

Archive als Informationsdienstleister und Infrastruktureinrichtungen

Positionspapier der Konferenz der Leiterinnen und Leiter der Archivverwaltungen des Bundes und der Länder (KLA) vom Dezember 2018 zum Aufbau einer Forschungsdateninfrastruktur für die historisch arbeitenden Geisteswissenschaften im Rahmen der Nationalen Forschungsdateninfrastruktur (NFDI)

Stand: 07.12.2018

Präambel

Archive stellen als Gedächtnisinstitutionen einen zentralen Baustein in der historisch arbeitenden, insbesondere in der geschichtswissenschaftlichen Forschungslandschaft dar. Sie verfügen über ein breites inhaltliches Angebot an interdisziplinär relevanten Forschungsdaten in Form von analogem, digitalisiertem und originär digitalem Archivgut einschließlich zugehöriger Erschließungsinformationen. Diesen Content machen sie der Forschung über verschiedene Kanäle wie Fachinformationssysteme oder übergreifende Portale frei zugänglich. Als Partner insbesondere der Geschichtswissenschaft beraten und begleiten die Archive seit Jahrhunderten Forschende bei der Recherche nach und der Auswertung von Forschungsdaten. Zugleich verfügen Archive über ein breites Wissen und konkrete Erfahrungen in Bezug auf die dauerhafte Sicherung dieser Daten bei gleichzeitiger Wahrung der Authentizität der Quellen – denn sie kann nur durch Archive gewährleistet werden. Archivische Inhalte, Services und Kompetenzen können daher nutzbringend bei der Etablierung einer Nationalen Forschungsdateninfrastruktur (NFDI) eingebracht werden. Im Sinne der vom Rat für Informationsinfrastrukturen (RfII) eingeforderten institutionellen Vernetzung und Zusammenarbeit streben die staatlichen Archive gemeinsam mit anderen historisch orientierten Kultur- und Forschungs(infrastruktur)einrichtungen und dem Verband der Historiker und Historikerinnen Deutschlands e.V. (VHD) an, ein Konsortium der historisch arbeitenden Geisteswissenschaftler zu bilden.¹ Die Archive sehen darin die Chance, gemeinsam mit anderen Institutionen und den Nutzerinnen und Nutzern Angebote und Services forschungs- und nutzungsorientiert auszubauen und weiter zu verbessern.

Ausgangslage

Archive sind als Forschungsinfrastruktureinrichtungen bekannt und angesehen. Neben den klassischen Aufgabenfeldern der Archive – Überlieferungsbildung, Erschließung, (analoge) Nutzung, Bestandserhaltung – gehören auch die Digitalisierung und Online-Bereitstellung von Kulturgut zu ihrem regulären Tätigkeitsspektrum. Die staatlichen Archivverwaltungen

¹ Vgl. die im Positionspapier des VHD zur Schaffung einer NFDI formulierte Zielstellung zur "Etablierung eines [...] fachspezifisch organisierten, kooperierenden Netzes institutioneller und lokal, regional oder national operierender Datenzentren und einer darauf bezogenen Koordinierungsstelle für die Geschichtswissenschaften, <https://www.historikerverband.de/verband/stellungnahmen/positionspapier-zur-schaffung-nationaler-forschungsdateninfrastrukturen-nfdi.html> (alle Links aufgerufen am 7.12.2018).

verfügen über langjährige Erfahrungen und haben einheitliche Standards für das Scannen, für Medienformate und für Metadaten entwickelt und etabliert.² Erschließungsinformationen und digitalisiertes Archivgut werden der Forschung bereits vielfach kostenfrei in den institutionseigenen Online-Findmittelsystemen zugänglich gemacht sowie in zahlreichen übergreifenden Portalen zur Verfügung gestellt. Angebote zur *Digitization on Demand* eröffnen weitere, zielgerichtete Zugangsperspektiven. Die Online-Bereitstellung von frei zugänglichem Archivgut wird durch derzeit entwickelte Konzepte für einen geschützten virtuellen Lesesaal erweitert. Hier sollen künftig auch zugangsbeschränkte Unterlagen zur Nutzung bereitgestellt werden.³

