Provenance Document: the Processing of Datasets by DANS

Once a dataset has been archived in EASY, we will ensure that the data will continue to be accessible and readable in a sustainable manner. In order to be able to meet this guarantee, several actions are required. You will read more about this in this document. Should you have any other questions about data processing or the reuse of data, please contact DANS at info@dans.knaw.nl.

1. Data Processing Protocol

After having deposited the data in EASY, they will be processed by a staff member at DANS in accordance with a standard data processing protocol. The purpose of this protocol is to ensure that the data will be findable, accessible, and comprehensible in the longer term, the last aspect also without the intermediary of the original researcher. A key element of this protocol is, where applicable, the verification of privacy-sensitive data. This applies, in particular, to survey data and interviews.

On the basis of this protocol, the following types of verification have been performed since the introduction of DANS in 2005:

- Verification of completeness of the dataset, with regard to both the data files deposited and the accompanying documentation files;
- Verification of the readability of the files;
- Verification of the file format. In the future as well, it should still be possible to open and use the data files as well as the documentation files. The verification is performed on the basis of a list of preferred file formats.
- Verification of the description of the dataset for completeness and accuracy, and improvement of the presentation of the description.
- Verification of the presence of privacy-sensitive data, both in the files and in the metadata (see the following section for further information)
- Verification of the clarity of the directory structure. If this structure is not sufficiently clear, it will be adjusted.
- Verification of completeness and correctness of the list of files in archaeological datasets submitted by the depositor. This list includes a short description of each file of the dataset.

In order to improve the clarity of a dataset, an archivist may apply minor changes to the metadata or the directory structure. Larger issues will be consulted with the data depositor.

During the archival procedure, an archivist may migrate files to preferred formats to ensure long-term preservation and accessibility. If files are migrated, the migrated files will be published with the dataset for use. The original files, as deposited, will always be archived with the dataset.

All actions performed on a dataset by a DANS employee will be registered in internal administration.

2. Privacy-Sensitive Data

If a file includes exact names and exact dates of birth of the respondents, these variables will be deleted. Exact contact data of the respondents will also be deleted, and only the digits of the postal codes will be maintained. Exact job names will not be made available either. It is, however, possible to deduce the respondents' jobs from job classifications. The general rule is consequently that all identifiable variables will be deleted. In this way, the datasets with privacy-sensitive data will be available only in anonymous formats. The non-anonymous datasets are still archived: this is done with future analysis options in mind. Qualitative data, such as audio files of interviews, cannot be made anonymous, but these data will be available as the interviewee has officially consented to the use of the interview for scientific research.

3. Persistent Identifier

Upon archiving, an automatically generated Persistent Identifier — composed of a combination of digits and letters — is attached to each dataset. This Identifier may be included in the references to the dataset in publications about analyses of the dataset. The Persistent Identifier guarantees permanent findability of the dataset, despite changes in internet addresses.

4. Access to Dataset

The rights of the users to view or download the files of a dataset are assigned by the depositor upon submission of the description of the dataset. DANS copies these rights upon archiving and making the data accessible.

5. Additional information

Read more on the DANS Preferred Formats at:
http://www.dans.knaw.nl/en/deposit/information-about-depositing-data