

Chancen und Herausforderungen des digitalen Wandels für Archive

Von MATTHIAS RAZUM

Einleitung: Der Digitale Wandel ist real!

*Der digitale Wandel hat Forschungsdaten und Forschungsmethoden erfasst, er verändert die Wissenschaft grundlegend,*¹ schreibt der Rat für Informationsinfrastrukturen in seiner Publikation *Leistung aus Vielfalt*. Vor dieser Entwicklung können und sollten sich die Archive nicht verschließen, denn neben großen Herausforderungen verbinden sich damit auch Chancen. Gerade kleine und mittlere Archive können ihre Bestände via Aggregatoren (z.B. Archivportal-D) über ihre bisherigen Publika hinaus sicht- und auffindbar machen. Neue Methoden wie OCR/HTR, Named Entity Recognition, semantische Technologien, maschinelles Lernen und Large Language Models unterstützen bei der Erschließung und Suche. Über Normdaten-Identifikatoren lassen sich die eigenen Bestände mit denen anderer Einrichtungen vernetzen. Dies sind wichtige Voraussetzungen für die sich gerade etablierenden Methoden und Werkzeuge der digitalen Geisteswissenschaften. Doch all diese Möglichkeiten erfordern trotz *Künstlicher Intelligenz* zusätzliche Ressourcen und neue Kompetenzen. Dieser Beitrag versucht mit einem Blick von außen Impulse zu geben, wie Archive den digitalen Wandel aktiv gestalten können.

Archive spielen als Teil der Informationsinfrastruktur eine wesentliche Rolle bei der Bewahrung des kulturellen Erbes, der Wahrung der Transparenz und Rechenschaftspflicht in Regierungen und Organisationen sowie der Unterstützung von Forschung und Bildung. Sie ermöglichen eine dauerhafte Aufbewahrung und Organisation von historisch, kulturell und administrativ wertvollen Dokumenten und Aufzeichnungen. Diese bilden unter anderem die Grundlage für die historisch arbeitende Forschung: als Quellen oder –generalisiert ausgedrückt – als Forschungsdaten.² Die digitale Verfügbarkeit dieser Forschungsdaten gewinnt zunehmend an Bedeutung. Die Deutsche Forschungsgemeinschaft schreibt in ihrem Impulspapier *Digitaler Wandel in den Wissenschaften* dazu: *Wissenschaftliches Arbeiten wird künftig in nahezu allen Disziplinen maßgeblich durch digitale Forschungspraktiken und Informationsinfrastrukturen geprägt sein.*³

¹ RfII – Rat für Informationsinfrastrukturen: Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland, Göttingen 2016, 160 S.URN: urn:nbn:de:101:1-201606229098.

² Damit sind Archivdaten (Erschließungs- und Primärdaten) gesamthaft als Forschungsdaten zu betrachten.

³ Matthias Katerbow u. a.: Digitaler Wandel in den Wissenschaften. Impulspapier, Bonn, Oktober 2020, 14 S. <https://doi.org/10.5281/zenodo.4191345> (aufgerufen am 02.08.2023).

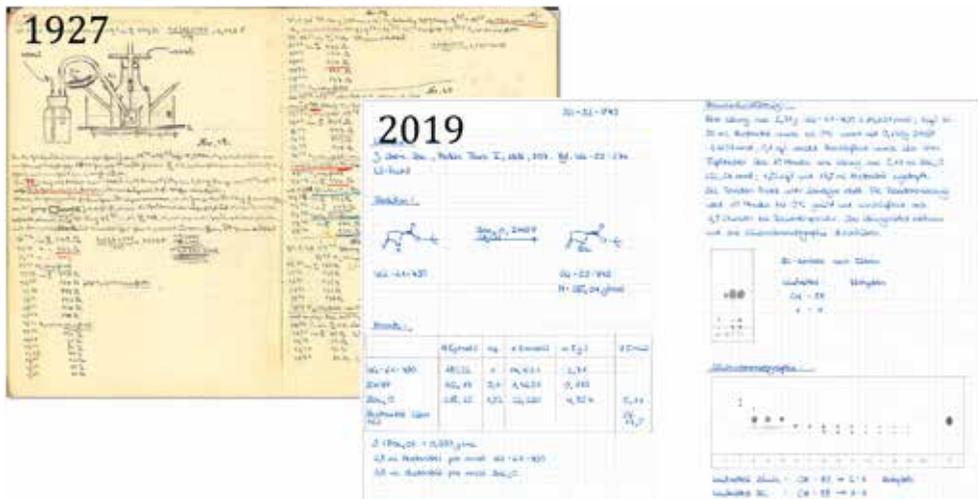


Abb. 1: Laborbücher in der Chemie einst und jetzt. Vorlage: Laborbuch am Kaiser-Wilhelm-Institut für Chemie 1929–1931 (Autor unbekannt. Archiv der Max-Planck-Gesellschaft, I. Abt., Rep. 11, Nr. 69) und Laborbuch am Institut für Organische Chemie des KIT des KIT 2009 (Autorin Simone Gräßle).

