
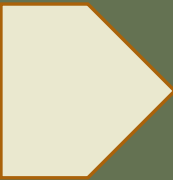

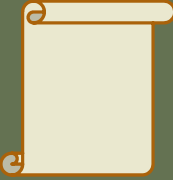


# Store it in a cool dry place

## Processing and long-term preservation of research data

The Finnish Social Science Data Archive is a national service resource for scientific research and teaching funded by the Ministry of Education and Culture. We archive digital research data from Finland and abroad and disseminate it for research, teaching and learning purposes.

Making research data accessible and reusable means preserving it properly. The recipe is heavy on metadata, and the data need to be carefully and systematically prepared for preservation. Without the right tools and a cool, dry place for storage, the mission is in jeopardy from the beginning.

TASK / PHASE	Task: choosing sustainable file formats, migrating existing content, and updating data processing policies and software accordingly	Task: creating archival packages including data and metadata and producing METS containers for transfer to preservation system	Task: Influencing and adapting to (inter)national preservation and documentation standards and services.	Training the data managers and IT- staff and carrying out related administrative tasks
Present state 	24 file formats in preservation, some unfit for long-term preservation  Rules have evolved over the years and current practises differ from earlier approaches  No major migrations	Archival packages created mostly manually, with a help from set of in-house tools  Operational database on provenance data  Own disk storage for data files  AIP = often SIP (no METS)	Ten years of involvement in national preservation efforts for cultural heritage data and research data  Compliance with OAIS and FAIR Data Principles	The Preservation Planning Team in charge of the preservation policy  Individual data managers not directly involved with all preservation questions
Goals 	Preservation policies are defined on the object level and maintained in a database  Changes in file formats are monitored and open formats preferred  Files do not include unnecessary private information	Archival packages created automatically as METS containers based on technical and provenience metadata  National Long-Term Preservation Solution for AIPs  Audited shared preservation responsibility	Leading expert role on a national level regarding SSH data  EU-level collaboration  Seamless integration standards and services to ensure sustainable long-term access	Regular training and quality control  Introducing new tools to make certain workflows less error-prone and more effective (e.g. PDF/A validation)
Steps taken 	Ongoing migration from RTF to ODT, PDF to PDF/A, XLS/DAT to CSV  Produced a recommend format list for ingest and delivery  Automated object level preservation status recognition	In-house system in place as an interim solution for creating AIPs  Transfer system and monitoring partly manual processes  More granular implementation for the new metadatabase needs to be designed	Partnered with a national preservation solution  DDI development, METS compl.  SPSS Portable submitted to PRONOM registry, PIDs for data  Aligning curation processes with (inter)national standards	Raising general awareness  Internal handbook updated with file formats information and guidelines on preservation status of data files  Data security and protection training and regular audits
Lessons learnt 	Nearly 17 years of activity have produced some unmonitored noise. Cleaning takes time and is not always straight-forward  Migration may lead to unexpected results  Exceptions are likely to happen	Need to decide which objects are preserved and which can be recreated easily  Need to keep in-house code safe too, or data replication may not be possible  One never has enough metadata	Field requiring standards and definitions is vast. So is the number of standards and definitions  Target is also moving fast “preserving a running train”  We know stuff and are a valued partner	Collaboration with IT and data managers is often needed; both make expert decisions  Privacy issues must be taken into consideration but anonymization procedures may make files unsuitable for long-term preservation (e.g. XMP-metadata)



- How is the change over time monitored?
  - When was the last migration? Was something lost in the process?
  - Do tools or needs dominate?
  - Is there a preservation policy?
- How is the code maintained?
  - Who stores your data?
  - How trustworthy they are?
  - Is there enough provenience metadata?
  - Is file size an issue?
  - How are the data versions maintained?
- What are the key standards and are they adhered to?
  - Is there sufficiently internal expertise?
  - Who would be your partner and what is the collaboration model (bilateral, distributed, institutional, commercial..)?
- How are data management and preservation skills maintained?
  - Are there administrative or legal hindrances?
  - Are instructions maintained and is there a technology watch in place?
  - Policy-based or ad hoc - solutions?

