

Kooperative Langzeitarchivierung in Hessen: LaVaH

Natascha Schumann, hebis Verbundzentrale, Goethe Universität Frankfurt am Main

Zusammenfassung

Im Projekt LaVaH – Langzeitverfügbarkeit digitaler Inhalte an hessischen Hochschulen – arbeiten die hessischen Universitäten und Hochschulen und die hebis Verbundzentrale an der arbeitsteiligen Langzeitarchivierung digitaler Daten. In einer verteilten Infrastruktur übernehmen die lokalen Datenkurator*innen vor Ort die Kuratierung der zu archivierenden Objekte und erstellen Transferpakete. Die zentrale Archivierung erfolgt durch die hebis Verbundzentrale. Die Speicherinfrastruktur wird von den Hochschulrechenzentren der Goethe-Universität Frankfurt und der TU Darmstadt zur Verfügung gestellt, die eine ortsunabhängige redundante Speicherung gewährleisten. Darüber hinaus wird spartenübergreifend mit einigen hessischen Universitätsarchiven zusammengearbeitet und langfristig eine Übersicht über die Bestände an hessischen Hochschulen erstellt. Das Projekt wird vom Hessischen Ministerium für Wissenschaft und Kunst (HMWK) gefördert.

Summary

In the project „Long-term availability of digital content at Hessian universities (LaVaH)”, Hessian universities and colleges work together with the head office of the hebis library network to provide a work-sharing long-term preservation of digital data in Hesse. Based on a distributed infrastructure, the local stakeholders are responsible for curating and selecting objects for preservation and creating the transfer packages. The archiving is carried out centrally by the hebis headquarters. The storage infrastructure is provided by the computing centers of the Goethe University Frankfurt and the TU Darmstadt, which ensure a location-independent and redundant storage. In addition, there is a cross-sectoral cooperation with some Hessian university archives. In the long term, an outline of the holdings at Hessian universities will be compiled. The project is funded by the Hessian Ministry of Science and Art.

Zitierfähiger Link: <https://doi.org/10.5282/o-bib/5887>

Autorenidentifikation: Schumann, Natascha: ORCID: <https://orcid.org/0000-0001-7897-9616>

Schlagwörter: Langzeitarchivierung, Langzeitverfügbarkeit, Verteilte Infrastruktur, Kooperation

Dieses Werk steht unter der [Lizenz Creative Commons Namensnennung 4.0 International](#).

1. Einleitung

Langzeitarchivierung (LZA) ist ein ressourcenaufwändiger und wiederkehrender Prozess, der sicherstellt, dass digitale Daten auch in (einer nicht näher definierten) Zukunft noch korrekt nutzbar sind. Das bedeutet, dass die Daten nicht nur unversehrt sein müssen, sondern auch noch mit geeigneten Werkzeugen korrekt darstellbar und nachnutzbar sind. Dazu braucht es eine gute Vorbereitung der

Daten, bevor diese in ein Langzeitarchiv überführt werden. Ebenso wichtig ist eine regelmäßige Überprüfung der Daten und gegebenenfalls die Überführung in andere, aktuellere Dateiformate.

Langzeitarchivierung muss unterschieden werden von einer Speicherung und Aufbewahrungsfrist von zehn Jahren (wie von der DFG empfohlen). Langzeitarchivierung bezieht sich auf einen nicht näher definierten Zeitraum in der Zukunft, der über die oft zitierten 10 Jahre weit hinausreicht.

„Langzeit‘ ist die Umschreibung eines nicht näher fixierten Zeitraumes, währenddessen wesentliche nicht vorhersehbare technologische und soziokulturelle Veränderungen eintreten, die sowohl die Gestalt als auch die Nutzungssituation digitaler Ressourcen in rasanten Entwicklungszyklen vollständig umwälzen werden.“¹

Es gilt also, viele verschiedene Aspekte zu berücksichtigen, um eine langfristige Nutzung von digitalen Objekten zu gewährleisten. Technologischer Wandel im Hinblick auf Soft- und Hardware ist dabei ebenso zu berücksichtigen wie ein Wandel im Verhalten von Nutzenden.