Die digitale Präsentation von Archivgut im Internet wird als strategische Fachaufgabe mit hoher Priorität verfolgt, sichtbar in den landesweiten bzw. regionalen Archivportalen und insbesondere in dem umfangreichen Datenangebot innerhalb des Archivportals-D.⁴ Dieses Informationssystem wurde in einem Gemeinschaftsprojekt als Spartenportal der Deutschen Digitalen Bibliothek (DDB) entwickelt und hat sich als zentraler Anlaufpunkt für die vernetzte Recherche nach digitalen Quellen und Erschließungsinformationen aus deutschen Archiven etabliert.⁵ Die erfolgreiche Umsetzung, das nachhaltige Betriebskonzept und die konkreten Weiterentwicklungsperspektiven dieser Plattform zeigen, dass die deutschen Archive als forschungs- und serviceorientierte Informationsanbieter zusammen mit ihren IT-Partnern leistungsfähig sind und einen wertvollen Beitrag beim Aufbau einer NFDI leisten können.⁶

² Vgl. die Ergebnisse des DFG-Produktivpiloten zur Digitalisierung von archivalischen Quellen (<https://www.archivschule.de/DE/forschung/forschungsprojekte/digitalisierung/dfg-projekt-archivgutdigitalisierung.html>). Zu nennen ist in diesem Zusammenhang ferner u.a. die Entwicklung und Kommunikation von Datenaustauschformaten und -standards (EAD(DDB), archivisches METS/MODS-Profil) wie auch die Weiterentwicklung des DFG-Viewers.

³ Siehe hierzu Aufbau eines virtuellen Lesesaals. Sachstandsbericht des KLA-Ausschusses „Archivische Fachinformationssysteme“ (Natascha Noll), in: *Archivar* Jg. 71, 2018, Heft 3, S. 275–283, auch online: http://www.archive.nrw.de/archivar/hefte/2018/Ausgabe-3/Archivar-3_2018.pdf. Außerdem Gerald Maier/Clemens Rehm/Julia Kathke: Nutzung digital. Konzepte, Angebote und Perspektiven eines „virtuellen Lesesaals“ im Landesarchiv Baden-Württemberg, in: *Archivar* Jg. 69, 2016, Heft 3, S. 237–248, auch online: www.archive.nrw.de/archivar/hefte/2016/Ausgabe_3/Archivar_3_2016.pdf. Peter Sandner: „Virtueller Lesesaal“. Originär archivspezifische Anforderungen an einen virtuellen Nutzungsbereich, in: *Neue Wege ins Archiv – Nutzer, Nutzung, Nutzen*. 84. Deutscher Archivtag in Magdeburg, hg. vom VdA – Verband deutscher Archivarinnen und Archivare e. V. Fulda 2016 (= Tagungsdokumentation zum Deutschen Archivtag Bd. 19), S. 37–45; Beate Dorfey/Marc Straßenburg: Open Access – Ein Konzept zur Gestaltung von Zugang zu Archivgut. Vortrag auf der 22. Tagung des Arbeitskreises zur Archivierung von Unterlagen aus digitalen Systemen am 7./8.03.2018 in Marburg, online abrufbar unter https://www.staatsarchiv.sg.ch/home/auds/22/_jcr_content/Par/downloadlist_692544306/DownloadListPar/download_87448353.ocFile/Dorfey_Strassenburg_OpenAccess.pdf.

⁴ Das Angebot des Archivportals-D umfasst aktuell bereits ca. 14 Mio. Erschließungsdatensätze von über 140 Archiveinrichtungen (<https://www.archivportal-d.de/>). Innerhalb der Deutschen Digitalen Bibliothek stellt die Sparte Archiv den größten Anteil an Daten.

⁵ Siehe dazu Gerald Maier/Christina Wolf: Das Archivportal-D und die Deutsche Digitale Bibliothek. Neue übergreifende Recherchemöglichkeiten nach Quellen für die historische Forschung, in: *Recherche und Weiterverarbeitung. Digitale Angebote der Archive für die historische Forschung im Netz*. Beiträge einer Sektion auf dem 51. Deutschen Historikertag 2016 in Hamburg, hg. von Rainer Hering und Robert Kretzschmar, Stuttgart 2017, S. 10–35 und die weiteren Beiträge in dem genannten Band.

⁶ So ist z.B. für das Archivportal-D das Leibniz-Institut für Informationsinfrastruktur – FIZ Karlsruhe (<https://www.fiz-karlsruhe.de/>) der maßgebliche IT-Partner als Software-Entwickler und technischer Betreiber.

