

BEREITSTELLUNG, AUFBEREITUNG, LANGZEITSICHERUNG

FUNKTIONEN DER ARCHIVE IN DER FORSCHUNGSDATEN- INFRASTRUKTUR

von *Gerald Maier, Daniel Föhle und Andreas Neuburger*

Als fest etablierte Gedächtnisinstitutionen und Forschungsinfrastruktureinrichtungen sind Archive unverzichtbare Akteure in der historisch arbeitenden und insbesondere der geschichtswissenschaftlichen Forschungslandschaft. Sie sind damit ein integrales Element der übergreifenden Forschungsinfrastruktur der Bundesrepublik. Hieraus ergibt sich die Herausforderung, auch den Aufbau der Nationalen Forschungsdateninfrastruktur (NFDI) mitzugestalten und einschlägige Expertise und Kompetenzen der Archive in den laufenden Prozess einzubringen.¹ Zudem erfordert die Etablierung einer übergreifenden, also interdisziplinären und gleichzeitig vernetzten Forschungsdateninfrastruktur, dass Archive ihre Angebote und Services weiterentwickeln und neuen Anforderungen gerecht werden müssen. Ziel des Beitrags ist es, zunächst die Handlungsfelder und Angebote zu umreißen, die das Archivwesen in eine NFDI einbringen kann. Ferner sollen Überlegungen zum Handlungsrahmen sowie dem konkreten Förderungsbedarf angestellt werden, über welche die Beiträge der Archive nachhaltig in die NFDI integriert und dort mit möglichst breiten Wirkungsperspektiven verankert werden können.

AUSGANGSLAGE: ARCHIVE ALS ANBIETER VON FORSCHUNGSDATEN

In den Überlieferungen der Archive findet sich ein umfangreiches Angebot an interdisziplinär relevanten Forschungsdaten. Archive beraten und unterstützen Forschende bei der Recherche und Auswertung dieser Forschungsdaten. In zahlreichen

Kooperationsprojekten agieren sie bereits als aktive Partner insbesondere der historischen Forschung, z. B. im Rahmen von Digitalisierungs- und Forschungsprojekten.² Mit Blick auf die konkrete Rolle des Archivwesens in der NFDI ist dabei zu betonen, dass die Archive bereits seit Jahren über umfangreiche, rasant wachsende und darüber hinaus zumeist digital frei zugängliche Forschungsdatenbestände verfügen.³ Neben den klassischen Aufgabenfeldern der Archive – Überlieferungsbildung, Erschließung, Nutzung, Bestandserhaltung – gehören auch die Digitalisierung und Online-Bereitstellung von Kulturgut zu ihrem regulären Tätigkeitsspektrum. Archive verfügen über langjährige Erfahrungen und haben einheitliche Standards für

¹ Mit dem Aufbau der NFDI verfolgen Bund und Länder das Ziel, „Datenbestände von Wissenschaft und Forschung systematisch [zu] erschließen, nachhaltig [zu] sichern und zugänglich [zu] machen sowie (inter-)national [zu] vernetzen“ (<https://www.dfg.de/foerderung/programme/nfdi/>, aufgerufen am 25.11.2019).

² 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 (aufgerufen am 25.11.2019).

³ Der Forschungsdatenbegriff des Rats für Informationsinfrastrukturen wäre dementsprechend teilweise zu erweitern, vgl. Rat für Informationsinfrastrukturen: Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland, Göttingen 2016, S. A-13; online unter: <http://www.rfii.de/?p=1998>, (aufgerufen am 25.11.2019).

das Scannen, für Medienformate und für Metadaten entwickelt und etabliert.⁴ Erschließungsinformationen und digitalisiertes Archivgut werden der Forschung vielfach kostenfrei in den institutionseigenen Online-Findmittelsystemen zugänglich gemacht sowie in übergreifenden Portalen zur Verfügung gestellt. Angebote zur Digitization on Demand eröffnen weitere, zielgerichtete Zugangsperspektiven ebenso wie Überlegungen zu Nutzungskonzepten für zugangsbeschränkte Unterlagen.

Im Unterschied zu den natur- und ingenieurwissenschaftlichen sowie einem Teil der sozialwissenschaftlichen Forschung basieren die historisch arbeitenden Geisteswissenschaften auf der Auswertung von Datenbeständen, die in ihrer analogen Ausprägung auf Pergament oder Papier oftmals seit Jahrhunderten vorhanden sind.⁵ Zu einem immer höheren Anteil werden diese Daten digital verfügbar, zunächst über die datenbankgestützte Erstellung beschreibender Metadaten im Rahmen der archivischen Erschließung. Diese bildet wiederum die Grundlage für die genannten umfangreichen Anstrengungen zur Digitalisierung von Archivgut und der informationswissenschaftlichen Aufbereitung sowohl der Metadaten als auch der Digitalisate.⁶ Neuerdings kommen originär digitale („born digital“) Objekte hinzu, die es zu übernehmen, dauerhaft zu sichern und ebenfalls in geeigneter Weise zugänglich zu machen und zu halten gilt.