An Universitäten und Hochschulen für angewandte Wissenschaften werden sehr unterschiedliche digitale Objekte erstellt, angeschafft und bewahrt. Nicht für alle gibt es bereits Regelungen, wie diese langfristig gesichert werden, z.B. durch Überführung in ein fachliches Langzeitarchiv² oder durch die Ablieferung an die Deutsche Nationalbibliothek. Daher erscheint es sinnvoll, dass sich Einrichtungen zusammenschließen, um gemeinsam Standards zu etablieren und ressourcenschonend ein gemeinsames Langzeitarchiv zu nutzen.

Vor diesem Hintergrund ist in Hessen das Projekt LaVaH³ – Langzeitverfügbarkeit digitaler Inhalte an hessischen Hochschulen – entstanden. Das Projekt ist bislang vom Hessischen Ministerium für Wissenschaft und Kunst gefördert und befindet sich inzwischen in der zweiten Projektphase. In der ersten Projektphase (2019–2021) wurden mit den hessischen Universitäten⁴ prototypische Workflows entwickelt und getestet. In der zweiten Phase (2022–2024) sind die hessischen Hochschulen für angewandte Wissenschaften⁵ dem Projekt beigetreten. Die prototypischen Workflows werden erprobt und die Ergebnisse des Projekts sollen in einen langfristigen Dienst überführt werden.

1 Schwens, Ute; Liegmann, Hans: Langzeitarchivierung digitaler Ressourcen, in: Kühlen, Rainer; Seeger, Thomas; Strauch, Dietmar (Hg.): Grundlagen der praktischen Information und Dokumentation, 5., völlig neugefasste Ausg. München 2004 (Handbuch zur Einführung in die Informationswissenschaft und -praxis 1), S. 567–570. Online: <<http://nbn-resolving.de/urn:nbn:de:0008-2005110800>> (Die Zählung der Abschnitte stimmt nicht mit der Druckausgabe überein).

2 Hiermit sind nicht fachliche Repositorien gemeint, sondern dezidiert digitale Langzeitarchive mit entsprechend dokumentierten Workflows. Während Repositorien i.d.R. zum Publizieren von Daten genutzt werden, stellen digitale Langzeitarchive die langfristige Nutzung durch verschiedene Erhaltungsmaßnahmen sicher.

3 <<https://www.lavah.de>>, Stand: 19.07.2022.

4 Technische Universität Darmstadt, Goethe Universität Frankfurt am Main, Justus-Liebig-Universität Gießen, Universität Kassel und Philipps-Universität Marburg.

5 Hochschule Darmstadt, Frankfurt University of Applied Sciences, Hochschule für Bildende Künste Städtelschule Frankfurt, Hochschule für Musik und Darstellende Kunst Frankfurt, Hochschule Fulda, Technische Hochschule Mittelhessen, Deutsches Dokumentationszentrum für Kunstgeschichte – Bildarchiv Foto Marburg und Hochschule RheinMain Wiesbaden.

Die beteiligten Einrichtungen arbeiten zusammen mit der hebis Verbundzentrale an der Langzeitverfügbarkeit ihrer Daten. Die jeweiligen Anforderungen der Institutionen, z.B. im Hinblick auf vorhandene Metadaten werden berücksichtigt. Gemeinsamkeit wird nicht nur über die Nutzung eines zentral genutzten Archivsystems hergestellt, sondern auch durch die Absprache über Workflows und Standards (z.B. in welcher Form Daten abgeliefert werden) und über die gemeinsame Erarbeitung von Tools und Best Practices in den verschiedenen Arbeitsgruppen.

2. Verteilte Infrastruktur

Bei einer verteilten Infrastruktur denkt man oft an digitale Objekte, die an unterschiedlichen physischen Orten gespeichert, aber zentral nachgewiesen werden. Bei LaVaH bezieht sich die verteilte Infrastruktur weniger auf die technische Organisation der Archivierung, sondern auf die Arbeitsteilung in den Bereichen Pre-Ingest⁶ und Ingest⁷. Die Archivierung erfolgt in einem gemeinsamen Archivsystem.