Nun gibt es bei Archiven die verbreitete Ansicht, im Bereich der Digitalisierung besonders weit zurückzuliegen, sowohl im Vergleich zu anderen Gedächtniseinrichtungen als auch zur Forschung. Sicherlich gehören Archive nicht zur Speerspitze, aber ein Beispiel aus den Naturwissenschaften, genauer der Chemie, soll aufzeigen, dass auch andere Disziplinen und Einrichtungen diesbezüglich noch vor erheblichen Herausforderungen stehen. Wie in allen laborgebundenen Disziplinen, dokumentieren Forschende in der Chemie die Planung, Durchführung und Auswertung von Experimenten und der dabei erzeugten Forschungsdaten in Laborbüchern. Zwar gibt es zwischenzeitlich eine Vielzahl an elektronischen Laborbüchern, aber in der gelebten Praxis hat sich in den vergangenen gut 90 Jahren wenig getan. Denn trotz weitgehender Digitalisierung aller Messungen erfolgt die Dokumentation immer noch meist analog auf – hoffentlich säurefreiem – Papier.

Das Beispiel zeigt, dass zwar viele Schritte im Forschungsprozess digitalisiert sind, aber gerade im Bereich der Dokumentation und nachhaltigen Bereitstellung von Ergebnissen noch ein erheblicher Nachholbedarf besteht. Ein wesentlicher Grund für dieses Desiderat besteht in der projekthaften Förderung der notwendigen Infrastrukturkomponenten. Der Rat für Informationsinfrastrukturen hat seit seiner Gründung 2014 wesentlich dazu beigetragen, auf wissenschaftspolitischer Ebene eine Veränderung herbeizuführen. Dies führte letztlich zur Etablierung der Nationalen Forschungsdateninfrastruktur (NFDI). Sie soll *Standards im Datenmanagement setzen und als digitaler, regional verteilter und vernetzter Wissensspeicher Forschungsdaten nachhaltig sichern*

*und nutzbar machen.*⁴ An der NFDI beteiligen sich zahlreiche Hochschulen sowie außeruniversitäre Forschungs- und Gedächtniseinrichtungen, die sich in insgesamt 26 fachlich orientierten Konsortien organisiert haben. Bund und Länder fördern die NFDI für zunächst zehn Jahre mit insgesamt 900 Millionen Euro.

Die historisch arbeitenden Wissenschaften (als wichtige Zielgruppe der Archive) haben sich in dem Konsortium NFDI4Memory⁵ zusammengeschlossen. Beteiligt sind hier neben weiteren Partnern das Landesarchiv Baden-Württemberg, die Archivschule Marburg, das Bundesarchiv, die Generaldirektion der Staatlichen Archive Bayerns, das Hessische Landesarchiv und der Verband deutscher Archivarinnen und Archivare, was einerseits die Bedeutung der Archive für die historisch arbeitenden Wissenschaften aufzeigt, andererseits aber die Notwendigkeit verdeutlicht, sich als Archiv auch den sich ändernden (digitalen) Forschungspraktiken anzupassen.

Herausforderungen: Der KI eine Chance!

Das Motto des 82. Südwestdeutschen Archivtags lautet *Smart und intelligent – Digitale Unterstützung für die Arbeit im Archiv*. Bei smart und intelligent drängt sich der Begriff *Künstliche Intelligenz* auf. Der BitKom e.V. und das Deutsche Forschungszentrum für Künstliche Intelligenz definieren sie als *Eigenschaft eines IT-Systems*, „*menschenähnliche*“, *intelligente Verhaltensweisen zu zeigen*. Tatsächlich gibt es aber bereits seit Jahren etablierte Verfahren, die einen weniger hohen Anspruch verfolgen, aber sinnvolle Arbeitserleichterungen für Archive bieten.

Named Entity Recognition (NER) ist eine Technik der Sprachverarbeitung zur Erkennung und Klassifizierung benannter Entitäten (wie Namen von Personen, Orten oder Organisationen) in einem Text. Sie nutzt meist linguistische oder statistische Regeln, um Entitäten zu klassifizieren.

Maschinelles Lernen ist ein breiterer Ansatz, der darauf abzielt, Algorithmen und Modelle zu entwickeln, die aus Beispielen (*Trainingsdaten*) lernen können. Auch hier kommen statistische Verfahren wie z. B. Gradient Descent oder Backpropagation zum Einsatz. Maschinelles Lernen ermöglicht es, Muster in großen Datenmengen zu erkennen und Vorhersagen über bislang unbekannte Daten zu treffen. Viele moderne Ansätze bei der Optical Character Recognition (OCR) und Handwritten Text Recognition (HTR) setzen auf Verfahren des Machine Learnings auf.