Digitale Findmittel, Digitalisate von Archivgut sowie originär digitale Unterlagen sind daher als Forschungsdaten oder auch „Forschungsrohdaten“ bzw. „primäre Forschungsdaten“ zu betrachten und dementsprechend im Rahmen einer NFDI zu berücksichtigen. Nur auf diese Weise ist gewährleistet, dass die historisch arbeitenden Geisteswissenschaften und zumal die innovativen Ansätze der Digital Humanities eine inhaltlich und technisch zeitgemäße Materialbasis für unterschiedlichste Forschungsansätze erhalten und diese Disziplinen nicht von der digitalen Weiterentwicklung der deutschen Forschungslandschaft ausgeschlossen werden.

HANDLUNGSFELDER

Bereits im Dezember 2018 hat sich die Konferenz der Leiterinnen und Leiter der Archivverwaltungen des Bundes und der Länder (KLA) als Organ der staatlichen Archivverwaltungen mit einem Positionspapier in den Aufbauprozess der NFDI eingeschaltet.⁷ Als Handlungsfelder werden darin konkret die folgenden Bereiche benannt:

- Übernahme und Langzeitarchivierung von Daten sowie Bereitstellung für die Nutzung
- Standardisierung, Normdaten und Interoperabilität
- Open-Access, Nachnutzbarkeit und Schnittstellen
- Datenanalyse, -auswertung und -anreicherung
- Recherchierbarkeit von Volltexten

Zur Umsetzung der skizzierten Anforderungen unterstützt die KLA die Einrichtung eines fachspezifischen Konsortiums aus Infrastruktur- bzw. Archiveinrichtungen sowie universitären und außeruniversitären Forschungseinrichtungen. Damit wird auch der vom Rat für Informationsinfrastrukturen (RfII) und dem Verband der Historiker und Historikerinnen Deutschlands e. V. (VHD) eingeforderten institutionellen Vernetzung und Zusammenarbeit Rechnung getragen.⁸ Ein solches Konsortium mit der Fokussierung auf die historisch arbeitenden Geisteswissenschaften formiert sich derzeit (u. a. mit Beteiligung des Landes-

archivs Baden-Württemberg) unter der Bezeichnung „NFDI-4Memory“. Ein bei der Deutschen Forschungsgemeinschaft (DFG) einzureichender Antrag auf Förderung von Projekten und Maßnahmen im Rahmen der NFDI ist für 2020 vorgesehen.⁹ Für die operative Implementierung eines Großteils der aus den angeführten Handlungsfeldern entstehenden Anforderungen bietet das Archivwesen mit dem Archivportal-D bereits eine geeignete Infrastruktur, die im Rahmen der NFDI weiterentwickelt werden kann. An dieser Stelle sollen schwerpunktmäßig zwei bedeutsame Bereiche betrachtet werden, die von archivischer Seite in den NFDI-Prozess eingebracht werden. Zum einen sind dies die wachsenden Anforderungen an die Qualität von Forschungsdaten und an deren Optimierung (etwa durch den Einsatz von Normdaten und die Nutzbarmachung der technischen Fortschritte beim Einsatz künstlicher Intelligenz). Zum anderen der Bereich der digitalen Archivierung mit der inzwischen mit vielfältigen praktischen Erfahrungen unterfütterten Expertise des Archivwesens. Der heterogenen Struktur der deutschen Archivlandschaft wird dabei ebenso Rechnung getragen wie der sehr unterschiedlichen Größe und Leistungsfähigkeit der einzelnen Einrichtungen.