An den beteiligten Einrichtungen sind lokale Datenkurator*innen dafür zuständig, die Daten, v.a. Texte und Digitalisate, aber auch Audio- und Videomaterialien für die Aufnahme in das Langzeitarchiv vorzubereiten. Diese Vorbereitungsphase (Pre-Ingest) beinhaltet auch die Auswahl der zu archivierenden Daten. Dies geschieht gemeinsam mit den Leitungen der Institution und wird beratend durch die hebis Verbundzentrale begleitet. Die ausgewählten Daten müssen die rechtliche Voraussetzung für eine langfristige Speicherung mitbringen. Die Datenkurator*innen entscheiden, in welchen Dateiformaten die Daten geliefert werden. Des Weiteren kümmern sie sich um die Metadaten. In der Regel sind zu den Objekten bereits Metadaten vorhanden, die teilweise oder ganz übernommen werden können und um spezifische LZA-Metadaten ergänzt werden. Die Metadaten und Daten müssen in abgestimmter Form abgeliefert werden. Im Detail wird das gemeinsam mit der hebis Verbundzentrale entschieden, damit die Datenpakete mit den Vorgaben des Archivsystems kompatibel sind.



Abb. 1: Verteilte Infrastruktur in LaVaH

6 Pre-Ingest meint die Phase vor der Ablieferung von Daten in das Archiv.

7 Ingest meint die Überführung eines Submission Information Packages (SIP) in das Langzeitarchiv.

Diese verteilte Struktur ist hilfreich, weil sie die Stärken der einzelnen Partnerinstitutionen nutzt und dazu beiträgt, gemeinsam bestehende Aufgaben zu lösen, z.B. bei der Formatvalidierung oder der Bestimmung von signifikanten Eigenschaften der zu archivierenden Objekte. Das Konzept der signifikanten Eigenschaften verfolgt das Ziel, diejenigen Eigenschaften eines digitalen Objekts zu bestimmen, die unbedingt über die Zeit erhalten bleiben müssen, um es korrekt nutzen zu können. Diese Eigenschaften hängen unter anderem von der Intention de*r Produzent*in ab und von den möglichen Nutzungszielen und -gruppen. Es gibt Texte, bei denen ist nur der Inhalt von Interesse und andere, bei denen auch die gestalterische Ausführung erhalten werden soll. In einer AG werden Checklisten erarbeitet, die es den Datenkurator*innen erleichtern sollen, die signifikanten Eigenschaften für ihre Bestände zu ermitteln. Dabei werden verschiedene Nutzungsziele, z.B. „Finden (im Sinne von Recherchieren)“ mit entsprechenden Eigenschaften wie „bibliographisch erfasst“ kombiniert.

Die erarbeiteten Workflows ebenso wie vorbereitende Skripte zur Erstellung von Datenpaketen oder Empfehlungen zur Erstellung von Dokumenten in archivtauglichen Formaten kommen allen Projektbeteiligten zu Gute und können dazu beitragen, bereits bei der Entstehung digitaler Dokumente darauf zu achten, dass die langfristige Nutzung sichergestellt wird.

Da die Datenkurator*innen in den jeweiligen Einrichtungen agieren, haben sie eine größere Nähe zu den Datenproduzierenden bzw. Abliefernden als die Betreiber*innen des Langzeitarchivs. Durch Informationen oder auch durch dezidierte Vorgaben kann die Qualität der Daten (z.B. im Hinblick auf langzeitarchivierungsfreundliche Dateiformate) und die der Metadaten verbessert werden. Und je bessere Daten produziert werden, desto einfacher wird die LZA, weil im Nachhinein Veränderungen an Dokumenten vorzunehmen aufwändiger und fehleranfälliger ist.

3. Schwerpunkte der ersten Projektphase: LaVaH I

Eine erste Projektphase, im Folgenden als LaVaH I referenziert, lief von 2019 bis 2021 mit den fünf hessischen Universitäten als Partnern und der hebis Verbundzentrale als Projektleitung.

