
UD Frisian

A Data Management Plan created using DMPonline

Creators: Wilbert Heeringa, First Name Surname

Affiliation: Other

Template: DCC Template

Project abstract:

The UD Frisian treebank is a manually annotated treebank for Frisian. The texts are tokenised, lemmatised, tagged and dependency parsing annotated according to the Universal Dependency guidelines (UD version 2) that are found at: <https://universaldependencies.org/guidelines.html> .

ID: 75227

Start date: 01-01-2018

End date: 31-12-2022

Last modified: 12-04-2021

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

UD Frisian

Data Collection

What data will you collect or create?

Frisian text

How will the data be collected or created?

Texts from Wikipedia and the Oersettercorpus are selected, the words are annotated either manual or semi-automatic.

Documentation and Metadata

What documentation and metadata will accompany the data?

The genres represented by the corpus are mentioned, per genre the number of tokens, words and sentences is given, and per genre the text included are mentioned. A reference to <https://universaldependencies.org/guidelines.html> will be given. ISLE Meta Data Initiative (IMDI) is used as a metadata standard.

Ethics and Legal Compliance

How will you manage any ethical issues?

There are no ethical issues.

How will you manage copyright and Intellectual Property Rights (IPR) issues?

The order of the sentences will be randomized. The data will become open source. The Fryske Akademy is owner. The data will be shared when the treebank is complete.

Storage and Backup

How will the data be stored and backed up during the research?

It will be stored on the server of the Fryske Akademy. Data stored on this server are backed up daily. The data is also mirrored on a server of Tresoar. The system manager is responsible for backup and recovery.

How will you manage access and security?

The data will be accessible via a repository at <https://bitbucket.org/fryske-akademy/> .

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

All data are of long-term value and are retained, shared, and preserved for ever.

What is the long-term preservation plan for the dataset?

The data will be stored in a DANS repository, see: <https://dans.knaw.nl/en> .

Data Sharing

How will you share the data?

Via: <https://bitbucket.org/fryske-akademy/> and via DANS. The data becomes available when the treebank includes at least 100,000 words that are completely annotated according to the UD guidelines.

Are any restrictions on data sharing required?

No.

Responsibilities and Resources

Who will be responsible for data management?

Wilbert Heeringa

What resources will you require to deliver your plan?

The data is open source, and will become freely available, no charges will be applied.