Weiterentwicklung des Archivportals-D zur zentralen Forschungsdatenplattform

Überlegungen und Strategien zur Bereitstellung von Archivgut für die Forschung haben sich längst von der aus dem analogen Zeitalter stammenden Fixierung auf den persönlichen Lesesaalbesuch von Wissenschaftlerinnen und Wissenschaftlern gelöst. Die frei zugängliche Onlinebereitstellung von Erschließungsdaten in Fachinformationssystemen oder übergreifenden Portalen zählt ebenso zum Standardangebot wie das Bemühen um die Präsentation von digitalisiertem Archivgut im Internet. Konkreter werdende Konzepte für virtuelle Lesesäle vervollständigen den digitalen Zugang auch für originär digitales Archivgut.¹⁰ Mit dem Archivportal-D¹¹ steht inzwischen ein zentraler Zugang für archivische Forschungsdaten aus der gesamten Bundesrepublik zur Verfügung. Es enthält Erschließungsinformationen, digitalisiertes Archivgut sowie Archivinformationen. Seit der Freischaltung 2014 wächst das Angebot rasch, so dass inzwischen ein die Archivsparten (unter anderem staatliche und kommunale Archive, Universitäts- und Wirtschaftsarchive sowie Kirchen- und Medienarchive) übergreifender deutschlandweiter Forschungsdatenpool bereitsteht. Aktuell finden sich bereits 21 Millionen Erschließungsdatensätze – davon 1,1 Mio. mit Digitalisaten – aus knapp 200 Archiven (Stand: November 2019) sowie Kontaktinformationen zu über 2.400 Archiveinrichtungen in Deutschland. Über die gemeinsame Datenhaltung mit der Deutschen Digitalen Bibliothek (DDB) ergibt sich eine Anschlussfähigkeit der im Archivportal vorhandenen Metadaten und Digitalisate an das Angebot anderer Sparten wie etwa dem Bibliotheks- und Museumswesen. Die erfolgreiche Umsetzung, das nachhaltige Betriebskonzept und die konkreten Weiterentwicklungsperspektiven dieser Plattform belegen, dass die deutschen Archive zusammen mit ihren IT-Partnern als forschungs- und serviceorientierte Informationsanbieter leistungsfähig sind und in diesem Zusammenhang auch einen wertvollen Beitrag beim Aufbau einer NFDI leisten können.¹² Die gegenwärtige Aufstellung von Archivportal und DDB erfüllt ferner zwei wesentliche Positionen des RfII: Zum einen den Bedarf zu einer genau überlegten Ergänzung dezentraler und

parallel gewachsener Strukturen um zentrale Infrastrukturen. Andererseits die Forderung, für eine nachhaltige und auskömmliche Finanzierung von Infrastrukturprojekten jenseits der Initialförderung zu sorgen.¹³ In Gestalt des Archivportals-D und des dort vorhandenen Forschungsdatenangebots sowie der Funktionalitäten zur Recherche und Präsentation der Inhalte kann das Archivwesen bereits einen substanziellen Beitrag in das geplante Konsortium NFDI4Memory und die übergeordnete NFDI einbringen.

Zu erwähnen ist auch, dass mit der Realisierung des zentralen archivischen Nachweissystems im DDB-Kontext erhebliche standardisierende Effekte für die gesamte Sparte zu verzeichnen sind. Dies betrifft gleich mehrere mit den sogenannten FAIR-Prinzipien¹⁴ assoziierte Aspekte: In rechtlicher Hinsicht wurden sämtliche Metadaten einheitlich unter CC0 lizenziert und sind damit ohne Beschränkungen nachnutzbar. Digitalisiertes Archivgut ist auf Objektebene transparent mit den jeweiligen Nutzungsmöglichkeiten ausgezeichnet. Hierbei kommt ein fester Kanon von Lizenzvarianten zum Einsatz, einschlägige und nutzerfreundliche Ausprägungen wie zum Beispiel CC-BY sind stark vertreten. Mit Blick auf die Datenflüsse erfolgte einerseits die Community-weite Verständigung auf das einheitliche archivistische Lieferformat EAD (DDB), einem XML-Anwendungsprofil. Workflows sichern die Einhaltung von Qualitätskriterien (z. B. Verwendung persistenter Identifikatoren, Normalisierung und Anreicherung erforderlicher Metadaten), nötigenfalls durch Aufbereitung der Daten. Hierdurch wird eine recht weitgehende Einheitlichkeit und Konsistenz des Datenbestands erreicht. Zum anderen erzielte gerade die normdatenbasierte Portalfunktionalität eine regelrechte Sogwirkung, die in eine starke Dynamik hinsichtlich der Verwendung kontrollierter Vokabulare wie der Gemeinsamen Normdatei (GND) in der archivischen Erschließung mündete. Archivdaten werden damit in zunehmendem Maße interoperabel.

Wie aber gestalten sich die konkreten Ausbauperspektiven des Angebots? Hier sei zunächst auf das aktuell laufende DFG-Projekt zur Etablierung sachthematischer Zugänge im Archivportal-D am Beispiel der Weimarer Republik verwiesen. Qualifizierte themenbezogene Recherchemöglichkeiten zu Kulturgut in übergreifenden Portalen sind aus wissenschaftlicher Sicht seit langem ein Desiderat. Im Projekt werden die technischen Grundlagen hierfür geschaffen. Insbesondere wird ein Verknüpfungstool realisiert, das eine zentrale Referenzierung vorhandener Datenbestände mit kontrollierten sachthematischen Vokabularen bzw. Systematiken in effizienter Weise ermöglicht. Außerdem wird in Kooperation mit dem FIZ Karlsruhe ein Anwendungsprototyp zur (teil-)automatisierten Verschlagwortung von Archivbeständen entwickelt. Hierbei kommen maschinelle Lernverfahren sowie semantische Technologien zum Einsatz. Eine gezielte Weiterentwicklung des Archivportals-D und der assoziierten Dienste und Werkzeuge im Rahmen der NFDI könnte an Szenarien zur Datenoptimierung, Auswertung und Nachnutzung anknüpfen. Gerade Technologiefelder wie Machine Learning bzw. 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 Bildmuster- oder Objekterkennung kann eingesetzt werden, um digitalisierte Archivalien automatisiert zu klassifizieren (Typ der Vorlage, Identifikation dargestellter Objekte etc.). Für historische Forschende könnte eine signifikant verbesserte Datengrundlage bereitgestellt werden, die zugleich durch erweiterte und neue Zugangsmöglichkeiten gekennzeichnet ist.