In LaVaH I waren durch den Projektplan drei Schwerpunkte gesetzt:

- Installation und Konfiguration des Archivsystems,
- Entwicklung von Ablieferungsmodulen für verschiedene Daten und Dateiformate und
- Erstellung einer Ablieferungsvereinbarung.

Bereits vor Projektbeginn hat die hebis Verbundzentrale eine Erhebung durchgeführt, die die Bedarfe der beteiligten Institutionen eruiert hat. Im Anschluss wurden unterschiedliche Lösungen betrachtet und Expert*innengespräche geführt. Als Ergebnis dieser Vorarbeiten wurde „Rosetta“ von ExLibris als technisches Archivierungssystem ausgewählt.

Die Installation und Konfiguration der Rosetta-Software erfolgte durch die hebis Verbundzentrale in Absprache mit dem Hochschulrechenzentrum der Goethe Universität Frankfurt.

Neben einer aus Datenschutzperspektive sicheren Installation der Software stand im Vordergrund, die Bedarfe der Partner möglichst umfassend über die Konfigurationsmöglichkeiten der Software abzubilden und entsprechende Workflows zu installieren. Für jede Institution wurde im Testsystem eine eigene Instanz angelegt, so dass die Daten jeweils getrennt vorliegen und dieser Bereich für jede Institution entsprechend konfiguriert werden kann.

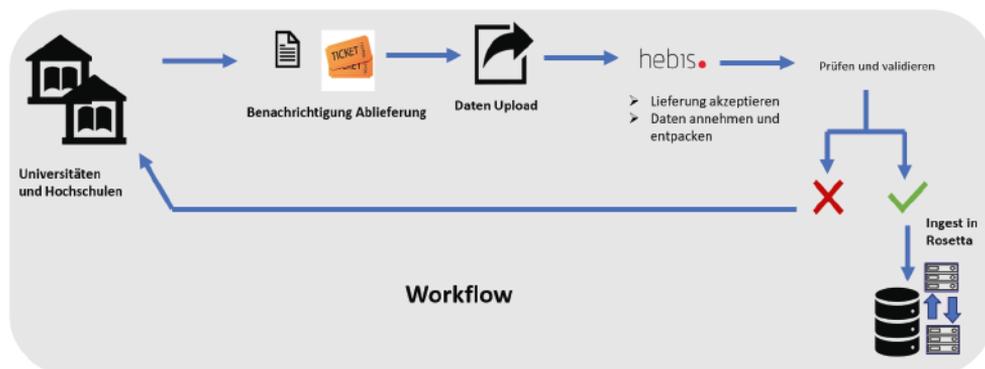


Abb. 2: Prototypischer Workflow

Für verschiedene Dateiformate wurde mit der Entwicklung von Ablieferungsmodulen begonnen. Die Projektbeteiligten teilten sich hierfür in verschiedene Arbeitsgruppen auf.

Dabei ging es nicht nur um die Dateiformate der Objekte, sondern auch um die Formate der Metadaten und die verschiedenen Workflows, wie diese in Rosetta eingespielt werden. Neben der Festlegung, ob die Datenpakete manuell oder automatisch eingespielt oder vom Archivsystem regelmäßig geholt werden sollen, wird auch die Struktur der Pakete inklusive der Metadaten festgelegt. Das bezieht sich darauf, wie die Metadaten abgeliefert werden, ob als METS, Dublin Core (DC), CSV oder BagIT⁸. Diese Wahl beeinflusst dann die Ordnerstruktur, die das Paket erfüllen muss.

Es wurde mit den Dateiformaten TIFF und PDF begonnen, da diesen eine herausgehobene Bedeutung für die bereits kurzfristig zu archivierenden Materialien der beteiligten Universitäten zugemessen wurde. In den Ablieferungsmodulen für TIFF und PDF wurden die Grundlagen für die Einrichtung von Workflows gelegt, die in den übrigen Ablieferungsmodulen nachgenutzt bzw. konzeptionell übertragen werden konnten. Im TIFF-Format liegen hauptsächlich Retro-Digitalisate vor. Für diese wurde ein Workflow ausgewählt und eingerichtet, der die Metadaten in einer METS-Struktur liefert.

