

Providing access to research data, publications and current research information at Data Archiving and Networked Services - DANS

Elly Dijk, Peter Doorn

elly.dijk@dans.knaw.nl, peter.doorn@dans.knaw.nl

Data Archiving and Networked Services – DANS, the Netherlands

Abstract

Data Archiving and Networked Services (DANS) promotes sustained access to digital research data in the Netherlands. Researchers can deposit their data through the online archiving system EASY. Via the portal NARCIS the research data are shown in context, namely in relation to e-publications, and other research information. Both EASY and NARCIS contain grey literature like archaeological reports, data documentation, doctoral thesis, conference papers, patents, and technical documentation of universities and other research institutes.

Keywords

Research Data, Archiving Systems, E-publications, Research Projects, Research Institutions, Researchers, Enhanced Publications, NARCIS, EASY, Dutch Dataverse Network

Introduction

The mission of the Dutch institute Data Archiving and Networked Services (DANS)¹ is to promote sustainable access to digital research data (such as databases, spreadsheets, text, images, audio, video, multimedia), research information (information on research, research organisations and researchers) and electronic publications (including preprints, reports).

To carry out its mission, DANS offers services for long-term archiving and re-use of completed research data, mainly in the humanities and the social sciences, and it supports data management during ongoing research (research data management). Additionally, DANS offers training and consultancy.

DANS supports its services with research into sustainable access to digital information. Research (Research & Innovation) is a derivative of the main objective and is focused on innovation and improving service to users.

DANS is an institute of the Royal Netherlands Academy of Arts and Sciences (KNAW)² and the Netherlands Organisation for Scientific Research (NWO)³ since 2005, but the first predecessors of DANS date back to 1964, the Steinmetz Foundation for the social sciences, and to 1989, the Historical Data Archive.

The emergence of institutes for data archiving

In the sixties of the 20th century we saw the emergence of data archives in the social sciences, like the Dutch Steinmetz Archive and the British UKDA. Later followed by text archives for linguistics and literary studies in the seventies (e.g. Oxford Text Archive), historical archives in the eighties and nineties (e.g. NHDA, HDS, IPUMS), and archaeological data archives at the beginning of this century (e.g. ADS, EDNA). About ten years ago the universities started to develop university repositories for publications and research data. General data sharing facilities, such as Zenodo, Figshare and the EUDAT B2-tools are even more recent.

In the past few years there is growing attention for the value of (big) data, data sharing and proper research data management. Data sharing has become an important issue among research funders, policy makers, in the scholarly world itself, among publishers, and even for the general public. In 2010 the influential EU report, “Riding the Wave”⁴ about the future of scientific data was published. In 2011 Neelie Kroes, Vice-President of the European Commission responsible for the Digital Agenda said that “Data is the new gold”⁵. In the EU programme Horizon 2020 one of the goals is “open research data”. This has a positive effect on the willingness of researchers and on their institutes to share research data.

DANS offers services for data curation, archiving and data management in the social sciences and humanities. The institute has been concerned with the awareness of the value of preserving data for re-use, so one can use the data for validating the results of earlier research. Also it is

¹ <http://www.dans.knaw.nl/en>

² <http://www.knaw.nl/en>

³ <http://www.nwo.nl/en>

⁴ Riding the wave: How Europe can gain from the rising tide of scientific data. Final report of the High Level Expert Group on Scientific Data. A submission to the European Commission, October 2010, 40 p. See:

<http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>

⁵ http://europa.eu/rapid/press-release_SPEECH-11-872_en.htm?locale=en

possible to use the data for comparative analysis, and researchers can use the data for secondary analysis, that is to say answering new research questions with existing data.

Federated data infrastructure in the Netherlands

On the initiative of Research Data Netherlands (RDNL)⁶, a partnership between DANS, 3TU.Datacentrum and SURFsara, a federated data infrastructure is being developed, with services, roles and responsibilities for various parties in the Netherlands. The “Collaborative Data Infrastructure”, a framework for a future data infrastructure from EU report “Riding the Wave”, formed the inspiration for this model. RDNL elaborates this framework in four instead of three layers: the layer Common Data Services from the 'Riding the Wave' report is split into a technical basic infrastructure and back-office data services (see Fig.1).