- 4 Vgl. die Ergebnisse des DFG-Produktivpiloten zur Digitalisierung von archivalischen Quellen (<https://www.archivschule.de/DE/forschung/forschungsprojekte/digitalisierung/dfg-projekt-archivgutdigitalisierung.html>, aufgerufen am 25.11.2019). 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.
- 5 Demgegenüber spielen die auf empirischen Tests, Versuchsreihen, statistischen Erhebungen etc. basierenden Forschungsdesigns anderer Disziplinen in den historisch arbeitenden Geisteswissenschaften eine weitaus geringere Rolle. Zum Einsatz kommen sie bereits im Bereich der Digital Humanities, deren Forschungsansätze und Fragestellungen im Rahmen einer NFDI-Beteiligung von Archiven wiederum zu berücksichtigen sind.
- 6 Anstrengungen zur Digitalisierung bislang ausschließlich analoger Daten bleiben im Blick zu behalten, wenn eine überlegte und priorisierte Vorgehensweise in der Erschließung im Fokus liegt. Hierbei handelt es sich jedoch nicht um eine Anforderung im Rahmen der NFDI, wo es darum geht, Möglichkeiten zur Optimierung digital vorhandener Daten umzusetzen.
- 7 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), (https://www.bundesarchiv.de/DE/Content/Downloads/KLA/positionspapier-forschungsdateninfrastruktur.pdf?__blob=publicationFile, aufgerufen am 25.11.2019).
- 8 Vgl. dazu das Positionspapier des VHD zur Schaffung einer NFDI und das dort formulierte Ziel 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> (aufgerufen am 25.11.2019).
- 9 Vgl. den Letter of Intent unter: https://www.dfg.de/download/pdf/foerderung/programme/nfdi/absichtserklaerungen/2020/2020_nfdi_4memory.pdf (aufgerufen am 25.11.2019).
- 10 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.archivar.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.archivar.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, hrsg. vom VdA – Verband deutscher Archivarinnen und Archivre e. V., Fulda 2016, 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 (alle Seitenabrufe am 25.11.2019).
- 11 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*, hrsg. von Rainer Hering und Robert Kretzschmar, Stuttgart 2017, S. 10-35 und die übrigen Beiträge in diesem Band.
- 12 So ist z. B. für das Archivportal-D das Leibniz-Institut für Informationsinfrastruktur – FIZ Karlsruhe (<https://www.fiz-karlsruhe.de/>, aufgerufen am 25.11.2019) der maßgebliche IT-Partner als Software-Entwickler und technischer Betreiber.
- 13 Vgl. RFII (wie Anm. 3), v. a. S. 10 f., 34 f., 37-39.
- 14 FAIR = Akronym für findable, accessible, interoperable, re-usable.

Weiterentwicklungsbedarf besteht neben dem Instrumentarium zur Metadatenoptimierung auch mit Blick auf die Nachnutzbarkeit von Forschungsdaten unabhängig vom primären Zweck ihrer Entstehung. Eine zentrale Rolle kommt hier der Datenbereitstellung über standardisierte und offene Schnittstellen (APIs) zu. Diese müssen so konzipiert sein, dass relevante Datenmengen für die jeweiligen Fragestellungen eines Forschungsanliegens zur Verfügung gestellt werden können. Das kann sich sowohl auf größere Datenvolumen beziehen als auch auf die unkomplizierte Einbindung einzelner Digitalisate unterschiedlicher Provenienz in Drittanwendungen wie IIIF-Viewer oder die Auslieferung von Erschließungsinformationen als Linked Data (RDF). Die Bereitstellung archivalischer Daten über Schnittstellen ermöglicht deren Integration, Auswertung und Anreicherung in andere(n) Forschungsumgebungen und -werkzeuge(n), die zum Beispiel auf Transkriptionen, Annotationen oder Georeferenzierung der digital vorliegenden historischen Quellen abzielen.

