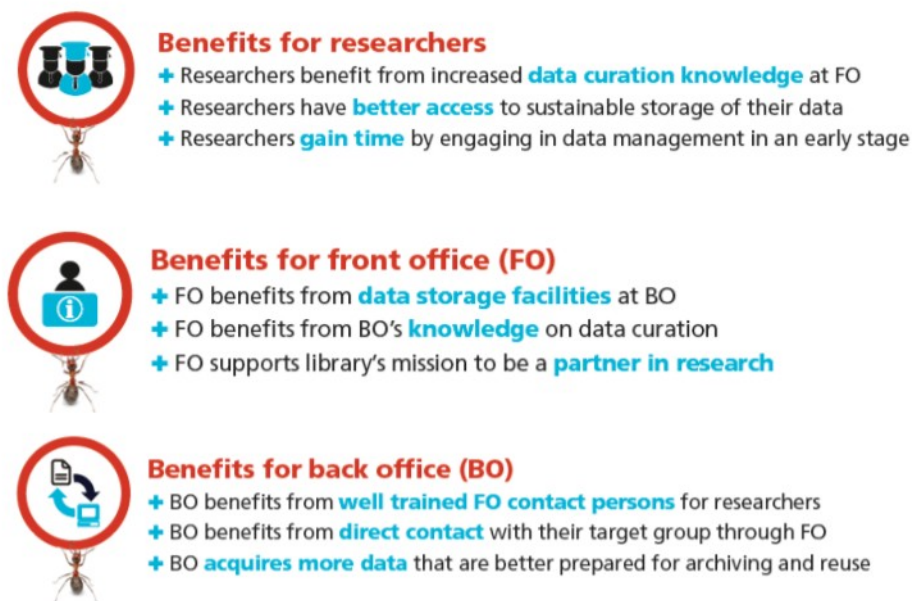
<sup>5</sup> Data stewardship is the management and oversight of an organization’s data assets to provide business users with high quality data that are easily accessible in a consistent manner. While data governance generally focuses on high-level policies and procedures, data stewardship focuses on tactical coordination and implementation. Data stewards can also be responsible for carrying out data usage and security policies as determined through enterprise data governance initiatives. For more information, see <http://searchdatamanagement.techtarget.com/definition/data-stewardship>

<sup>6</sup> NWO data contracts: <http://www.dans.knaw.nl/en/content/data-archive/data-contracts>

## Benefits of the FO-BO Model

The FO-BO model offers benefits for all stakeholders because it provides for an optimum division of labour based on the respective expert competencies of those stakeholders and their various roles in the data infrastructure.

Figure 2 summarises these benefits for both back and front office, and for the research community.



**Figure 2.** Stakeholder benefits of the FO-BO model.

## DANS Business Model

As a trusted digital repository, DANS performs back office functions, supplemented by a number of national and international front office duties. In its business model, DANS aims at institution-wide, formal agreements with the front offices at the universities in the context of the FO-BO model. Such framework agreements set out the details of data management and data storage services based on the FO-BO model outlined above. They may involve the distribution of responsibilities, but also the establishment and maintenance of the required technical infrastructure, for example.

DANS will charge basic data storage costs including back-up. Storage is currently supplied by Vancis<sup>7</sup>, a subsidiary of SURFsara. There is a Service Level Agreement with Vancis in place which guarantees a very high level of security and data availability.

In return for a one-off payment of these costs for five years in advance, DANS ensures conservation of the data “forever”. This will enable the long term safekeeping of data from projects with temporary funding.

<sup>7</sup> Vancis: <https://www.surfsara.nl/vancis>

This arrangement assumes that data and metadata are supplied in the agreed format. Where this is not the case, DANS will charge extra per hour for processing data and organizing documentation. For larger consultancy projects DANS will also charge on an hourly basis.

## Challenges in the Federated Data Infrastructure

In the federated data infrastructure, we have to deal with questions concerning both ingest and dissemination of data. Who is to take care of which data throughout the research life cycle, and where will it be stored? And how can we provide access to research data that are stored in distributed repositories, temporary and permanent, which are increasingly interwoven with publications and other information researchers need?

### Infrastructure for Automatic Data Ingest

We envisage a situation in which universities and research organizations are putting increasing demands on the quality of data management in different phases of the scientific process: during data collection, processing and analysis, it is the researcher or research group who is responsible for data management. A growing number of universities are offering collaborative tools for data management. Some universities offer Microsoft SharePoint to keep track of projects and for content management.

A group of universities has set up the Dutch Dataverse Network (D-DVN)<sup>8</sup> for sharing and storing research data, based on the Dataverse software<sup>9</sup> developed at the Institute for Quantitative Social Science at Harvard University. It is a web application for publishing, citing, analysing and preserving research data.

Dutch DVN is created for researchers and lecturers of universities in the Netherlands. The service makes it possible to store a wide variety of scientific data (texts and raw research data, but also video material and complete databases) in an online environment, safely and sustainably.

Dutch DVN allows researchers to index the data they have stored in a user-friendly way, offering them the opportunity to share their data with other scientists or interested parties. The researchers can determine themselves who gets access to which data, and they can refer to stored datasets by means of a persistent identifier.