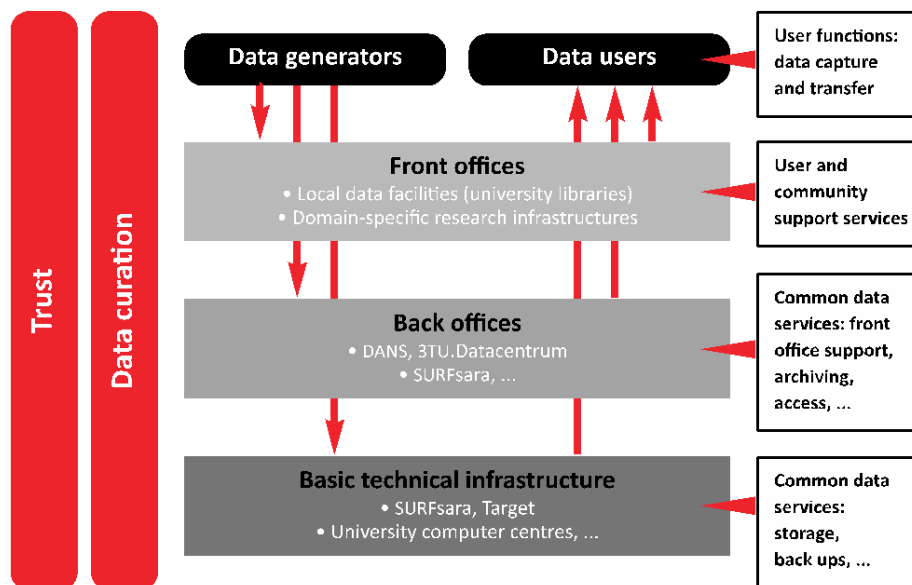


Fig. 1. Federated data infrastructure in the Netherlands

The foundation of the federated model is formed by the technical basic infrastructure, that are the computer centres which take care of data storage and backup facilities. Above this is a layer of back-office data services, which provides facilities and support for the long-term archiving and accessibility of the research data in trusted digital repositories, like the DANS archiving system EASY. Another level up one can find the front-office services, which are performed by university libraries as the first contacts with producers and users of research data. The European Research infrastructures such as CLARIN⁷, DARIAH⁸ and CESSDA⁹ also perform front-office functions in the domains they represent. The vertical bar "Trust" in the figure refers to the importance of integrity, transparency and/or certification at all levels. "Data curation", or

⁶ <http://www.researchdata.nl/en/>

⁷ <http://clarin.eu/>

⁸ <http://www.dariah.eu/>

⁹ <http://www.cessda.net/>

the care that data needs to remain available for the long term covers all phases of the life cycle of research data and it is a task to which all levels of the model contribute¹⁰.

The services of DANS

In order to support the research community with research data management DANS offers different services (more about these services in the underlying sections):

- EASY¹¹: electronic archiving system for self-deposit of research data. Researchers and institutions in the social sciences and humanities can submit their data in EASY. A majority of the datasets are under open access licences available for reuse.
- Dutch Dataverse Network (DDN)¹²: researchers can use this service to store scholarly data during the research process in a safe and sustainable manner in an online environment.
- NARCIS¹³: provides access to scholarly information, including (open access) publications from the repositories of all the Dutch universities, the Academy, the national research council (NWO), and a number of research institutes as well as datasets from some data archives (e.g. DANS). NARCIS also contains descriptions of research projects, researchers, and research institutes.
- Data Seal of Approval (DSA)¹⁴: DANS has developed minimum requirements for (scholarly) data archives.

Electronic archiving system EASY

Contents

EASY's collection currently (October 2014) consists of 28,130 datasets. These datasets are distributed among the following disciplines (a dataset can belong to more than one discipline): humanities – 25,933 datasets; socio-cultural sciences – 3,623 datasets; social sciences – 2,554 datasets; behavioural sciences – 993 datasets; life sciences and medicine – 276 datasets; and geospatial sciences – 40 datasets.