Optimierung vorhandener und Generierung zusätzlicher Metadaten

Um das Potenzial der Recherche- und Präsentationsmöglichkeiten in übergreifenden Informationssystemen ausschöpfen zu können und zugleich neue Anwendungskontexte beispielsweise aus den Digital Humanities zu unterstützen, wird eine deutlich höhere Datenqualität und auch eine tiefere Erschließung der Archivbestände unumgänglich sein. Folgt man den FAIR-Prinzipien als Leitgedanken für die Umsetzung der NFDI, so gilt es nämlich, die Auffindbarkeit von Informationen als Basis-Anforderung zu gewährleisten. Dem lässt sich strategisch durch einen höheren Standardisierungsgrad und zugleich eine inhaltlich umfassendere Verzeichnung begegnen – beides ist mit Blick auf einen überwiegenden Teil der heute verfügbaren archivalischen Erschließungsdaten als stark ausbaufähig zu bezeichnen. Betrachten wir zunächst das Handlungsfeld Standardisierung, so kommt dem Einsatz von Normdaten eine herausgehobene Bedeutung zu. Die Verwendung von Normdaten ist ein zentrales Charakteristikum einer guten und zeitgemäßen Metadatenqualität – dies gilt domänenübergreifend und damit auch für die archivalische Erschließung. In den Regelwerken ISAD(G) (Erschließung), ISAAR(CPF) (Beschreibung von Bestandsbildnern) sowie ISDIAH (Beschreibung von Archiven als Institutionen) ist eine Normdatenreferenzierung vorgesehen. Auch wird eine solche inzwischen immer häufiger von Drittmittelgebern als verbindliches Förderkriterium verlangt. Zu verweisen ist hier insbesondere auf die entsprechenden Regelungen in den Digitalisierungsrichtlinien der DFG. Zudem erscheint der Einsatz kontrollierter Vokabulare geradezu als zwingende Voraussetzung für die Beteiligung von Archiven an vernetzten Forschungsinfrastrukturen, denn standardisierte und normierte Metadaten bilden die Grundlage der Interoperabilität von Daten. Durch die Verständigung auf die Nutzung beispielsweise der GND kann ein gemeinsames Vokabular beim Aufbau der NFDI genutzt werden. Archivbestände werden damit im Verbund sichtbar und recherchierbar. Gerade auch für kleinere Archiveinrichtungen birgt die Nutzung von Normdaten ein großes Potenzial. Sie können mit einzelnen interessanten Beständen von überregionaler Relevanz (etwa bestimmten Nachlässen) prominent aufscheinen und somit eine höhere Sichtbarkeit im Web erlangen, anstatt im Volltext-Datenozean unterzugehen.

Die GND gilt als bedeutendste institutionell gesicherte Normdatei im deutschsprachigen Raum. Längst handelt es sich bei ihr um ein spartenübergreifend einschlägiges Angebot, das als Gemeinschaftsvorhaben von Gedächtniseinrichtungen sehr nah an deren spezifischen Bedürfnissen agiert.¹⁵ Die gestiegene Bedeutung der GND lenkt aber auch einen Blick auf Probleme und noch bestehende Defizite dieses Großprojekts. Dazu gehören heterogene Datenqualität und fehlende Einträge, gerade mit Blick auf archivische Bedürfnisse. So ist die Nutzung der GND bei der Erschließung von Archivgut derzeit noch maßgeblich davon abhängig, ob die benötigten Normdaten zuvor im bibliothekarischen Kontext erfasst worden sind.¹⁶ Daher sollen nach dem Willen der Deutschen Nationalbibliothek (DNB) und ihrer Verbundpartner weitere Sparten und Institutionen aktiv und mit ihrer jeweiligen fachlichen Kompetenz in die Qualitätssicherung und den inhaltlichen Ausbau der GND einbezogen werden. Dem Thema Öffnung der GND widmet sich seit 2018 ein DFG-Projekt mit dem Titel „GND4C – GND für Kulturdaten“.¹⁷ Ausgangspunkt des Projekts ist die Überlegung, dass für eine großangelegte Partizipation anderer Kultursparten zunächst systematisch die nötigen Voraussetzungen geschaffen werden müssen – organisatorisch, inhaltlich, technisch und im Bereich der Kommunikation. Besondere Bedeutung kommt hierbei dem nachhaltigen Aufbau einer sparten- und fächerübergreifenden Organisation durch die Etablierung von sogenannten GND-Agenturen zu. Diese Agenturen sollen mit ihrem koordinativen und operativen Aufgabenportfolio sowohl Normdaten-Kompetenzzentren als auch Scharnierstellen zu den Fach-Communities sein. Damit könnten sie eine wichtige Rolle innerhalb der NFDI als zentrale Akteure und Ansprechpartner im Bereich Daten-Standardisierung und Normierung übernehmen. Wie lässt sich nun neben der Qualität der archivalischen Erschließung auch der Umfang der digital vorhandenen Erschließungsdaten substantiell und bedarfsgerecht erweitern? Ansätze hierzu wurden bereits angesprochen: Über automatisierte Verfahren, den Einsatz von Algorithmen und von künstlicher Intelligenz lassen sich zusätzliche Metadaten vor allem in Form von Schlagwörtern beziehungsweise Indexbegriffen generieren. Neben dieser punktuellen Erweiterung der digitalen Datenbasis sollte ein nächster großer Ausbauschritt im Kontext der Digitalisierung und Erschließung von Archivgut angegangen werden. Galt bisher für die Verfügbarmachung von Archivinformationen beziehungsweise von Archivgut das Vier-Stufen-Modell (1. Bereitstellung einer Online-Präsenz, 2. von Online-Bestandsübersichten, 3. von Online-Findmitteln, 4. von Digitalisaten), muss nun konsequenterweise die Bereitstellung von Volltexten in den Blick genommen werden. Denn die Zugänglichmachung von digitalisiertem Archivgut wird längst erweitert um die nutzerseitige Anforderung, auch die Volltexte zu erfassen und durchsuchbar aufzubereiten. Technologisch ist dies durch die erhebliche Verbesserung entsprechender OCR-Verfahren zur optischen Zeichenerkennung für gedrucktes jüngerer Archivgut bereits möglich, und auch für ältere, handschriftliche Dokumente durch den Einsatz von Mustererkennung mit KI-unterstützter Machine-Learning-Technologie etwa im Transkribus-Projekt¹⁸ zunehmend realistisch. Hierdurch wird der für die Auswertung verfügbare Datenumfang diversifiziert und zugleich maßgeblich ansteigen. Bereits ein „schmutziges“ OCR wäre für die Recherche nach relevantem Archivgut ein erheblicher Fortschritt, muss sich diese doch bisher auf die Suche in den Titelaufnahmen