Maschinelles Lernen ist neben symbolischer Wissensrepräsentation und Logik, Robotik, Natural Language Processing etc. ein Teilgebiet der *künstlichen Intelligenz* (KI). Bei ihr unterscheidet man zwischen starker und schwacher KI. *Starke KI* zielt auf eine dem menschlichen Denken ebenbürtige Intelligenz. Sie ist bislang und bleibt in absehbarer Zukunft ein theoretisches Konstrukt. Geforscht wird an *schwacher KI*, die auf die Ausführung bestimmter Aufgaben

⁴ Gemeinsame Wissenschaftskonferenz: Pressemitteilung: Forschungsdaten nachhaltig sichern und nutzbar machen – Startschuss für eine Nationale Forschungsdateninfrastruktur. Hg. von GWK. Bonn 16. November 2018. <https://www.gwk-bonn.de/fileadmin/Redaktion/Dokumente/Pressemitteilungen/pm2018-13.pdf> (aufgerufen am 02.08.2023).

⁵ <https://4memory.de/>

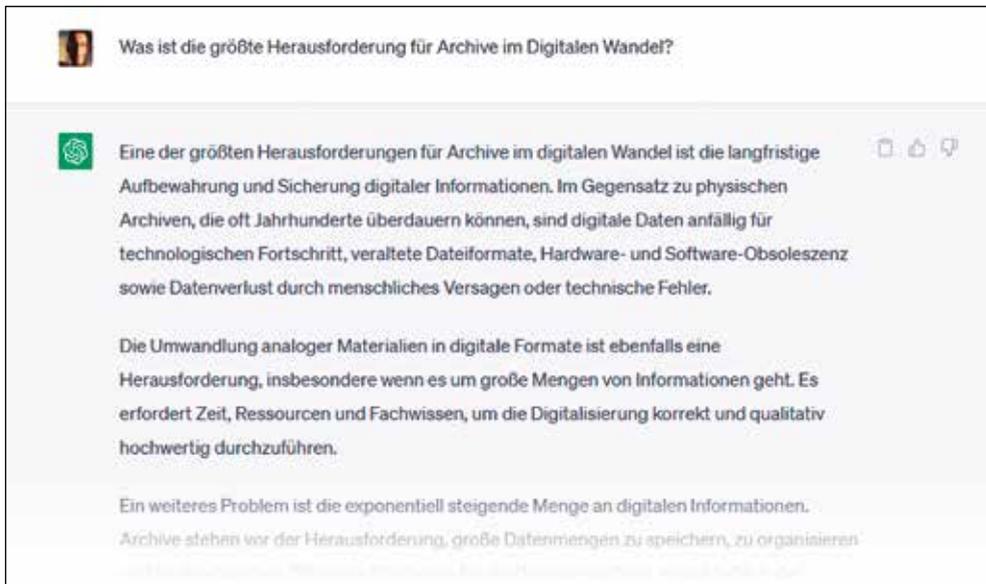


Abb. 2: Screenshot von chat.openai.com (Vorlage: Autor)

trainiert und ausgerichtet ist. In letzter Zeit stößt der Begriff der künstlichen Intelligenz auch jenseits der Fachwelt auf breite Resonanz, insbesondere durch die rasanten Entwicklungen bei den sogenannten *Large Language Models*, die mit extrem umfangreichen Daten vortrainiert und anschließend (oft in aufwändigen intellektuellen Prozessen) für ihre jeweilige Aufgabenstellung feinabgestimmt werden. Eine der bekanntesten Anwendungen eines solchen Large Language Models ist ChatGPT der Firma OpenAI.⁶

Was liegt also näher, als ChatGPT zu fragen, worin denn nun die größte Herausforderung für Archive im digitalen Wandel besteht? Da ChatGPT ein Dialogsystem ist, muss zunächst ein passender Prompt, also letztlich eine Aufgabenstellung, formuliert werden: Was ist die größte Herausforderung für Archive im digitalen Wandel? Das Ergebnis ist ein längerer Text (Abb. 2) mit gleich mehreren Herausforderungen, die sich so zusammenfassen lassen:

- Langfristige Aufbewahrung und Sicherung digitaler Informationen
- Umwandlung großer Mengen analoger Materialien in digitale Formate
- Exponentiell steigende Menge an digitalen Informationen
- Datenverwaltung, geeignete Metadatenstandards und Suchfunktionen
- Sicherung der Integrität digitaler Informationen

⁶ <https://chat.openai.com/> (aufgerufen am 02.08.2023).