The collection Humanities consists primarily of a historical collection and an archaeological collection. The historical collection contains a broad range of historical datasets, which relate to the Netherlands and the former Dutch colonies. The archaeological collection, that forms a large part of the Humanities collection, is also known as the e-Depot for Dutch Archaeology, which consists of archived excavation reports and related datasets of archaeological excavation and exploration projects. There is also a substantial oral history and qualitative social science collection, consisting of recorded and/or transcribed interviews. Oral history is the collection and study of historical information using sound recordings of interviews with people having personal knowledge of past events, like the Interview project Dutch Veterans.

¹⁰ See: <http://www.dans.knaw.nl/sites/default/files/file/DANS%20FO-BO-model%20UK%20V1%20DEF%20web.pdf>

(An update will be published later this year)

¹¹ See: <https://easy.dans.knaw.nl/ui/home>

¹² See: <http://www.dataverse.nl>

¹³ <http://www.narcis.nl>

¹⁴ <http://www.datasealofapproval.org>

The social and behavioural sciences data collection consists of a large number of social survey data relating to the Netherlands from the late 1950s to the present, including longitudinal surveys of the Netherlands Institute for Social Research (SCP). The collection also includes National Election Studies (*Nationale Kiezersonderzoeken*) and protected microdata files of Statistics Netherlands (CBS).

The Geospatial Sciences collection mainly consists of files from the Dutch Land Registry Office (Kadaster).

Reuse of data

In 2013, about 30,000 datasets were downloaded. The reviews of the datasets show that users highly appreciate the reuse of data¹⁵.

Besides the download of the datasets deposited in EASY, there are many national and international projects where research data sharing concerns collaborative work. DANS is involved for example in international European infrastructure projects like Holocaust studies (European Holocaust Research Infrastructure - EHRI¹⁶), ARIADNE (for integration of archaeological datasets)¹⁷, Dendrochronology (Digital Collaboratory for Cultural Dendrochronology – DCCD)¹⁸, Global inequality (Clio Infra)¹⁹ and Dutch collaborations for instance on Dutch Census Data 1795-1971 (a cooperation of four institutes)²⁰ and Historical Shipping (bringing together shipping records)²¹. (See also chapter 4).

Grey literature in EASY

EASY not only contains datasets but also grey literature. These publications, around 14,000, are connected to the datasets. One can find in EASY archaeological reports (in which the excavation results are described and documented), publications belonging to research datasets, and data documentation.

Dutch Dataverse Network - DDN

In May 2014 DANS took over the hosting of the Dutch Dataverse Network (DDN) from Utrecht University, where it was installed in 2010. The Dataverse Network is a scholarly data storage and sharing platform developed by the IQSS at Harvard University and is used worldwide. Researchers and lecturers from universities and other academic institutions can store and share their data in DDN during the course of their research. It also makes possible persistent data citation, and it enables reproducible research. With this service DANS supports researchers with research data management (RDM). Since May the number of users was extended from 7 to 10 institutions. Most of the 13 Dutch universities are currently formulating and implementing RDM policies. We expect this will have a boosting effect on the use of Dataverse by Dutch scholars.

¹⁵ <http://datareviews.dans.knaw.nl/index.php>

¹⁶ <http://www.ehri-project.eu/>

¹⁷ <http://www.ariadne-infrastructure.eu/>

¹⁸ <http://dendro.dans.knaw.nl/>

¹⁹ <http://www.clio-infra.eu/>

²⁰ <http://www.volkstelling.nl/>

²¹ <https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:33898>

The scholarly portal NARCIS

Since 2011 the portal NARCIS is a service of DANS. DANS took over this portal with the idea that it is not only important to provide open access to data and publications, but that it is also necessary to connect them to research information. Research information can be described as information about researchers, research institutes and research projects. With this connection the research data and publications are shown in their context.