beschränken. Die Perspektive wäre eine deutlich bessere Basis sowohl für unterschiedlichste Forschungsfragestellungen als auch beispielsweise für Editionsprojekte, wenn man die Möglichkeiten einer nachfolgenden Tiefenerschließung der Volltexte via TEI-XML einbezieht.

Langzeitarchivierung, Bereitstellung, Nachnutzung von Daten

Für den Aufbau einer NFDI in hohem Maße relevant ist ferner die in den Archiven vorhandene Expertise bezüglich der vom RFIH einmal mehr als anspruchsvoll anerkannten Langzeitarchivierung¹⁹ digitaler wie auch digitalisierter Objekte. Neben einer umfassenden theoretischen Fundierung sind zum Teil langjährige praktische Erfahrungen der Archive vorhanden und substantielle Erfolge auf diesem komplexen Handlungsfeld zu verzeichnen. In diesem Zusammenhang können die Archive neben der Absicherung gegen Datenverluste zudem ihr breites Wissen und ihre konkreten Erfahrungen in Bezug auf die Wahrung der Integrität und Authentizität der digitalen Daten in eine NFDI einbringen.

Mit Blick auf die kleinteilig strukturierte deutsche Archivlandschaft ist dabei zu betonen, dass die Herausforderungen der digitalen Archivierung zwar auf Grundlage etablierter internationaler Standards,²⁰ jedoch mit unterschiedlicher Herangehensweise und mit variablen technischen Lösungen angegangen wurden und bewältigt werden. Heute stehen in Bund, Ländern und Kommunen unterschiedliche Systeme zur Verfügung.²¹ Abhängig von der Region sowie dem institutionellen Status von Einrichtungen oder Forscherinnen und Forschern ergeben sich somit unterschiedliche Anschlussperspektiven an bestehende Infrastrukturen der digitalen Archivierung. Letztere sind dabei entweder bereits in der Lage, die Anforderungen aus der Langzeitsicherung digitaler Forschungsdaten zu erfüllen – oder zumindest hinreichend entwicklungsfähig, damit entsprechende Strukturen nicht im Rahmen der NFDI von Grund auf etabliert werden müssen. Unter sachgerechter paralleler Berücksichtigung von Strategien zur Migration und Emulation gilt dies auch für die Archivierung komplexer Forschungsdatenerzeugnisse wie etwa Datenbanken oder Onlineanwendungen unterschiedlicher Art.

Zu berücksichtigen sind in diesem Zusammenhang jenseits der Datenarchivierung auch die Erwartungen der Nutzerinnen und Nutzer bezüglich der dauerhaften Auswertbarkeit von Daten. Dabei sollen diese einerseits in ihrer ursprünglichen technischen Umgebung zugänglich und lesbar bleiben, andererseits von dort aber auch reibungslos in diverse neue Tools zur Auswertung, Datenaufbereitung und -interpretation übergehen können.