⁸ METS steht für Metadata Encoding and Transmission Standard und ist ein XML-basierter Metadatenstandard zur Beschreibung digitaler Objekte. Dublin Core ist ein Metadatenstandard für deskriptive Metadaten. CSV steht für Comma-separated values und ist ein Standard zur Beschreibung einfacher Textdateien. BagIT beschreibt eine Verzeichnisstruktur, die zur Übertragung von Dateien genutzt werden kann.

Rosetta nutzt ein eigenes METS-Profil, welches das eigens entwickelte DNX-Schema verwendet. Es ist orientiert am METS-Standard⁹, weicht aber an einigen Punkten davon ab, so dass bereits in METS vorliegende Metadaten entsprechend angepasst werden müssen. Das bezieht sich z.B. auf den Header. Diese Ingest-Methode bietet u.a. die Möglichkeit, bereits vorhandene Metadaten zu den Objekten zu übernehmen. Diese werden dann als Provenance-Metadaten aufgenommen. Außerdem ermöglicht sie die Abbildung von Strukturen, so dass innere Zusammenhänge und Hierarchien einzelner Archivobjekte abgebildet werden können, womit eine zukünftige Nutzung erleichtert, gegebenenfalls auch überhaupt erst gewährleistet werden kann.

Die Projektpartner haben nach den Vorgaben, die sich aus den Anforderungen der Rosetta-Konfiguration ergeben, Datenpakete erstellt, die neben den TIFF-Dateien eine Rosetta-konforme METS-Datei und eine einfache XML-Datei mit einem DC-konformen Titeleintrag enthalten mussten. Zusätzliche Dateien, wie beispielsweise OCR-Dateien oder Metadaten in unterschiedlichen Formaten können optional mitgeliefert werden.

Die Datenkurator*innen haben aus ihren Beständen unterschiedliche Testpakete zusammengestellt, die prototypisch für eine spätere Ablieferung im Regelbetrieb stehen. Dabei wurden die vorhandenen Metadaten der Ausgangsrepositorien in das Rosetta-konforme METS-XML konvertiert. Dafür wurden entsprechende Stylesheets entwickelt, die den jeweils anderen Institutionen zur Nachnutzung zur Verfügung gestellt wurden und werden.

Zum Testen wurden sowohl einzelne Datenpakete als auch – in einem halb-automatischen Verfahren – mehrere hundert Pakete eingespielt. Beide Verfahren verliefen nach einigen Anpassungen erfolgreich.

Für Textdokumente im PDF-Format wurde eine DC-Struktur ausgewählt und getestet. Die Auswahl für eine Struktur hängt nicht vom Dateiformat ab, sondern wird aufgrund der vorhandenen Metadaten und in Absprache mit den abliefernden Universitäten getroffen und kann für unterschiedliche Bestände und Sammlungen differieren. Für weitere Dateiformate (epub, jpeg, Audio- und Videoformate) wurden ebenfalls die genannten Ablieferungsworkflows eingerichtet und getestet.

Innerhalb von Rosetta erfolgt eine Validierung des jeweiligen Dateiformats. Dabei wird geprüft, ob das Dateiformat konform zur jeweiligen Spezifikation ist. Ist dies nicht der Fall, kann es bei späteren Dateimigrationen zu Problemen kommen und im schlimmsten Fall die Nachnutzbarkeit der Datei gefährden. Daher ist eine genaue Überprüfung der Dateien zu Beginn des Archivierungsprozesses sehr wichtig.

Im Falle von Fehlermeldungen wird – wenn möglich – geprüft, ob und wenn ja wie gravierend der Fehler im Hinblick auf die langfristige Nutzung sein könnte. Ist dies der Fall wird die abliefernde Institution um Prüfung und Behebung des Problems gebeten, bevor das Objekt erneut ins System überführt wird.