NARCIS can be seen as the national aggregator of the Current Research Information Systems and repositories of the Dutch science system. It harvests the metadata of 7 repositories with datasets, for example EASY, and 33 repositories with (open access) publications. Besides that NARCIS has a controlled database with information about (current) research, researchers, and research institutes. The information in NARCIS is connected via the Digital Author Identifier (DAI) that is given to researchers, authors and depositors of datasets. The DAI is a unique national identifier assigned to every author working at a Dutch university or research institute. (In the Netherlands there are plans to replace the DAI by the international author identifiers ORCID and/or ISNI). We aim to present the information in an increasingly integrated and interconnected way.

The metadata of the publications and datasets within NARCIS, such as title or author's name, are also available via other services²², like Google Scholar²³, DART-Europe E-theses Portal²⁴, OpenAire²⁵, and WorldWideScience.org²⁶.

Contents

In October 2014 NARCIS contains over 900,000 publications, of which 41% are open access publications. It features nearly 175,000 datasets, over 51,000 researchers including 8,500 professors and associate professors, almost 60,000 research projects (25% current research) and 2,900 research institutes²⁷.

In addition NARCIS contains more than 1,800 enhanced publications. Dr. Leen Breure, researcher at DANS and Utrecht University, has made a typology of 80 types of enhanced publications. This was made in the context of the Xpos're, research project on scholarship and multimedia²⁸. Enhanced publications contain additional materials that have been attached to the full text version, such as underlying research data, models, images, conference papers, ranking data, and comments. The enhanced publications contained in NARCIS originate from various projects and cover a wide range of research fields, like archaeology, humanities, and social sciences. The presentation of these enhanced publications is still a beta version. DANS has however plans to change the way enhanced publications will be displayed. We will also soon launch a Data Journal in which data depositors can publish data papers, which describe their datasets.

²² See for the overview of these services: <http://www.narcis.nl/dataprovider/Language/en>

²³ <http://scholar.google.com/>

²⁴ <http://www.dart-europe.eu/basic-search.php>

²⁵ <https://www.openaire.eu/>

²⁶ <http://worldwidescience.org/>

²⁷ This number is including all the parts of the 13 main universities in the Netherlands.

²⁸ See: <http://xposre.nl>

Grey literature in NARCIS

Over 300.000 (open access) publications in NARCIS (33%) are grey literature: reports, doctoral theses, conference papers, contributions to periodicals, book reviews, working papers, annotations, patents, lectures et cetera.

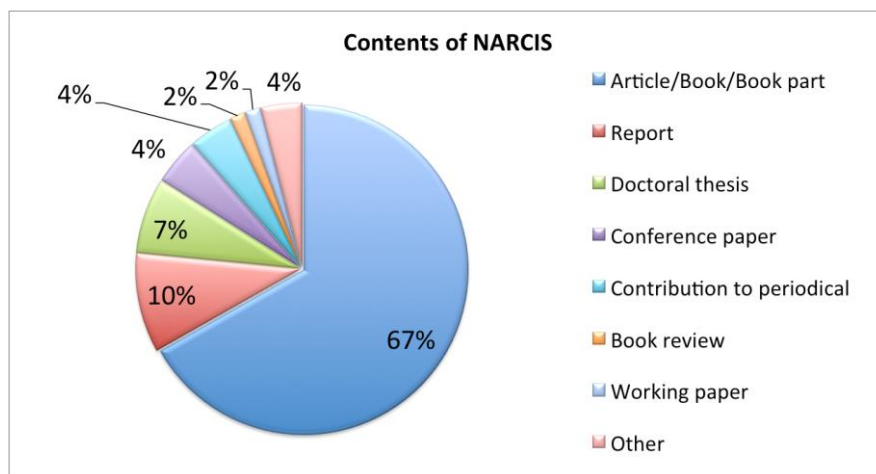


Figure 2: Contents of NARCIS, October 2014

It depends on the sort of publication if the publication is available as open access or not. For example 82% of the doctoral theses is open access, 83% of the patents, and 58% of the conference papers. On the other hand “only” 37% of the contributions to periodicals are openly available.