FÖRDERUNGSBEDARF

Das Förderprogramm zum Aufbau der NFDI adressiert das zentrale Dilemma der bislang zumeist projektförmigen Infrastrukturentwicklung: Jeweils mit dem Ende der überwiegend drittmittelfinanzierten Anschubförderung fehlen in der Regel tragfähige Modelle, die Projektergebnisse nicht nur zu verstetigen, sondern dauerhaft weiter zu entwickeln. Daher muss die Förderlandschaft zwingend um Komponenten zur nachhaltigen Bereitstellung und Sicherung der Projektergebnisse erweitert werden. Dazu gehören im Rahmen der geplanten NFDI adäquate Ressourcen für Infrastruktur zur Langzeitspeicherung von

Daten, für Personal zur Pflege und gegebenenfalls technischen und inhaltlichen Weiterentwicklung von Diensten sowie zur Gewährleistung von Beratungsangeboten.²² Bezogen auf den Beitrag der Archive zur NFDI könnte einer Förderung der Forschungsdatenplattform Archivportal-D eine bedeutende Hebelwirkung zukommen, würde doch nicht nur eine gesamte Sparte davon profitieren, sondern auch im übergreifenden, interdisziplinären Kontext ein vernetzter und ausbaufähiger Infrastrukturbaukasten genutzt. Bislang ist zwar durch ein Memorandum of Understanding mit der DDB der Betrieb des Archivportals-D gesichert, nicht aber die für den NFDI-Prozess notwendige stetige Fortentwicklung, um die genannten Perspektiven realisieren zu können.

Ein konkreter Bedarf ergibt sich außerdem in Anknüpfung an das DFG-Projekt „GND4C“. Die neu aufgestellte spartenübergreifende Kooperative als Trägermodell der Normdatei steht vor einer Herausforderung, für die bislang keine Lösung absehbar ist: Anders als bei der DNB und den großen Bibliotheksverbänden, wo die Pflege der GND immer Bestandteil der Regelaufgaben im Bereich der Erschließung war, gibt es aufseiten derjenigen Einrichtungen, die als „Agenturen“ für die neu einbezogenen Sparten (Archive, Museen, Denkmalpflege, Forschungseinrichtungen) agieren sollen, abseits der initialen Projektförderung keine dauerhaften Mittel für ein Betriebsmodell der „geöffneten GND“. Die Finanzierung zumindest eines

¹⁵ Die GND entstand im April 2012 aus der Zusammenführung der Personennormdatei (PND), Gemeinsamen Körperschaftsdatei (GKD), Schlagwortnormdatei (SWD) und DMA-EST-Datei (Einheitssachtitel des Dt. Musikarchivs) mit dem Ziel, die Online-Recherchemöglichkeiten zu verbessern. Vgl. den Webaufruf der GND unter https://www.dnb.de/DE/Professionell/Standardisierung/GND/gnd_node.html (aufgerufen am 25.11.2019).

¹⁶ Vgl. hierzu die Ausführungen von Susanne Laux: Von VIPs und Durchschnittsbürgern. Überlegungen zur Systematisierung von relevanten Personengruppen zur Erweiterung der Gemeinsamen Normdatei für die archivische Erschließung Transferarbeit, Archivschule Marburg 2017 (https://www.landesarchiv-bw.de/sixcms/media.php/120/63604/Transferarbeit2018_Laux.pdf, aufgerufen am 25.11.2019).

¹⁷ Detaillierte Informationen zum Vorhaben finden sich im Projektwiki unter: <https://wiki.dnb.de/pages/viewpage.action?pageId=134055796>, (aufgerufen am 25.11.2019).

¹⁸ Projektseite unter: <https://transkribus.eu/> (aufgerufen am 25.11.2019).

¹⁹ Vgl. RFIH (wie Anm. 3), S. 45-48.

²⁰ Vgl. v. a. das OAIS-Referenzmodell (<http://www.oais.info/>) sowie die Materialien des nestor-Kompetenznetzwerks Digitale Langzeitarchivierung (<https://www.langzeitarchivierung.de/Webs/nestor/DE/>, alle Seitenabrufe am 25.11.2019).

²¹ Als Beispiele für verbreitete Systeme in den Ländern und Kommunen seien exemplarisch genannt der DIMAG-Verbund (dazu übergreifend zuletzt Christian Keitel, Robert Kretzschmar: The Archiving System DIMAG and its Development through Cooperation and Collaboration: Policy: Professional and Technical Aspects, in: Born Digital in the Cloud: Challenges and Solutions. Contributions to the 21. Archival Science Colloquium/International Symposium of InterPARES Trust, hrsg. von Karen Anderson, Irmgard Becker, Luciana Duranti, Marburg 2018, S. 83-97), das Digitale Archiv NRW (<http://www.danrw.de/>) sowie auf kommunaler Ebene u. a. in NRW außerdem DIPS (vgl. Martin Hoppenheit, Christoph Schmidt, Peter Worm: Die Digital Preservation Solution (DiPS). Entstehung, Grundlagen und Einsatzmöglichkeiten eines Systems zur elektronischen Langzeitarchivierung, in: Archivar Jg. 69, 2016, Heft 4, S. 375-382, online unter: http://www.archive.nrw.de/archivar/hefte/2016/Ausgabe_4/Ausgabe_4-16.pdf (alle Seitenabrufe am 25.11.2019).

