LaVaH

Distributed long-term preservation with diverse roles and responsibilities

Natascha Schumann

*hebis Verbundzentrale,
Goethe University Frankfurt
Germany*

schumann@hebis.uni-frankfurt.de

https://orcid.org/0000-0001-7897-9616

Martina Sinkovic

*hebis Verbundzentrale,
Goethe University Frankfurt
Germany*

sinkovic@hebis.uni-frankfurt.de

https://orcid.org/0000-0002-4558-8974

Abstract – Distributed long-term preservation with diverse roles and responsibilities is the goal of the LaVaH project. The poster describes the division of work between stakeholders from the university libraries, the data centers, and the service provider for library infrastructure. Data management is done locally at the universities, central digital archiving tasks by the service provider (hebis) and storage infrastructure is hosted by two data centers.

Keywords – collaboration, data management, distributed preservation, metadata

Conference Topics – 2.3 Sub-theme 3: Enhancing the Collaboration

I. INTRODUCTION

The project „Langzeitverfügbarkeit an Hessischen Hochschulen (LaVaH)" provides long-term accessibility for the universities of Hessen and is funded by the Hessen State Ministry for Higher Education, Research and the Arts (HMWK). The goal is the development of an infrastructure for the long-term preservation of digital resources. Currently the project deals with text-based file formats as well as images and audio-visual files.

In collaboration with the universities, we will progressively build up a distributed infrastructure for long-term preservation. This does not only mean physical storage of the digital objects but includes division of responsibilities and tasks. All parties use the same archiving system but have different workflows for their holdings.

II. DISTRIBUTED DIGITAL PRESERVATION

Based on the conviction that long-term preservation should be a common effort, the tasks were divided in the following three sections:

Data management is done locally at the universities. hebis VZ, the service provider for Hessian universities and state libraries, is responsible for central archiving tasks, including ingest, central data management as well as preservation planning. It hosts the archiving software (Rosetta). The storage infrastructure is hosted by the data centers of the universities of Frankfurt and Darmstadt.

Local data management includes the selection of collections that should be preserved. Local data managers at the universities take care that there are no legal restrictions and they ensure that only recommended data formats are used. If the original document is not available in suitable format it may be converted before transferring it to the archival system. As it requires an enormous effort to correct or change data after submission to the archiving system it seems to be useful to start at an early stage. Best would be – obviously – just when producing a digital object. That is not always possible. Local data managers may validate data formats and – in case they are located in the university library - they can affect the way data producers deliver their work in terms of regulations of formats and other requirements that enable long-term preservation.

As agreed with the hebis VZ and according to further processing, local data managers build a transfer package that contains the data and a metadata file. Depending on their needs, that metadata file might have a METS or DC or some other structure. The main point is that this package conforms the policy both parties agreed on.

The role of hebis VZ is to manage all the preservation related issues. The affiliates build up knowledge about the archiving system and communicate the requirements to the local data managers. Together with the partners appropriate workflows for different collections were chosen, developed and tested. Transferred package(s) from the partners are validated and ingested.

An organizational issue of the hebis VZ is the configuration of contracts and agreements for delivery and archiving.

Even if preservation planning is an (ongoing) task for the hebis VZ, the definition and identification of significant properties should be done by the local data curators. These decisions are made in coordination with the partners and with colleagues from the university archives. As they are more experienced with regard to selection and appraisal, this collaborative effort is very beneficial.

The technical infrastructure of storage and backup is organised by a cooperation of two data centers. It conforms to the requirements of data security and redundancy.

III.   CONCLUSION

The project and its structure with diverse roles on the one hand, and an extensive exchange about metadata, workflows, and significant properties on the other, offers an effective infrastructure for the universities. All partners could focus on their special expertise and benefit from the others´ expertise. It offers not only a common understanding of digital preservation but also supports a digital preservation infrastructure that goes beyond technical sharing of an archival system. Our goal is to establish a multi-disciplinary Hessian digital preservation consortium, where stakeholders from the libraries, university archives, data centers, infrastructure and others work together.

The next steps will be to set up a subsequent project with focus on more automated workflows for SIP generation, metadata file creation and validation. It will include universities of applied sciences, as well.