In zahlreichen Kooperationsprojekten agieren staatliche Archiveinrichtungen bereits als aktive Partner der historischen Forschung, z.B. im Rahmen von Digitalisierungs- und Forschungsprojekten.⁷ Diese Projekte, in die Vertreterinnen und Vertreter der *Digital Humanities* sowie der klassischen Geschichtswissenschaften einbezogen sind, illustrieren das Potenzial von Archiven bei der Etablierung neuer Zugänge zu Archivgut.

Die dauerhafte Speicherung, Lesbarhaltung und Bereitstellung digitaler historischer Forschungsdaten muss ein zentraler Anforderungskomplex innerhalb der NFDI sein. Die staatlichen Archive verfügen über bewährte Technologien und einschlägiges Know-how, die in ein Konsortium historisch arbeitender Geisteswissenschaften eingebracht und dort weiterentwickelt werden können. Die in den staatlichen Archiven eingesetzten Systeme folgen dem international anerkannten und maßgeblichen Referenzmodell für ein Offenes Archiv-Informationssystem (OAIS).

Handlungsfelder

Aus Sicht der staatlichen Archive sind folgende Handlungsfelder von besonderer Bedeutung für das zu bildende Konsortium:

- 1) **Übernahme und Langzeitarchivierung von Daten sowie Bereitstellung für die Nutzung:** Ziel einer dauerhaften Sicherung von Forschungsdaten muss es sein, diese auch über den Lebenszyklus der bei der Produktion der Forschungsdaten eingesetzten Technologien hinaus langfristig erhalten, auswerten und nutzen zu können – und dies unabhängig vom Primärzweck ihrer Entstehung. Für die Nachvollziehbarkeit, Überprüfung und Nachnutzung von Forschungsergebnissen ist diese Anforderung unabdingbar. Nutzerinnen und Nutzer erwarten einerseits, dass die ursprünglichen Daten in ihrer originären technischen Dienstumgebung nutzungsfähig bleiben. Andererseits sollen die Daten reibungslos vom Ausgangsdienst in die eigenen Auswertungstools von Nutzerinnen und Nutzern übergehen können. Verlust von Daten schädigt die Forschung und muss folglich mit der einschlägigen Strategie der Migration, ergänzt durch Emulation, verhindert werden. Die von den staatlichen Archiven bereits realisierten Technologien und Verfahren bieten erfolgversprechende Lösungsansätze zur Bewältigung dieser Herausforderung und werden kontinuierlich weiterentwickelt.
- 2) **Standardisierung, Normdaten und Interoperabilität:** Die Bedeutung von Maßnahmen zur Homogenisierung und Angleichung von Informationen und Datenstrukturen/-formaten steigt mit dem Bedarf der übergreifenden Auswertbarkeit. Gleichförmigkeit setzt jedoch die Verständigung auf gemeinsame Regelwerke voraus, die sich z.B. in Formatstandards manifestieren. Die deutschen Archive haben sich auf solche Standards verständigt, um die großen Portale nachhaltig betreiben und der Forschung und anderen Nutzergruppen ein verlässliches Informationsangebot zur Verfügung stellen zu können.

⁷ Als Beispiele sind aktuell das DFG-geförderte Forschungsprojekt „Überwachung. Macht. Ordnung – Personen- und Vorgangskarteien als Herrschaftsinstrument der Gestapo“ des Niedersächsischen Landesarchivs und der Universität Osnabrück (<https://hvos.hypotheses.org/583>), das DFG-geförderte Vorhaben zum „Aufbau sachthematischer Zugänge im Archivportal-D am Beispiel Weimarer Republik“ (<https://www.landearchiv-bw.de/web/63525>) und das EU-Projekt COOP zur Verbesserung des Zugangs zu Archivgut (<https://coop-project.eu/>) zu nennen.

Zukünftig wird das noch wichtiger werden, etwa wenn Spartengrenzen in übergreifenden Projekten überschritten werden. Eine zentrale Rolle spielen hier Normdaten und die Verwendung kontrollierter Vokabulare. Auf der Basis normierter Daten lässt sich das Linked-Data-Konzept realisieren, das eine Vernetzung und einen Transfer der Informationen über Repositoriengrenzen hinaus eröffnet.