- Datenschutz und Informationssicherheit
- Langfristige Nachhaltigkeit digitaler Archive

Das Ergebnis ist zunächst beeindruckend. Alle aufgeführten Punkte sind valide, aber bei genauem Hinsehen fällt auf, dass einige aus meiner Sicht zentrale Aspekte fehlen, die ich hier ergänze:

- Sichtbarkeit von Archiven im digitalen Raum
- Weiterentwicklung der analog/digitalen Schnittstelle von historischem Quellenmaterial und Daten
- Rasanz der technischen Entwicklung im Widerspruch zur langfristigen Ausrichtung von Archiven
- Bereitstellung und Unterhaltung notwendiger technischer Systeme
- Qualifikation der Mitarbeitenden
- Limitierte zeitliche und finanzielle Ressourcen

Abstrahierend, kann man aus den Herausforderungen fünf Handlungsfelder ableiten: Erschließung, Interoperabilität, technische Infrastruktur, Sichtbarkeit und fehlende Ressourcen. Im Folgenden werde ich auf diese detaillierter eingehen.

Handlungsfelder: Abwarten ist keine Lösung!

Erschließung

*Simply making the data available is insufficient for the coherent sharing and interpretation of that data.*⁷ Die Erschließung von Archivgut und damit dessen Kontextualisierung ist archivischer Alltag. Sie beschreibt (Verzeichnung) und ordnet es (Klassifizierung) mit dem Ziel, Findmittel für die Recherche bereitzustellen.

Archive zeichnen sich durch ihre typischerweise unikalen Bestände aus. Entsprechend haben sie lange Zeit ihre Erschließungsregeln aus einer Binnensicht heraus entwickelt. Da in der analogen Welt die Nutzenden jedes Archiv einzeln aufsuchen mussten, war das auch eine nachvollziehbare und aus der lokalen Perspektive logische Vorgehensweise. ISAD(G) beförderte dann eine auf dem Provenienzprinzip basierende Standardisierung bei der Erschließung, die eine wichtige Grundlage für den digitalen Wandel in Archiven bildet. Die Recherche erfolgt nicht mehr vor Ort, sondern im Netz. Sie ist nicht mehr auf ein Archiv beschränkt. Das nächste Archiv ist nur ein Browserfenster entfernt oder über Aggregatoren wie Archive NRW, Archivportal-D oder Archives Portal Europe⁸ direkter Bestandteil der Recherche. Die Vernetzbarkeit der Erschließungs-

⁷ Fran Berman, Ron Wilkinson, John Wood: Guest editorial: Building global infrastructure for data sharing and exchange through the Research Data Alliance. In: D-Lib Magazine 20(1/2). DOI: <https://doi.org/10.1045/january2014-berman> (aufgerufen am 06.08.2023).

⁸ <https://www.archive.nrw.de/>, <https://www.archivportal-d.de>, <https://www.archivesportaleurope.net/> (aufgerufen am 06.08.2023).

informationen über Archiv- oder sogar Spartengrenzen hinweg wird damit essentiell: benannte Entitäten wie Personen oder Orte müssen übergreifend und eindeutig identifizierbar sein. *Data sharing and exchange allow us to uncover connectedness in what was previously unconnected.*⁹ Das kann nur funktionieren, sofern sich Rechercheergebnisse aus unterschiedlichen Archiven in einer Treffermenge sinnvoll zusammenführen und verarbeiten lassen. Hierfür müssen dieselben Entitäten über Bestands- und Archivgrenzen hinweg als identisch erkannt werden können. Die Benennung der Entitäten alleine (also die Zeichenkette) ist dafür ungeeignet. Für eine Entität gibt es oft unterschiedliche Schreibweisen des Namens. Andererseits gibt es gleichlautende Namen unterschiedlicher Personen. Normdaten-Identifikatoren¹⁰ ermöglichen es, Entitäten unabhängig von der Schreibweise des Namens eineindeutig zu identifizieren und gleichzeitig von namensgleichen anderen Entitäten zu disambiguieren. Erst mit der Auszeichnung zentraler Entitäten mit Normdaten-Identifikatoren gelingt eine Vernetzung von Archivgut über Archivgrenzen hinweg. Sie ist ein wesentlicher Schritt, um Bestände interoperabel und damit auch maschinenlesbar zu gestalten (siehe dazu den nächsten Abschnitt).

