Preservation Plan Taalmaterialen at INT

1. Scope of the plan

This plan is restricted to language data in the Taalmaterialen archive. It does not pertain to data that is produced by INT in the regular work processes and other activities.

The preservation plan is based on a variety of external guidelines and standards for digital preservation such as OAIS, the FAIR principles, CoreTrustSeal (formerly Data Seal of Approval), nestorSeal (based on DIN 31644) and ISO 16363).

2. Mission of the archive

An important part of the mission of the Dutch Language Institute (Instituut voor de Nederlandse Taal – INT) is providing long-term access to Dutch source material in the form of historical and contemporary corpora, dictionaries, lexical digital databases, grammars, including the required technical tools. Apart from the data we produce ourselves we also accept resources from other organisations.
More information:
- Information about deposition
- Dutch Language Institute

3. Objectives of the archive

The task of the archive is to make these materials mentioned in the previous section available to third parties, including the scientific community, commercial enterprise and the general public and to make sure that they remain accessible and usable in the long-term.

4. Tasks

It is important to identify factors that can jeopardize these objectives.
The main tasks of the archive is to remedy these jeopardies.

4.1 obsolete data formats

We advice contributors to use standards from our list of preferred formats as much as possible (see Information on Deposition). If necessary, staff will convert data for distribution purposes.
Every two years the archive is scanned for obsolete formats and, if necessary, formats are replaced.

4.2 legal claims that make the data unaccessible to the intended audience

At ingest we check legal clearance of the data. The Data Protection Officer checks accordance with all privacy regulations. Contributors of data sign a deposit agreement. Appropriate licenses are attached to the products in consultation with the contributors.

4.3 data loss through calamities

System management of the INT is responsible for the technical infrastructure. The functioning and condition of the technical components is permanently monitored.

Maintenance takes place according to extensive documentation and numerous Standard Operating Procedures and checklists.

The procedures include: Weekly full back-ups to tape. The full back-ups are followed by incremental back-ups on a daily basis. Quarterly full back-ups to tape, to be stored at another location.

These tapes will eventually be reused, but one set of each year is retained for at least seven years. A restore can be carried out upon request. Essential information for disaster recovery is stored on paper and digitally in a vault at a great distance from the data centre.

Installation of security patches and updates on a monthly basis.

4.4 loss of data integrity

On ingest, the original files are preserved. Modifications required for distribution are documented and preserved.

Checksums are calculated of all distributed versions of products.

Change logs are kept recording all modifications by INT of the archived products.

4.5 Oblivion

It is often said that data is better preserved by being used than by being stored. Therefore we make an effort to stimulate the use of our products.

- By providing adequate metadata and advertising the content of our archive through metadata aggregators like CLARIN VLO
- By providing a search facility to the Taalmaterialen website.
- By providing PID's to all products and by maintaining those PID's.
- By running a servicedesk that assists users in selecting and using products.

4.6 loss of interpretability of the data

At ingest we check if the provided documentation allows sufficient understanding of the data. If necessary additional documentation is added.

Furthermore, the servicedesk with specialised staff maintains a knowledge base related to the data products, including all communications with the contributors of the data.

5. Roles and responsibilities

The prime responsibility lies with the director of Dutch Language Institute.

There are several roles associated with the INT Taalmaterialen archive.

5.1. Project manager

The project manager oversees all archiving operations. Negotiates submission agreements with Producers. Establishes policies for maintenance, preservation, standards and hardware and software planning, customer support, appoints staff to procedures, monitors the workflows, etc.

5.2. Acquisition management

The acquisition manager is responsible for ingest of data. This procedure is not automated. It therefore involves direct communication with depositors in which they together run through the check-list. In that process also legal issues will be cleared.

5.2. Data management
The data manager is responsible for content integrity of the data. He sees that data and descriptive information is connected and is responsible for version management. This involves e.g. allocation of proper storage locations for data, looking after data integrity, version control, format conversions, and maintenance of the metadata set.

5.3. Data Protection officer
The DPO reviews ingested data according to standards of privacy protection.

5.4. Servicedesk staff
Advises and assists users of the data archive.

5.5. System manager
The system manager is responsible for the functioning of the technical infrastructure