Dare to share! How the DANS Data Station SSH supports FAIR data in the ODISSEI Community

Short article appeared on the ODISSEI website on 22/09/2023[link]

Author	Organisation	ORCID
Ricarda Braukmann	Data Station Manger Social Sciences and DANS.	0000-0001-6383-7148

Brief summary: In the beginning of June DANS launched the DANS Data Station Social Sciences and Humanities (SSH). Researchers and data stewards from the SSH domain can use the Data Station to archive and publish data as well as to find data for reuse. This blog post outlines the features of the Data Station and how it can be used to make data reusable for new research while allowing to restrict data access where needed.

Data Archiving and Networked Services (DANS) is the national center for expertise and repository for research data. For more than 15 years DANS exists with a mission to make research data reusable. In this role DANS provides data archiving and publishing services for the Dutch research community with a long history in providing services for the social sciences and humanities (SSH) domain. The repository system - called EASY - that DANS has been using for many years is now being replaced by a new system based on open source software and tailored to a specific domain. In June 2023,

Dare to share! How the DANS Data Station SSH supports FAIR data in the ODISSEI Community

Short article published on the ODISSEI website

DANS launched the [DANS Data Station Social Sciences and Humanities](#) - the place where Dutch researchers and data stewards from the SSH domain can archive and publish their data, storing it for the long term and making it available for others.

A trustworthy digital repository

The DANS Data Station SSH can be considered a trustworthy digital repository. Now what does it mean? A digital repository is a place where (research) data can be stored and information about the data can be found to facilitate its reuse. A *trustworthy* digital repository, or TDR, is a digital repository that has gone through a process of certifying the procedures and policies that the repository has in place. This certification process ensures that the repository follows archiving and community-specific standards. A user can be sure that the data is managed well, curated, stored safely and that the repository has procedures in place to ensure data availability over a longer period of time. In contrast to generic repositories that are not certified, like for instance Zenodo which does not curate the deposited information and does not promise long term preservation, a trustworthy digital repository like the DANS Data Station SSH provides more elaborate services ensuring good data quality and availability of the data for the long term.

As open as possible as closed as necessary

DANS supports the Open Science movement which provides a framework that puts knowledge sharing, collaboration and transparency at the centre of our work. The Open Science vision is that research papers and results, as well as research data, software and tools are made available for reuse by others so that everyone can benefit from the knowledge that is produced.

A common misconception is that following the Open Science principles means that all data needs to be shared openly for anyone to access and re-use. This is, however, not the case. The Open Science community acknowledges that there are valid reasons why data cannot be shared openly. Datasets containing sensitive information about individuals, companies or the environment that could potentially be misused and harmful need to be protected!
This is the reason why the DANS Data Station SSH - like many TDRs - offers the possibility to restrict access to datasets. Data published under restricted access can be downloaded only after a user has received permission from the data owner. In this

Dare to share! How the DANS Data Station SSH supports FAIR data in the ODISSEI Community

Short article published on the ODISSEI website

way, the researcher or data steward depositing data remains in control and can decide whether the data can be released to a particular user.

In the spirit of Open Science, the principle to follow is to publish data *as open as possible* and *as closed as necessary*. DANS provides guidance on the depositing process and how to choose access rights and licences through [a manual for depositing data](). Recently,additional guidance was published particularly addressing the challenges that come with [working with qualitative data](). The guide describes different aspects to consider when working with qualitative data and the various options available to make data reusable even if it cannot be openly shared.

How you can use the Data Station to make your data FAIR

The DANS Data Station SSH is designed to make it easier for you to deposit data and make it available for reuse, either openly or with access restrictions as discussed above. The Data Station has several features that help you to make data Findable, Accessible, Interoperable and Reusable (FAIR).

Findable
For datasets to be findable, you need to provide sufficient metadata and the Data Station SSH provides various fields to describe the data in a lot of detail. Every field is accompanied by a help text to guide you in providing the information. Importantly the Data Station SSH offers SSH-specific metadata, where you can add methodological information, as well as keywords specifically relevant for your peers to understand how the data was collected and to be able to find it quickly. This includes for instance the European Language Social Sciences Thesaurus ([ELSST]()) and topics from the [CESSDA Topic Classification](). The findability of your data in the Data Station SSH is enhanced as the metadata can be harvested by external portals and aggregators. This way the metadata is for instance made available in the [ODISSEI Portal]()for Dutch researchers and internationally in the [CESSDA Data Catalogue]().

Accessible
For datasets to be accessible, your dataset needs to have a Persistent Identifier (PID). A PID is "a long-lasting reference to a document, file, web page, or other object" (Wikipedia, 2021, see [this blog post]()). The DANS Data Station SSH provides PIDs, specifically DOIs, for every dataset. With that PID you can always retrieve your dataset and access the metadata that you have provided about your dataset. The metadata is openly accessible and includes information about the licence and the access

Dare to share! How the DANS Data Station SSH supports FAIR data in the ODISSEI Community

Short article published on the ODISSEI website

conditions that you may want to apply to your dataset. As outlined above, the access category for a dataset can either be open access or restricted access for datasets that need protection. If you choose restricted access for your dataset the Data Station lets you handle requests easily in the system.

Interoperable

Interoperability is likely the most complex aspect of the FAIR acronym. It is defined as "the ability of data or tools from non-cooperating resources to integrate or work together with minimal effort" (Wilkinson et al., 2016) . You realise interoperability of your dataset by using standards to describe and organise the (meta)data. Using PIDs to refer to datasets is an important feature as well as using available controlled vocabularies that allow for standardised annotation of datasets. DANS has a list of preferred formats and encourages everyone to archive data in open formats that are independent of certain types of software. If you cannot provide a preferred format yourself directly, DANS also converts data into preferred formats to ensure the long term availability of the data.

Reusable

Using open formats also enhances the reusability of your data. DANS supports you in the depositing process to ensure that the data is well documented and available in formats that make reuse as easy as possible. If you use the standardised licences and access categories that the DANS Data Station SSH provides, you ensure that users know how the data can be reused. For anyone who wants to find datasets in the DANS Data Station SSH, DANS provides a manual for reusing data.

The DANS Data Station SSH currently features more than 7000 datasets from a variety of topics. This includes survey data, public use files, for instance from CBS, as well as a large collection of Oral History interviews.
DANS invites everyone to have a look at our data station at https://ssh.datastations.nl and check out our depositing manual to add data to our collection.

For questions anyone can reach out to DANS and from August 21st onwards there will be a weekly open hour again held on Monday mornings from 10:00-11:00.

References

Wikipedia (2021). Persistent Identifier.
https://en.wikipedia.org/wiki/Persistent_identifier

Dare to share! How the DANS Data Station SSH supports FAIR data in the ODISSEI Community

Short article published on the ODISSEI website

Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

Relevant links

- The data station is available at https://ssh.datastations.nl/
- The data Station SSH is harvested by the ODISSEI portal where metadata from various Dutch data providers is available: https://portal.odissei.nl/
- This video gives a short explanation of the main features of the data station: https://youtu.be/6hwZxTFmnS0
- If you have any questions about the Data Station or Open Science and FAIR data in general, visit our Open Hour on Monday mornings (starting again after the summer on August 21st) : https://dans.knaw.nl/en/agenda/open-hour-ssh-live-qa-on-monday-2/

Featured image by DANS.