Vor der Auszeichnung von Entitäten mit Normdaten-Identifikatoren steht deren Erkennung. Im nächsten Schritt muss der richtige Identifikator gefunden werden. Beides sind aufwändige intellektuelle Prozesse, die Zeit und damit Ressourcen kosten. Mit der *Named Entity Recognition* (NER) existiert ein Verfahren der Computerlinguistik, um benannte Entitäten in einem Text zu identifizieren und zu klassifizieren. Diese Entitäten können Personen, Orte, Organisationen, Daten oder ähnliches sein. Der NER-Algorithmus analysiert den Text, erkennt bestimmte Muster und Zusammenhänge, um die Entitäten zu extrahieren und sie entsprechenden Klassen zuzuordnen.

Zunächst benötigt man aber einen Text, auf dem die Algorithmen arbeiten können. Typischerweise funktionieren die Algorithmen am besten auf Fließtext. Verfahren der *Optical Character Recognition* (OCR) und der *Handwritten Text Recognition* (HTR) können in Scans von gedrucktem oder handschriftlichem Archivgut Text erkennen und maschinenlesbar extrahieren.¹¹ Hierauf kann nun die NER laufen. Alternativ kann dieser Schritt der Entitätenerkennung auch intellektuell im Rahmen der Erschließung erfolgen.

Im nächsten Schritt gilt es, den passenden Normdaten-Identifikator zu finden. Hierfür existieren Recommender-Systeme, die mögliche Kandidaten über eine Recherche suchen, oft unter Einbeziehung weiterer Metadaten wie Laufzeit, Geburtsdaten o.ä., um zu besseren Ergebnissen bei der Disambiguierung zu gelangen. Als Ergebnis erhält man eine Liste mit Vorschlägen von Normdaten-Identifikatoren. Die letztendliche Auswahl bleibt den Archivarinnen und Archivaren vorbehalten.

⁹ Fran *Berman*, Ron *Wilkinson*, John *Wood*, wie Anm. 7.

¹⁰ Ein prominentes Beispiel für Normdaten-Identifikatoren ist die Gemeinsame Normdatei (GND), siehe <https://gnd.network/> (aufgerufen am 07.08.2023).

¹¹ Für einen tiefergehenden Einblick in die Thematik sei auf den Beitrag von Dr. Regina *Keyler* und Dorothee *Huff*, in diesem Band verwiesen.

Alle genannten Werkzeuge können die Erschließung unterstützen und beschleunigen. Bei ihrem Einsatz sollte man sich aber immer gewahr sein, dass die eingesetzten Verfahren fehlerbehaftet arbeiten. Eine intellektuelle Prüfung der Ergebnisse ist aktuell noch angezeigt.

Interoperabilität

Die FAIR-Prinzipien¹² formulieren vier Grundsätze mit dem Ziel, Forschungsdaten (also auch die Bestände der Archive) für Forschende zugänglicher und wiederverwendbarer zu machen. FAIR steht für *findable* (auffindbar), *accessible* (zugänglich), *interoperable* (interoperabel) und *reusable* (wiederverwendbar). Die Vernetzbarkeit von Daten und technischen Systemen im Sinn der Interoperabilität ist also ein wesentliches Prinzip einer digitalen Informationsinfrastruktur.

Die Voraussetzungen in Archiven sind dabei gar nicht so schlecht: mit ISAD(G) gibt es einen internationalen Anwendungsstandard zur Verzeichnung archivischer Unterlagen. Mit EAD und dem abgeleiteten Profil EAD(DDB) existiert ein standardisiertes Metadaten-Schema für den Austausch archivischer Erschließungsinformationen. Schwieriger wird es bei der Interoperabilität der eingesetzten Systeme zur Erschließung und Archivverwaltung. Sie mussten bislang nur innerhalb des Archivs funktionieren. Eine Anbindung an das Netz spielte lange Zeit keine Rolle. Um etwa Digitalisate über das Netz abzurufen, kommen heute meist der DFG-Viewer (via METS/MODS) oder ein IIIF-Viewer (via IIIF Image API)¹³ zum Einsatz. Die wenigsten archivischen Fachinformationssysteme (AFIS) bieten jedoch schon die notwendigen Schnittstellen an. Erschwerend kommt hinzu, dass einige Archive veraltete AFIS-Versionen einsetzen, die z. B. noch Probleme mit dem fehlerfreien Export von EAD(DDB)-Daten haben. Diese Probleme lassen sich auf der Ebene einzelner Archive nur schwer lösen. Der Wechsel auf aktuelle Softwareversionen ist mit erheblichem Ressourcenaufwand verbunden. Der Funktionsumfang der eingesetzten AFIS lässt sich nur schwer beeinflussen.