9 <<https://www.loc.gov/standards/mets/>>, Stand: 19.07.2022.

Eine vorgelagerte Validierung hat den Vorteil, dass Dateien, die solche Fehler enthalten, die Auswirkungen auf eine spätere Nutzung enthalten, nicht im System aufgenommen werden. Vielmehr werden die Probleme bereits im Vorfeld erkannt, die Daten überprüft und neu erstellt und erst dann im Archivsystem abgelegt.

In LaVaH I wurden prototypische Workflows entwickelt und getestet. Textdokumente und Digitalisate mit standardisierten Metadaten können in mehreren Verfahren eingespielt werden. Je nachdem, wie umfangreich die vorhandenen Metadaten sind, wird eine passende Struktur ausgewählt. Dabei sind weniger die Dateiformate die ausschlaggebenden Faktoren als vielmehr die Metadaten. Bei den Dateiformaten steht die Validierung im Fokus und die Frage, wie mit Fehlermeldungen umgegangen werden soll.

Die Ergebnisse der Arbeitsgruppen und der Tests flossen in die Spezifikationen für die verschiedenen Ablieferungsformate ein. Diese sind Teile der Ablieferungsvereinbarung, in welcher die Rechte und Pflichten der beteiligten Einrichtungen geregelt sind.

4. Schwerpunkte der zweiten Projektphase: LaVaH II

In der nun gestarteten zweiten Projektphase wird auf die Ergebnisse aus LaVaH I erweiternd aufgebaut. Zum einen bezieht sich diese Erweiterung auf den Kreis der teilnehmenden Einrichtungen. Neben den bisher schon beteiligten hessischen Universitäten sind nun auch die Hochschulen für angewandte Wissenschaften in LaVaH eingebunden. Auch dort entstehen Daten, die langfristig gesichert und verfügbar gehalten werden müssen. So besteht z.B. in Hessen ein (befristetes) Promotionsrecht für forschungsstarke Fachrichtungen an den HAWs.¹⁰

Zum anderen wird auch das Spektrum der zu archivierenden digitalen Objekte erweitert. Forschungsdaten und digitale Sammlungen der beteiligten Einrichtungen, die erhaltenswert sind und nicht in anderweitigen fachlichen Langzeitarchiven untergebracht werden, sollen in LaVaH eingebracht werden. Bei Forschungsdaten und digitalen Sammlungen besteht die Herausforderung nicht nur darin, dass weitere Dateiformate zu archivieren, sondern vor allem auch, Bewertungsmodelle für die Auswahl der zu archivierenden Daten zu entwickeln sind.

Die Langzeitarchivierung von Forschungsdaten ist aufwändig und ressourcenintensiv, u.a., weil es disziplinspezifisches Wissen voraussetzt. Was bislang fehlt, sind Mechanismen zur Bewertung und zur Übergabe von Daten, z.B. aus institutionellen Repositorien an digitale Langzeitarchive. Diese sollen in LaVaH II entwickelt werden.

Die bereits in LaVaH I entstandenen Ablieferungsmodule werden in LaVaH II in den Produktivbetrieb überführt. Die bereits erprobten Workflows werden für automatisierte Massenablieferungen so ausgerichtet, dass sukzessive für alle Partner eine Ablieferung erfolgen kann. Automatisiert werden

¹⁰ Siehe auch: <<https://wissenschaft.hessen.de/studieren/hessens-hochschulstrategie/promotionsrecht-haw>>, Stand: 19.07.2022.

kann zum Beispiel die Erstellung eines Datenpakets, welches den Anforderungen der jeweiligen Ingest-Variante entspricht. Das kann bedeuten, dass bereits vorhandene Metadaten geharvestet und in das geforderte Metadatenformat umgewandelt werden. In einem weiteren Schritt wird dann ein Paket mit Daten und Metadaten (und eventuell weiteren, wie OCR-Daten) erstellt, das zum Ingest in das digitale Langzeitarchiv zur Verfügung steht. Jedoch geht es nicht nur um die automatische Erstellung von Datenpaketen, sondern auch um Dateivalidierung und um die Überführung der Datenpakete in das Langzeitarchiv. Dabei kann auf bestehende Skripte von ExLibris und anderen Rosetta-Nutzenden zurückgegriffen werden, die entsprechend angepasst werden müssen.