- 3) **Open-Access, Nachnutzbarkeit und Schnittstellen:** Aus dem freien Zugang zu Informationen und Kulturgut (Open Access) und den Möglichkeiten datengetriebener Forschung entstehen neue Nutzungskontexte, aus denen sich wieder neue Anforderungen ergeben. Das betrifft etwa den Bedarf zur Nachnutzbarkeit von Forschungsdaten unabhängig vom primären Zweck ihrer Entstehung, für die einerseits der Einsatz einheitlicher Lizenzierungsmodelle (z.B. Creative Commons) und andererseits die Bereitstellung der Daten über offene Schnittstellen (APIs) Voraussetzung ist. Sie müssen so konzipiert sein, dass relevante Datenmengen für die jeweiligen Fragestellungen eines Forschungsanliegens zur Verfügung gestellt werden können. Dies kann sich sowohl auf größere Datenvolumen beziehen als auch auf die unkomplizierte Einbindung einzelner Digitalisate unterschiedlicher Provenienz in einen IIIF-kompatiblen Viewer. Die Bereitstellung archivalischer Daten über Schnittstellen ermöglicht deren Integration, Auswertung und Anreicherung in andere(n) Forschungsumgebungen und -werkzeuge(n), die z.B. auf Transkriptionen, Annotationen oder Georeferenzierung der digital vorliegenden historischen Quellen abzielen.
- 4) **Datenanalyse, -auswertung und -anreicherung:** Übergreifende Datenrepositorien und neue Methoden der Datenauswertung auf der Basis von Konzepten wie Machine Learning oder Deep Learning ermöglichen perspektivisch die automatisierte Generierung zusätzlicher Erschließungsinformationen. Damit kann trotz begrenzter personeller Ressourcen in den Kultureinrichtungen eine erheblich höhere Datenqualität erreicht werden. Über entsprechende Verfahren des Text-/Data-Minings lassen sich Entitäten oder sachthematische Referenzen in vorhandenen Textkorpora identifizieren und extrahieren, um zusätzliche Recherchemöglichkeiten zu unterstützen und die semantische Vernetzung der Inhalte voranzutreiben. Auch Bildmustererkennung kann eingesetzt werden, um digitalisierte Archivalien automatisiert zu klassifizieren (Typ der Vorlage, Identifikation dargestellter Objekte etc.). Für historisch Forschende soll damit eine signifikant verbesserte Datengrundlage bereitgestellt werden, die zugleich durch erweiterte und neue Zugangsmöglichkeiten gekennzeichnet ist.
- 5) **Recherchierbarkeit von Volltexten:** Die Bereitstellung von digitalisiertem Archivgut wird längst erweitert um die nutzerseitige Anforderung, auch die Volltexte zu erfassen und durchsuchbar zu gestalten. Technologisch ist das durch Fortschritte bei OCR-Verfahren für jüngeres Archivgut bereits möglich. Auch für ältere, handschriftliche Dokumente wird Volltextrecherchierbarkeit durch den Einsatz von Mustererkennung etwa im Transkribus-Projekt⁸ zunehmend realistisch. Hierdurch wird der Datenumfang maßgeblich ansteigen und somit eine deutlich bessere Basis für mögliche Forschungsfragestellungen und beispielsweise Editionsprojekte (Tiefenerschließung bzw. Binnenstrukturierung der Volltexte via TEI-XML) geschaffen.

⁸ Siehe <https://transkribus.eu/>.

Perspektive

Ein für die historisch arbeitenden Geisteswissenschaften einzurichtendes Konsortium im Rahmen einer Nationalen Forschungsdateninfrastruktur kann insbesondere zu einem verbesserten Austausch zwischen den historisch orientierten Kulturinstitutionen und der geschichtswissenschaftlichen Fachcommunity führen. Für Archive und weitere Kultureinrichtungen bedeutet dies zunächst eine verbesserte Qualität und breitere Nutzung ihres bestehenden Angebots, für die Wissenschaft eine verstärkte Wahrnehmung ihrer konkreten Bedürfnisse durch die Kulturinstitutionen. In der Folge können Anforderungen und Angebote künftig noch besser aufeinander abgestimmt und Forschungsdaten breitgefächert produziert, bereitgestellt, ausgewertet, genutzt sowie schließlich nachhaltig archiviert und für neue Forschungskontexte nachgenutzt werden. Der Aufbau eines fachspezifischen Konsortiums im Kontext der Errichtung einer Nationalen Forschungsdateninfrastruktur ist daher seitens der Archive in höchstem Maße zu begrüßen. Die staatlichen Archivverwaltungen sind sehr daran interessiert, ihre Kompetenzen und Erfahrungen in diesen Prozess einzubringen und sich an der Schaffung eines entsprechenden Konsortiums aktiv zu beteiligen.