Sofern möglich, können notwendige Anpassungen durch Einbeziehung von Dienstleistern (idealerweise den Herstellern der AFIS) beauftragt werden. Eine Alternative ist, Teile der Aufgaben an Aggregatoren zu übertragen. Die Fachstelle Archiv der Deutschen Digitalen Bibliothek (und damit des Archivportals-D) hat viel Erfahrung mit der Aufbereitung nicht vollständig standardkonformer Daten, um sie in normalisierter Form in das Portal einzuspielen und abrufbar zu machen. Die Bereitstellung von Digitalisaten über standardisierte Schnittstellen kann eventuell ebenfalls ein Aggregator oder Dienstleister übernehmen. Und dann gibt es noch die NFDI mit ihren fachlichen Konsortien. Das Konsortium NFDI4Memory integriert erstmals historische Forschungs-, Gedächtnis- und Informationsinfrastrukturen in einer digitalen Forschungsdateninfrastruktur und ermöglicht durch die Entwicklung innovativer Methoden zum Forschungs-

¹² Mark D. *Wilkinson* u. a.: The FAIR Guiding Principles for scientific data management and stewardship. In: *Scientific Data*, 3 (1), 160018. 2016. DOI: <https://doi.org/10.1038/sdata.2016.18>.

¹³ International Image Interoperability Framework, siehe <https://iiif.io/> (aufgerufen am 06.08.2023).

datenmanagement und Services die nachhaltige Transformation historisch arbeitender Geisteswissenschaften. NFDI4Memory verfolgt zwei Ziele:

- Es erschließt auf der Basis zunehmend digitalisierter und digitaler Daten sowie digitaler Methoden durch optimierte Forschungsdateninfrastrukturen und Dienste völlig neue Bereiche für die historische Forschung. Das bildet die Voraussetzungen dafür, dass eine digital heterogene historische Forschung sich zukunftsfähig erneuert.
- Es verhindert eine methodologische Lücke, indem es gezielt digitale und analoge Methoden zusammenführt, Kompetenzen im Bereich der digitalen Transformation auf- und ausbaut (z. B. digitale Quellenkritik) sowie einen grundlegenden fachkulturellen Wandel vorantreibt.

Eine der zentralen Maßnahmen ist die Etablierung eines sogenannten Data Spaces, der vielfältige Datenquellen interoperabel virtuell zusammenführt und recherchierbar macht. Hier wird also die Forschungsdateninfrastruktur von NFDI4Memory die notwendigen Schnittstellen zur Sicherstellung von Interoperabilität bereitstellen.

Abhängig wird der Erfolg aber von der Qualität der bereitgestellten Erschließungsdaten sein. Wenn Archive aus Ressourcengründen nicht in der Lage sind, Interoperabilität auf Systemebene herzustellen, sollten sie sich auf ihre Kernkompetenzen fokussieren (Verzeichnung, Überlieferungsbildung und Bestandserhaltung) und in anderen Bereichen auf Kooperationspartner setzen. Das soll nun nicht als Aufruf zur Inaktivität missverstanden werden, sondern als Brücke auf dem Weg zu einem zeitgemäßen AFIS, welches die Herausforderungen im Bereich der Interoperabilität adäquat adressiert.

Technische Infrastruktur

Die Verwaltung von Tektonik, Beständen und Verzeichnungseinheiten mit einem Programm zur Textverarbeitung oder Tabellenkalkulation zu organisieren, reicht nicht aus, um den Anforderungen des digitalen Wandels gerecht zu werden. Genau das ist aber für viele Archive gelebte Realität – nicht aus freien Stücken, sondern vor allem durch fehlende finanzielle und technische Ressourcen bedingt.¹⁴ Selbst viele existierende AFIS kämpfen noch mit der Digitalisierung. Sie sind primär auf die interne Nutzung im Archiv vor Ort und nicht über das Internet ausgelegt. Meist sind nachträglich aufgesetzte Zusatzmodule und -dienste erforderlich, die eigenen Bestände im Internet anbieten zu können. Dabei bleiben grundlegende Anforderungen an Vernetzung, persistente Identifikatoren, sinnvoller Zugriff auf Digitalisate und eine befriedigende User Experience (also ein benutzungsfreundliches Interaktions- und Webdesign) auf der Strecke.

Die Hersteller kommerzieller Archivverwaltungssysteme passen ihre Systeme zunehmend den neuen Anforderungen an, aber es ist ein langsamer Prozess. Zudem erfordert er, dass die Archive

¹⁴ Thekla Kluttig: Umfrage zur Erschließung: IT-Fachpersonal und Erschließungszielgruppe(n). 2019. <https://saechsischer-archivtag.vda-blog.de/2019/05/01/umfrage-zur-erschliessung-it-fachpersonal-und-erschliessungszielgruppen/> (aufgerufen am 04.08.2023).

auch in der Lage sind, auf neuere Software-Versionen umzusteigen. Ein wirklich *digitales* AFIS sollte neben den archivfachlichen mindestens folgende technische Anforderungen abdecken:

- eine Digitalisaterverwaltung, die die Verknüpfung von Erschließungsinformationen mit Digitalisaten vereinfacht und den Speicherort der Digitalisate flexibel handhaben kann, also auch außerhalb des AFIS;
- eine Präsentationsplattform für Digitalisate, die den Zugriff über das Netz erlaubt, Zugriffskopien auch in höheren Auflösungen bereitstellen kann, standardisierte Schnittstellen wie IIIF unterstützt und automatisch aus den Masterdateien der Digitalisate geeignete Zugriffskopien erzeugt (etwa in Kachelformaten wie Pyramid Tiff oder JPEG2000);
- neue unterstützende Methoden bei der Erschließung wie NER (mit Schnittstellen zu OCR/HTR-Pipelines) und Recommender-Funktionen; sowie
- Vernetzbarkeit mit Beständen anderer Einrichtungen und Internetressourcen, was die Unterstützung von Normdaten-Identifikatoren, die Einbindung von normierten Vokabularen und Ontologien, die Verwendung stabiler Identifier und die Zugriffsmöglichkeit über stabile URIs auf möglichst alle Knoten der Tektonik und Bestände bedingt.

Darüber hinaus sind Schnittstellen zur Langzeitarchivierung und die entsprechende Aufbereitung der Erschließungsdaten für eine zeitgemäße (semantische) Suche in einem digitalen Lesesaal oder bei Aggregatoren wie dem Archivportal-D wünschenswert.

Gerade für kleine und kleinste Archive stellt sich hier unweigerlich die Frage, wo ein solches AFIS herkommen soll. Meist sind die Systeme kommerzieller Anbieter zu mächtig (und zu teuer). Auch fehlen oft die technische Infrastruktur und die Kompetenz, solche Systeme im eigenen Haus zu betreiben. Hier sei auf den Beitrag von Andreas Neuburger in diesem Band zum einfachen Erschließungs- und Zugriffssystem EEZU verwiesen.

Sichtbarkeit

Nur was Google findet, existiert,¹⁵ so titelte Adrian Lobe etwas überspitzt in der Süddeutschen Zeitung. Das diese Behauptung nicht von der Hand zu weisen ist, zeigt die wirtschaftliche Bedeutung der Optimierung von Webseiten für Suchmaschinen (Search Engine Optimization – SEO). Hier hat sich in kurzer Zeit eine neue Dienstleistungssparte etabliert, um die Sichtbarkeit im digitalen Raum zu gewährleisten und zu verbessern. Auch wenn die heutigen Archivbenutzerinnen und -benutzer überwiegend noch anders sozialisiert sind und für sie die gezielte Recherche vor Ort selbstverständlich ist, werden sich die Archive dem Trend der zunehmenden Abhängigkeit von Suchmaschinen auf Dauer nicht entziehen können.

Die Minimalvoraussetzung für die Sichtbarkeit im Internet sind digitalisierte Findmittel. Schon 1999 forderte der ehemalige Präsident des Bundesarchivs: *Die Basisinformationen über Archive*

¹⁵ Adrian Lobe: Nur was Google findet, existiert. In: Süddeutsche.de (2. März 2019). <https://sz.de/1.4349158> (aufgerufen am 14.08.2023).

und die Nutzerhinweise gehören zu 100% ins Internet, Beständeübersichten ebenfalls zu 100%, Findmittel über uneingeschränkt nutzbare Bestände im Prinzip ebenfalls.¹⁶ Zwanzig Jahre später können laut einer Umfrage des Landesverbands Sachsen im Berufsverband VdA¹⁷ mehr als 75 % der befragten Archive immer noch nicht ihre Beständeübersichten und Findmittel online zur Recherche anbieten. Hier besteht also noch ein erheblicher Nachholbedarf auf Seiten der Archive. Hinzu kommt dann die für Suchmaschinen optimierte Darstellung von Beständen und Verzeichnungseinheiten in Online-Systemen. Während das für große Archive noch leistbar sein mag, sind mittlere und kleine Archive damit meist überfordert. Erneut bietet sich hier aber die Kooperation mit Partnern wie den schon mehrfach genannten Aggregatoren an. Ihre Portale schaffen Sichtbarkeit – beteiligen Sie sich also durch die Lieferung digitaler Findmittel!

