

# DANS Data Stations, DataverseNL and the Vault

Presentation for the UKB working group Research Data

Marion Wittenberg, Service manager DataverseNL

Tuesday December 14, 2021, online

# About DANS

- DANS is the national expertise center and repository for research data.
- We help researchers to make their data available for reuse. In this way scientists can use the data for new research and published research is verifiable and repeatable.
- With more than 150,000 data sets and sixty employees, DANS is one of the leading repositories in Europe.
- DANS is an institute of the Royal Netherlands Academy of Arts and Sciences (KNAW) and the Netherlands Organization for Scientific Research (NWO).

# EASY and DataverseNL

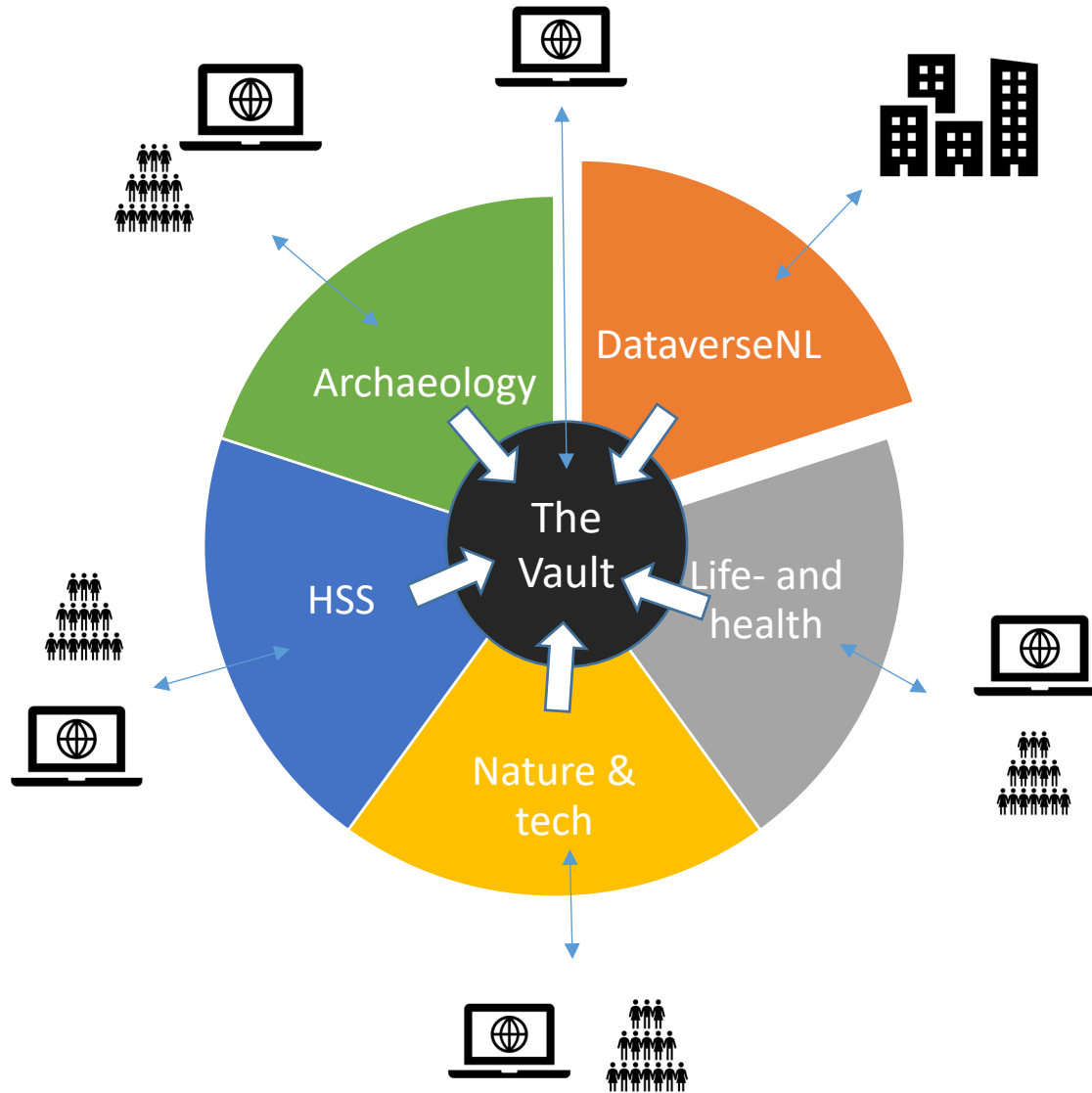
Until now: Two data repositories

EASY: Discipline independent repository

DataverseNL: Platform service for institutes that want to establish their own repository