Although Dataverse is a repository for research data that can also be used for long term preservation, the organizational setup of the Dutch-DVN makes it impractical to guarantee good archival practices: these will only be pursued as long as the researchers in charge take this responsibility. Therefore the partners in D-DVN have chosen to involve DANS and 3TU.Datacentrum (working together in the earlier mentioned Research Data Netherlands coalition) for the long term storage. By means of an interface based on the SWORD<sup>10</sup> protocol, the data in D-DVN can be automatically ingested into the permanent archives at these institutions. A similar system will be implemented to ingest the data from SharePoint. In the near future, researchers will be able to simply indicate when the data are to be permanently archived (and where).

<sup>8</sup> Dutch Dataverse Network: <https://www.dataverse.nl/dvn/>

<sup>9</sup> Dataverse: <http://thedata.org/>

<sup>10</sup> SWORD (Simple Web-service Offering Repository Deposit) is an interoperability standard that allows digital repositories to accept the deposit of content from multiple sources in different formats (such as XML documents) via a standardized protocol.

Whether Dataverse and/or SharePoint perform the functions researchers “really want” is as yet an open question. It is certainly true that both systems are only suitable for the long tail of small data. At meetings with researchers we often hear that they would like “something with the ease of use of Dropbox”. However, such a lightweight cloud service for data storage and sharing lacks the descriptive metadata that would be necessary to let others understand what your research data are about. Perhaps new developments, such as the iDROP interface to upload, manage and download files in an iRODS<sup>11</sup> repository will offer the next solution.

### **Integrating Access to Research Data**

The research data, publications and other scholarly information researchers need for their scientific work are increasingly distributed and interwoven. How to provide researchers with the necessary overview, and how to provide access, or at least make it transparent what is accessible, and under which conditions, are key issues.

DANS takes the stance that publicly funded research data and publications should be open by default, although we accept restrictions for a variety of reasons (ranging from privacy of respondents to the protection of archaeological heritage). Rather than taking a dogmatic view on openness, we use the slogan “open if possible, protected if needed”.

Our NARCIS<sup>12</sup> portal already is a gateway to scholarly information in The Netherlands, and our aim is to further develop NARCIS into the most important national portal for scientific information in the sciences, social sciences and humanities. Of course, national boundaries are not particularly relevant for science and scholarship, and this is why we are giving priority to incorporating the information in NARCIS, which essentially consists of metadata describing research data and publications in academic repositories, as well as information on research (projects, organizations, researchers), in as many international information service providers as possible.<sup>13</sup>

We see two trends that we expect to be dominant in the next few years: the linked open data movement and the increasing integration of data and digital publications in “enhanced publications”.

With respect to the first, NARCIS harvests the (meta) data from a growing number of academic (or more generally: “higher education”) repositories and research information systems. By the end of 2013, NARCIS contained information on 850,000 publications, 30,000 data sets, 60,000 research projects, 50,000 researchers, and 3,000 research organizations (down to the institute/faculty/department level) in The Netherlands. The NARCIS information is exposed as Linked Open Data (published as RDF with dereferenceable URIs for easy data retrievability and provided as a SPARQL endpoint for flexible data access).

DANS started the second development with a number of pilot projects in 2011, funded by the SURFshare programme on enhanced publications. NARCIS contains over 1,700 of such enhanced publications. The relations among the components are graphically represented by an “in context visualizer”<sup>14</sup>.

<sup>11</sup> Integrated Rule-Oriented Data-management System (iRODS), is a community-driven, open source, data grid software solution. iRODS helps researchers and archivists manage (organize, share, protect, and preserve) large sets of computer files, ranging in size from moderate to a hundred million files or more totalling petabytes of data.

<sup>12</sup> National Academic Research and Collaborations Information System (NARCIS): <http://www.narcis.nl>

<sup>13</sup> One of these is OpenAIRE: <http://www.openaire.eu>

<sup>14</sup> See: <http://www.surf.nl/en/themes/research/research-data-management/enhanced-publications/index.html>

We are currently preparing for a next step: making it possible to publish one or more datasets in the archives as data articles, and afterwards, to collect these data articles in a digital volume similar to a data journal. Such a data article may be much shorter than a normal research paper (a text of 2-4 pages will be sufficient, but it may be longer if the researcher wishes): the research data should be introduced by a description of the project (including the research question), the procedure of data gathering and processing, illustrated with graphs and tables, and perhaps a few significant features of the data set. A full analysis and conclusions are not required. Because a data paper will be a digital document published on the web, in an advanced version the graphs and tables could be interactive widgets embedded in the text, which allow the user to preview and to explore the data as part of the reading.

The data article will be citable, and we will extend our existing community review procedure of data sets into a more full-fledged peer review system (Grootveld and van Egmond, 2011). What remains is a matter of presentation, branding and marketing, the planning for which is currently underway.

## References

- High Level Expert Group on Scientific Data. (2010). *Riding the wave: How Europe can gain from the rising tide of scientific data*. Retrieved from European Union website: <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>
- Grootveld, M., & van Egmond, Jeff. (2011). *Data reviews, peer reviewed research data* (DANS Studies in Digital Archiving No. 5). The Hague: DANS.