Da ein Großteil der zu archivierenden Daten bereits in verschiedenen Systemen (Repositorien, Portalen etc.) verwaltet und vorgehalten wird, soll die Überführung in das Langzeitarchiv in diesen Fällen automatisiert erfolgen. Hier können Synergien genutzt werden, weil mehrere Einrichtungen die gleiche Software z.B. für ihre Repositorien nutzen. Es gab bereits erste Tests, wie eine automatisierte Ablieferung aus bestehenden Systemen, z.B. aus Publikationsrepositorien (auf der Software DSpace basierend) oder Digitalisierungsplattformen wie der Visual Library organisiert werden kann. Per Skript wurden dazu Daten und Metadaten geharvestet und in Rosetta konforme Pakete gepackt, welche dann in das Archiv überführt wurden.

Die Abfolge der Ablieferung wird zentral gesteuert. Für die neuen Partner in LaVaH II wird eruiert, ob bereits vorhandene Workflows nachgenutzt werden können.

Während in der ersten Projektphase der Schwerpunkt darauf lag, Daten in das Testsystem einzuspielen, soll in LaVaH II auch die Bereitstellung von Daten aus dem Archivsystem realisiert werden. Dazu soll zunächst eine Bedarfserhebung bei den Partnern durchgeführt werden und eine Arbeitsgruppe wird Verfahren für die Bereitstellung und Auslieferung von Archivalien, inkl. Bereitstellungsformaten und -fristen entwerfen und testen. Besonderes Augenmerk wird dabei auf die möglichen Nutzungsrechte gelegt werden.

4.1 Spartenübergreifende Infrastruktur

Eine zentrale Frage bei der Langzeitarchivierung ist die nach der Auswahl der zu archivierenden digitalen Objekte. Auch im Hinblick auf die anfallenden Kosten (Personal, Infrastruktur) lohnt es sich, Kriterien für die Auswahl zu erstellen. Aus diesem Grund entsteht in LaVaH II eine übergreifende Arbeitsgruppe „Bewertungsmodelle“, an der auch Kolleg*innen mit archivarischem Hintergrund aus dem Digitalen Archiv der Hochschulen in Hessen¹¹ (DAH) beteiligt sind. Diese soll sich damit beschäftigen, wie die lokalen Datenkurator*innen eine Auswahl treffen und begründen können. Als Unter-Arbeitsgruppe wird die AG „Signifikante Eigenschaften“, die schon im Laufe von LaVaH I ihre Arbeit aufgenommen hat, weitergeführt. Die Bestimmung dieser Eigenschaften soll möglichst einfach erfolgen. Definierte Eigenschaften sollen in Form von Metadaten dem Archivpaket beigelegt werden. Im Falle einer notwendigen Formatmigration kann mit ihrer Hilfe ein Zielformat ausgewählt werden, welches die Anforderungen zum Erhalt der festgelegten signifikanten Eigenschaften erfüllt.

11 <<https://www.dahh.de/>>, Stand: 19.07.2022.

Darüber hinaus sollen Schulungsmodule und Online-Schulungen entwickelt und zur Nutzung innerhalb der Partnereinrichtungen bereitgestellt werden. Grundlagen der Langzeitarchivierung und die Bedeutung einer guten Vorbereitung der Daten für eine spätere Archivierung sollen besonders hervorgehoben werden. Es soll ein Bewusstsein geschaffen werden, möglichst früh darauf zu achten, dass z.B. langzeitarchivierungsfreundliche Formate gewählt werden.

Zusammen mit dem Digitalen Archiv der Hochschulen in Hessen (DAHH) wird an Kriterien zur Bewertung und Auswahl von Daten gearbeitet und langfristig soll eine Übersicht über die Bestände an hessischen Hochschulen entstehen.