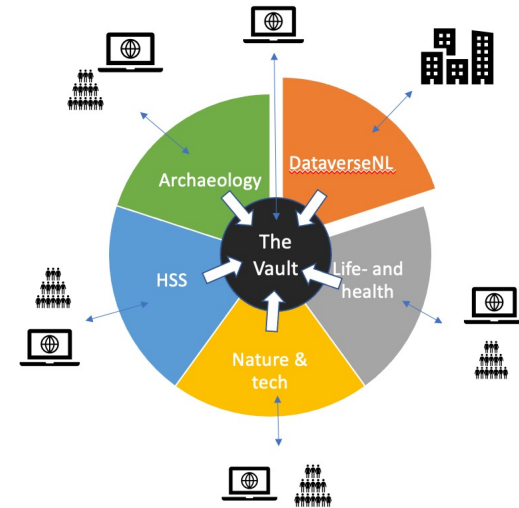
EASY needed a major technical upgrade

# DANS Services



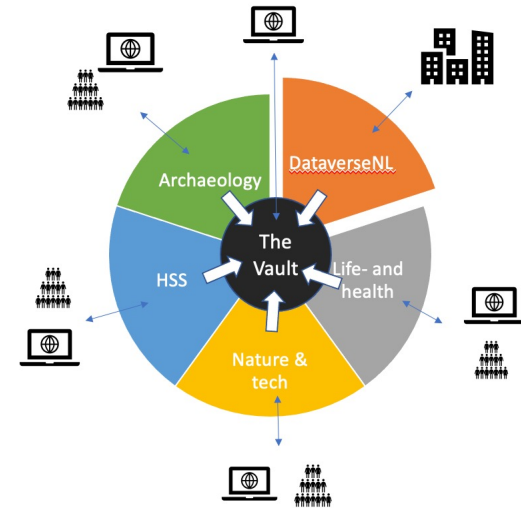
# DataverseNL (basic principles)

- Platform service for institutes that want to establish their own repository.
- The institutes form a user community, which together with DANS determines the (further) development.
- Hybrid financing model: KNAW basic financing + subscription fee.
- The metadata of the datasets in DataverseNL can also be found via the Data Stations.
- The datasets in DataverseNL are automatically sent to The Vault for long-term storage



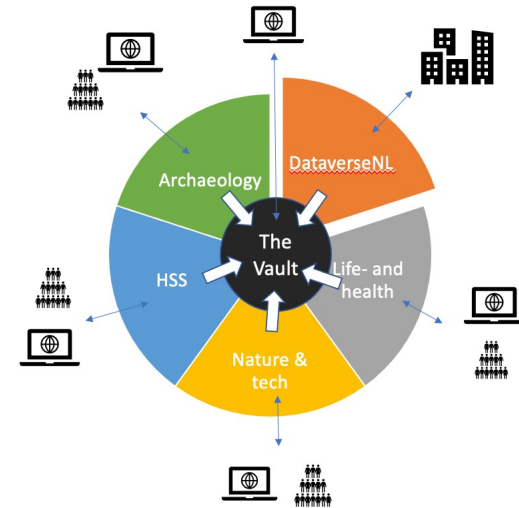
# DANS Data Stations (basic principles)

- Research communities have a need for digital environments focused on their specific research. The needs of different communities can differ.
- In addition to safe and reliable archiving of finished datasets, DANS also offers the option of working with 'living' datasets.
- The services must be reliable and certifiable, in combination with each other, or in combination with external services.
- A station is a complex of services, with Dataverse instance at its core.
- The datasets in the Data Stations are automatically sent to The Vault for long-term storage



# The DANS Data Vault

- Service for Long Term Preservation
- Connected to Data Stations and DataverseNL
- Also available for external repositories who need a LTP solution
- Tape storage



# Comparison of Data Station and dataverseNL

## *Can you explain how DataverseNL compares to the other Data Stations?*

Both the DANS Data Stations and DataverseNL are built with the dataverse software from Harvard University.

- DataverseNL is a platform service for institutes that want to establish their own repository, DANS manages the technical infrastructure, The participating institutes themselves are responsible for the content, data curation and the related work processes.
- For the DANS data Stations, DANS is responsible for the content and data curation.



# Retention period

*DataverseNL is seen as mid-term storage (10 years). Will this change with this new construction?*

This ten year period is no longer correct. Since DataverseNL switched to DOIs as a Persistent Identifier, the 10-year term has been dropped. Because as the word implies, a Persistent Identifier must be persistent.

*What is the retention period for the other Data Stations and the Vault?*

The retention period is also not limited for the Data Stations or the Vault.

Both the data from the Data Stations and from DataverseNL will automatically be pushed to the Vault.

# File formats

*Will anything change regarding (sustainable) file formats?*

For data files that are uploaded in a preferred format, DANS will ensure that these are always accessible. We cannot guarantee this for file formats that are not on the preferred format list.

# Costs and agreements

*What does the new situation mean for the costs and agreements with researchers and/or institutions?*

## Costs

- Initially, the situation will remain as it is now. A contribution from the participating institutes is requested for DataverseNL, because we adapt this repository (within the limitations of dataverse) to their needs. Institutes can also organize their dataverses and use a branding for their own folders.
- No contribution is requested for the DANS Data Stations, although for uploads above 50 GB a contribution will be requested from the depositor in connection with the storage costs.

## Agreements

- For DataverseNL, the agreements are concluded per institution
- For the DANS Data Stations, the agreements are concluded with the individual end users

# Choice of repository

*What should a researcher or institution base his/her choice on when selecting a repository?*

- A researcher should rely on reliability (certification), service and customer friendliness (can you speak to someone in person, for example)
- In addition, it is the policy of the institution which determines in which repositories researchers must or can place their data

# Link between DataverseNL and the Data Stations/Vault

*Is there a link between DataverseNL and the Data Stations/Vault?*

- All data from DataverseNL and the DANS Data Stations will also be archived in the DANS Vault (Tape storage) for long-term preservation.
- Metadata of DataverseNL will be harvested in the domain-specific Data Stations

# Comparison with other repositories

*How does this solution compare to other repositories like Zenodo, and other discipline specific repositories?*

DANS focuses on Social Sciences and Humanities

Our aim is to complement the already existing services, not to compete with them.

Thank you for your attention!

Questions?

[marion.wittenberg@dans.knaw.nl](mailto:marion.wittenberg@dans.knaw.nl)