²² Konkret wären die aus Förderprogrammen insbesondere des Bundes und der Länder finanzierten Projektvorhaben um Ressourcen zur Finanzierung dauerhafter Folgekosten (Speicherkapazität, Kosten für Softwareanpassungen etc.) zu ergänzen. Geschieht dies nicht, drohen immer mehr Einrichtungen vom Fortschritt der Digitalisierung abgehängt zu werden, da sie zwar erfolgreich Projektmittel einwerben, nicht aber die nach Projektende anfallenden Dauerkosten tragen können.

Teils der GND-Agenturaufgaben, nämlich der Koordinationsaufgabe für die gesamte Sparte, wäre ein Antragsgegenstand für eine NFDI-Förderung.

Schließlich erscheint es notwendig, Mittel für die Langzeitar Archivierung von Forschungsdaten nicht nur in allen Konsortien vorzusehen, sondern die Archive auf diesem Handlungsfeld als zentrale Akteure zu etablieren. Zusätzlich zu einer dem gerecht werdenden organisatorischen Ausgestaltung der NFDI wären die Archive mit den entsprechenden Mitteln zur Wahrnehmung der damit verbundenen Koordinations- und Beratungsaufgaben auszustatten.

FAZIT

Archive verfügen Stand heute über einen der inhaltlich breitesten digitalen Forschungsdatenpools und zugleich über das Know-how, diesen dauerhaft zu erhalten und zugänglich zu machen. In dem Großvorhaben zum Aufbau der NFDI bietet sich Archiven die Chance, gemeinsam mit anderen Institutionen sowie den Nutzerinnen und Nutzern Angebote und Services forschungs- und nutzungsorientiert auszubauen und weiter zu verbessern. Archive wollen aber auch Teil der NFDI werden, um ihrem Selbstverständnis als Forschungsinfrastruktureinrichtungen weiterhin gerecht werden zu können. Dies geschieht mit der Überzeugung, dass sie auf Grundlage ihres Forschungsdatenangebots, vorhandener (sowie weiterzuentwickelnder) Services und Kompetenzen die Etablierung einer NFDI an zentraler Stelle unterstützen und bereichern können. Daher dürfte im Rahmen der NFDI kein Weg an den Archiven vorbeiführen! Ein Teil der archivischen Angebote, die methodisch die Langzeitsicherung, Zugänglichmachung und Nachnutzung von Daten betreffen, kann auch für die konsortiumsübergreifende Strukturbildung von Interesse sein. Diesbezüglich sollte von Seiten der Archive keine ausschließliche Fixierung auf die historisch arbeitenden Geisteswissenschaften erfolgen.

Umgekehrt besteht die Gefahr, dass die historisch arbeitenden Geisteswissenschaften abgehängt werden, wenn die in den Archiven vorhandenen Forschungsdaten keine Berücksichtigung im Rahmen der NFDI finden. Hier wird die erfolgreiche Realisierung stark davon abhängen, wie kleinere Einrichtungen einbezogen werden können. Denn das Archivwesen ist strukturell sehr heterogen und es wird entscheidend sein, auch im Rahmen der NFDI möglichst die gesamte Vielfalt der Sparte abzubilden. Hervorragende Anknüpfungspunkte als Dreh- und Angelpunkt der archivischen NFDI-Beteiligung bietet daher das Archivportal-D, das als gemeinschaftliches Vorhaben der gesamten Archivsparte bereits fest etabliert und zugleich organisatorisch-technisch so aufgestellt ist, dass das Angebot bedarfsgerecht ausgebaut und um benötigte Services erweitert werden kann.

ACCESS, PROCESSING, LONG-TERM PRESERVATION. FUNCTIONS OF THE ARCHIVES IN THE RESEARCH DATA INFRASTRUCTURE

As well-established memorial institutions and research infrastructures, archives are indispensable actors in research using historical methods and, in particular, in historiographical research. They are thus an integral element of the overarching research infrastructure of the Federal Republic of Germany. As a result, there is the challenge of helping to shape the development of the National Research Data Infrastructure (NFDI) and of integrating the relevant expertise and competences of the archives into the ongoing process. In addition, the establishment of a comprehensive, i. e. interdisciplinary and at the same time networked research data infrastructure requires that archives further develop their offers and services and meet new requirements. The aim of the article is to outline the activities and offers that the archives' sector can bring into an NFDI. In addition, considerations are to be given on the scope of action and the concrete need for funding. Via this scope and the related funding the contributions of the archives can be sustainably integrated into the NFDI and anchored there with the broadest possible perspectives of impact.

Professor Dr. Gerald Maier

Landesarchiv Baden-Württemberg
Eugenstraße 7, 70182 Stuttgart
E-Mail: gerald.maier@la-bw.de

Daniel Fähle

Landesarchiv Baden-Württemberg
Olgastraße 80, 70182 Stuttgart
E-Mail: daniel.faehle@la-bw.de

Dr. Andreas Neuburger

Landesarchiv Baden-Württemberg
Eugenstraße 7, 70182 Stuttgart
E-Mail: andreas.neuburger@la-bw.de