Fehlende Ressourcen

Damit kommen wir zum Elefanten im Raum: fehlende Ressourcen. Kleine Archive können mit dem schnellen technischen Wandel kaum Schritt halten. Das liegt insbesondere an fehlender IT-Unterstützung. Die beiden Beiträge von Florian Spiess und Tobias Hodel in diesem Band zeigen die Dynamik bei der Entwicklung des maschinellen Lernens auf, aber auch die Komplexität der Thematik. Ein Landesarchiv kann ein *Data Lab* etablieren, um neue Verfahren zu erproben, aber welches kleinere Archiv vermag das zu leisten? Wenn also Archive, wie oben ausgeführt, Anbieter von Forschungsdaten sind, dann ist *enge Zusammenarbeit zwischen Informationsanbietern und Forschung [...] unabdingbar, um die infrastrukturellen Voraussetzungen für die Zugänglichkeit und Nutzbarkeit zu schaffen*.¹⁸ Die Forschung als eine zentrale Nutzungsgruppe nimmt eine neue Rolle ein: *Kennzeichnend für den digitalen Wandel ist, dass die herkömmliche Aufteilung zwischen Anbietern und Nutzern nicht mehr trägt*.¹⁹ Das bereits erwähnte Konsortium NFDI4Memory plant den Aufbau von übergreifenden Strukturen für die Erprobung neuer Methoden (*4Memory Methods Lab*) und zugehöriger Algorithmen (*4Memory Data Lab*) in enger Kooperation unter anderem mit Archiven und Aggregatoren. Mit einer *Virtual Indexing Layer* will das Konsortium langfristig Erschließungsdaten, die in Forschungsprojekten entstehen und bestehende Findmittel ergänzen, für die datengebenden Archive, aber auch Dritten verfügbar

¹⁶ Harmut Weber: Digitale Repertorien, virtueller Lesesaal und Praktikum im WWW – neue Dienstleistungsangebote der Archive an die Forschung. In: Fundus – Online Forum für Geschichte, Politik und Kultur der späteren Neuzeit 2/2000. S.1–17, <http://webdoc.sub.gwdg.de/edoc/p/fundus/4/weber.pdf> (aufgerufen am 18.02.2020).

¹⁷ Thekla Kluttig: Archive im Umbruch. Zur Situation in Sachsen. In: Archive im Umbruch, 22. Sächsischer Archivtag. Hg. vom Landesverband Sachsen im Verband deutscher Archivarinnen und Archivare e.V. Dresden 2018. S.21–32, https://www.vda.lvsachsen.archiv.net/fileadmin/user_upload/Tagungsband_LAT_Sachsen_2017.pdf (aufgerufen am 18.02.2020).

¹⁸ Matthias Katerbow u. a., wie Anm. 3.

¹⁹ Ebd.

machen. Hier bieten sich Chancen gerade für kleine Archive, von der bereitgestellten technischen Infrastruktur zu profitieren und so eigene Erschließungsdaten mit neuartigen Verfahren anzureichern.

Die Archive sitzen auf einem Datenschatz. Diesen zu heben, verschafft der Forschung die notwendige Grundlage, um neue Verfahren zu entwickeln, um etwa die Erschließung zu vereinfachen oder in Teilen sogar zu automatisieren. Die Verfahren werden noch auf Jahre hinaus fehlerbehaftet arbeiten. Dabei sollte man aber zwei Dinge im Auge behalten: auch die manuelle Erschließung ist keineswegs fehlerfrei. Und digital verfügbare Findmittel mit einer gewissen Fehlerquote sind für die Anwenderinnen und Anwender hilfreicher als nicht existierende Findmittel. Insofern lohnt es sich in jedem Fall, die Entwicklung neuartiger Verfahren für die teilautomatisierte Erschließung von Archivgut zu beobachten. Sie mögen nicht *ab kommendem Montag* (wie auf dem 82. Südwestdeutschen Archivtag aus dem Auditorium gelegentlich gefordert), aber doch mittelfristig Archive bei einer ihrer Kernaufgaben unterstützen und Ressourcen freisetzen. Dazu müssen diese Verfahren aber auch in den von Archiven eingesetzten Systemen *ankommen*. Hier kann nur dazu ermutigt werden, das Gespräch mit den Anbietern von AFIS zu suchen und Innovationen einzufordern, aber gleichzeitig den Kontakt zu und Austausch mit überregionalen Initiativen und Projekten zu suchen.

Fazit: Die Chancen wahrnehmen!

Es gibt keinen Grund, vor diesen zugegebenermaßen dynamischen und technisch oft komplexen Entwicklungen zu resignieren. Die digitale Transformation fordert Archive heraus, aber sie bietet gleichzeitig auch viele Chancen. Um diese trotz knapper Ressourcen nutzen zu können, bietet sich die Kooperation mit anderen Archiven, Infrastruktureinrichtungen, kommerziellen Dienstleistern und vor allem Aggregatoren an. Schließen möchte ich daher mit den Worten von Friedrich Schiller: *Verbunden werden auch die Schwachen mächtig.*²⁰

²⁰ So die Figur Werner Stauffacher in Friedrich Schiller: *Wilhelm Tell* I,3.