4.2 Kooperatives Bestandsmanagement

Über die Objekte aus den hessischen Universitäten und Hochschulen hinaus gibt es Überlegungen, wie besondere Objekte am besten geschützt werden können. Gemeint sind Objekte, die besonders wertvoll, einmalig, schutzbedürftig etc. sind. Das können z.B. besondere Sammlungen oder Nachlässe sein. Geplant ist, eine an Bedeutung und Gefährdung der betreffenden Objekte ausgerichtete und überörtliche Redundanz in der Archivierung zu schaffen. Dazu bedarf es einer Anforderungsanalyse, um dann über geeignete Verteilungs- und Bereitstellungsverfahren zu diskutieren.

Für die elektronischen Pflichtexemplare (ePflicht) wird eine gemeinsame Infrastruktur zur Sammlung und Langzeitarchivierung aufgebaut.

Es wird jedoch auch über den Tellerrand der Landesgrenzen hinausgeschaut und der Austausch und die Kooperation mit anderen Initiativen hergestellt, z.B. durch die aktive Beteiligung an *nestor*¹² und an der deutschsprachigen und internationalen *Rosetta-Community*. Darüber hinaus besteht eine Mitgliedschaft beim *Digital Curation Center*¹³ (DCC).

Derzeit wird eine Kooperation mit LZA-Initiativen in anderen Bundesländern¹⁴ aufgebaut, in der eine Zusammenarbeit bei konzeptionellen und technischen Aspekten der Langzeitverfügbarkeit angestrebt wird.

5. Zusammenfassung

Der in LaVaH praktizierte Ansatz einer dezentralen Aufgabenverteilung hat sich bewährt. Die Datenkurator*innen an den Universitäten und Hochschulen sind mit ihren Beständen vertraut und haben Zugang zu den Portalen oder Repositorien, in denen diese registriert und z.T. präsentiert werden. Gemeinsam mit der *hebis VZ* werden passende Transformationen erarbeitet und genutzt, um *Rosetta*-konforme Transferpakete zu schnüren. Auch wenn die Institutionen mit unterschiedlichen

12 <https://www.langzeitarchivierung.de/Webs/nestor/DE/Home/home_node.html>, Stand: 19.07.2022.

13 <<https://www.dcc.ac.uk/>>, Stand: 19.07.2022.

14 Neben Hessen mit LaVaH sind derzeit das Projekt „Digitale Langzeitverfügbarkeit für Wissenschaft und Kultur in Bayern“ und die Landesinitiative Langzeitverfügbarkeit LZV.NRW beteiligt.

Objekttypen und Systemen arbeiten und ihre Daten selbst verwalten, werden die erarbeiteten Lösungen für andere Einrichtungen angepasst und nachgenutzt. Die gemeinsame Nutzung der Archivierungssoftware hat neben der ressourcensparenden Komponente auch den Effekt, dass ein gemeinsames Verständnis von zentralen Aspekten der Langzeitarchivierung (Metadaten, signifikante Eigenschaften, Formatvalidierung etc.) entstanden ist. In der aktuell laufenden zweiten Projektphase werden die im ersten Projektabschnitt prototypisch erarbeiteten Workflows erweitert und in den Produktivbetrieb überführt.

LaVaH zeigt eine Langzeitarchivierungslösung auf Landesebene auf, die alle Stakeholder mit einbezieht, eine gemeinsame technische Infrastruktur nutzt und auf abgestimmte Standards setzt.

Literatur

- Schwens, Ute; Liegmann, Hans: Langzeitarchivierung digitaler Ressourcen, in: Kuhlen, Rainer; Seeger, Thomas; Strauch, Dietmar (Hg.): Grundlagen der praktischen Information und Dokumentation, 5., völlig neugefasste Ausg. München 2004 (Handbuch zur Einführung in die Informationswissenschaft und -praxis 1), S. 567–570. Online: <<http://nbn-resolving.de/urn:nbn:de:0008-2005110800>> (Die Zählung der Abschnitte stimmt nicht mit der Druckausgabe überein).