Data Seal of Approval (DSA)

For researchers, who want to share or want to use data, it is essential to know that the quality and integrity of the data is guaranteed. Since 2005 DANS has developed the so-called Data Seal of Approval (DSA)²⁹; in 2009 DSA was handed over to an international board. DSA prescribes 16 minimum requirements concerning the creation, storage, use and reuse of digital data, that data archives have to meet. Data archives can perform a self-assessment through an online tool on the DSA website. These self-evaluations are reviewed by the DSA board, before the DSA seal label is issued. In October 2014 there are 36 repositories with DSA seal, including EASY. Furthermore, there are another 35 DSA seals requested.

DSA is part of the European Framework for Audit and Certification of Digital Repositories. The Data Seal of Approval offers a basic certification, the next level is extended certification DIN 3164430 and the third level is the formal certification according to ISO 1636331.

DANS in a national and international context

National context

Research Data Netherlands (RDNL) is a coalition between DANS, 3TU.Datacentrum and SURFsara. The mission of this cooperation is the promotion of sustained access and responsible re-use of digital research data. RDNL provides training for librarians (Essentials 4 Data Support

²⁹ See for an overview of DSA:

http://datasealofapproval.org/media/filer_public/2014/10/03/20141003_dsa_overview_defweb.pdf

³⁰ http://www.langzeitarchivierung.de/Subsites/nestor/EN/nestor-Siegel/siegel_node.htm

³¹ <http://www.iso16363.org/>

training) and awards every two years the Dutch Data Award to researchers who gather and document data, and make it accessible to others.

With NARCIS, DANS cooperates with all the Dutch universities, the Academy, the National Research Council, the National Library, and a number of research institutes to make datasets, open access publications and an overview of research projects, researchers and research institutes accessible for the academic world, the media, policy makers et cetera.

In the Netherlands DANS contributes to the research infrastructures of the national roadmap CLARIN.NL³², European Social Survey (ESS)³³, CLARIAH³⁴ and The Language Archive³⁵.

International context

The aim of DANS is to be a leading building block of the data infrastructure in Europe. In October 2014 DANS participates in data activities of 21 international and European research infrastructures, and is an active contributor to a substantial number of European research infrastructure proposals for Horizon 2020.

Examples of the international research infrastructures in which DANS is involved are CESSDA, CLARIN ERIC and DARIAH ERIC. European projects in which DANS takes part are Data without Boundaries, EHRI, DASISH, APARSEN, OpenAirePlus, NeDiMAH, EUDAT, ARIADNE, Europeana Cloud, 4C, EuroCRIS, and IMPACT-EV³⁶.

Furthermore, DANS is member of the Research Data Alliance (RDA)³⁷ and organised the last RDA Plenary in September in Amsterdam. The goal of the RDA is that researchers around the world share and use research data without barriers. The Research Data Alliance wants to accelerate international data-driven innovation and discovery by facilitating research data sharing and exchange, use and re-use, standards harmonization, and discoverability.

References

Data Archiving and Networked Services - DANS. *A federated data infrastructure for the Netherlands: The front office - back office model*. Den Haag: Data Archiving and Networked Services, October 2013. 7 p.

High Level Expert Group on Scientific Data. *Riding the wave: How Europe can gain from the rising tide of scientific data*. Final report of the High Level Expert Group on Scientific Data. A submission to the European Commission. European Union: October 2010. 40 p.

DILLO, Ingrid and DE LEEUW, Lisa. Het Data Seal of Approval: keurmerk voor duurzame en betrouwbare databewaarplaatsen. In *Handboek Informatiewetenschap en in de IWA-database*. Alphen aan de Rijn: B+B Vakmedianet, 2014. 29 p. Translated in English: Data Seal of Approval: Certification for sustainable and trusted data repositories. The Hague: DANS, 2014. 20 p.

³² <http://www.clarin.nl/>

³³ <http://www.europeansocialsurvey.org/>

³⁴ <http://www.clariah.nl/>

³⁵ <https://tla.mpi.nl/>

³⁶ See for an overview of projects: <http://www.dans.knaw.nl/content/projecten> (in Dutch)

³⁷ <http://rd-alliance.org/>