

## Konferenz der Leiterinnen und Leiter der Archivverwaltungen des Bundes und der Länder (KLA)

### **Zum Umgang mit Digitalisaten: Sachstand und Perspektiven**

#### ***Eine Bestandsaufnahme innerhalb des staatlichen Archivwesens 2017/18***

verabschiedet von der 129. KLA am 17.09.2019 in Suhl

Mit dem im Jahr 2016 abgeschlossenen DFG-Projekt zur Entwicklung eines Produktivpiloten zur Digitalisierung von archivalischen Quellen wurde die Grundlage zur Einrichtung einer Förderlinie zur Digitalisierung von Archivgut erarbeitet. Darauf aufbauend konnten Fördermittel für Digitalisierungsprojekte bei der DFG beantragt werden. Ergänzend zu den ausführlich dokumentierten Projektergebnissen und daran anknüpfend legt der KLA-Ausschuss AFIS mit dem vorliegenden Bericht zum Thema der Verknüpfung von Digitalisaten in archivischen Fachinformationssystemen eine Bestandsaufnahme zum derzeitigen Umgang mit Digitalisaten vor.

Das vorliegende Papier gibt im ersten Teil einen Überblick über die aktuelle Praxis; darüber hinaus wird im zweiten Teil herausgearbeitet, welche Perspektiven hinsichtlich der Verknüpfung von Digitalisaten mit Erschließungsinformationen in Zukunft erwachsen könnten. Um Basisinformationen über den Umgang mit Digitalisaten zu sammeln, hat der KLA-Ausschuss AFIS im Herbst 2017 eine Umfrage unter den Archivverwaltungen des Bundes und der Länder durchgeführt und verschiedene Fragen zur Praxis der jeweiligen Häuser gestellt.<sup>1</sup>

### **1. Sachstand**

Die im Folgenden vorgestellten Strategien und Vorgehensweisen beim Umgang mit Digitalisaten wurden auf Grundlage der Umfrage herausgearbeitet. Dabei liegt der Fokus auf der Einbindung dieser Digitalisate in archivische Fachinformationssysteme.

#### **Digitalisierungsstrategie (allg.)**

Die Frage, ob die Digitalisierung als Methode der Bestandserhaltung (statt Mikroverfilmung) betrachtet wird, oder ob sie v. a. der Bereitstellung von Nutzungsrepräsentationen bzw. zur Verbesserung der Nutzbarkeit des analogen Archivguts dient, wird von den meisten Archivverwaltungen ähnlich beantwortet: Immer werden die Bestandserhaltung *und* die Nutzung als wesentliche Motive für die Digitalisierung genannt. Obgleich die Schutzdigitalisierung ursprünglich von der Bestandserhaltung motiviert war, wird aber in den meisten Fällen die Nutzbarkeit von online bereitgestellten Digitalisaten deutlich gegenüber den Zielen der Bestandserhaltung betont. Einen zusätzlichen Vorteil der Digitalisierung sehen einige Archive darin, dass auch in den Lesesälen die als im Komfort reduziert empfundenen Mikroformen durch die Digitalisate abgelöst werden können.

---

<sup>1</sup> Folgende Archivverwaltungen haben sich an der Umfrage beteiligt: Berlin, Baden-Württemberg, Brandenburg, Bremen, Bund, Hamburg, Hessen, Niedersachsen, Nordrhein-Westfalen, Sachsen, Thüringen.

## **Digitalisierung von Gesamt- oder Teilbeständen**

Auf die Frage, ob nur ganze Bestände oder auch Teile von Beständen digitalisiert werden, antworten die teilnehmenden Archivverwaltungen sehr unterschiedlich. Als idealer Weg wird die Komplettdigitalisierung ganzer Bestände genannt, doch ist es oft unvermeidlich, dass auch Teilbestände digitalisiert werden. Schließlich gibt es Fälle, in denen man aus grundsätzlichen Erwägungen nur einen Teil eines Archivaes in die Digitalisierungsmaßnahme einbezieht. Dies wird beispielsweise im Fall von Meldebögen der Spruchkammern oder bei in Akten befindlichen Karten praktiziert. Auch wirtschaftliche Gründe spielen bei der Erstellung von Digitalisaten eine Rolle. Ist z. B. die Digitalisierung „on demand“<sup>2</sup> des gesamten Archivaes wirtschaftlicher als die Erstellung von Einzelseiten, wird das ganze Archivale digitalisiert. In diesem Zusammenhang stellt sich jedoch stets die Frage nach der Übernahme der Kosten.

## **Datenhaltung**

Die Datenhaltung der Digitalisate erfolgt in der Regel außerhalb des AFIS an einem externen Speicherort (z. B. im digitalen Magazin); die Daten werden per Link mit dem AFIS verknüpft. Die meisten Archivverwaltungen nutzen (noch) keine Persistent Identifier oder Permalinks für die Adressierung der Digitalisate.

## **Erschließungstiefe**

Eine Festlegung darüber, wie tief die zu digitalisierenden Bestände oder Teile von Beständen erschlossen sein müssen, gibt es in den Archivverwaltungen nicht. Als Minimalkonsens aus der Befragung ist festzustellen, dass mindestens ein Verzeichnungsdatensatz vorhanden sein muss. Digitalisate werden in der Regel an die Verzeichnungseinheit angebunden, d.h. eine tiefere Anbindung (z. B. pro Amtsbucheintrag, pro Vorgang, pro Aktenblatt) ist derzeit bei den meisten Archivverwaltungen nicht umgesetzt.

## **Anwendung des Repräsentationenmodells**

In einigen Archivverwaltungen wird das Repräsentationenmodell in der Form angewendet, dass die vorhandenen Repräsentationen im Archivinformationssystem nachgewiesen werden. Einige der Befragten geben an, das Repräsentationenmodell in absehbarer Zukunft einführen zu wollen bzw. darüber nachzudenken. Momentan werden aber Nutzungsdigitalisate in den meisten Fällen mit dem Erschließungsdatensatz des Archivaes verknüpft. Bislang nutzt dafür niemand die Ebene der einzelnen Repräsentationen.

## **Umgang mit Hybridakten**

Zum Umgang mit Hybridakten liegen noch wenige Erfahrungswerte vor. Eine Digitalisierung des analogen Teils der Akten ist aktuell nicht geplant. Eine technische Verknüpfung mit dem

---

<sup>2</sup> Digitalisierung „on demand“ wird in den Archivverwaltungen durchaus unterschiedlich verstanden und gehandhabt. Meist wird sie im Rahmen von kostenpflichtigen Reproduktionsaufträgen angeboten. Selten wird sie dazu genutzt, Digitalisate auf Nutzer\*innenwunsch kostenfrei zu erstellen, und sie im Anschluss direkt für einen Zugriff über das Internet zur Verfügung zu stellen.

genuin elektronischen Teil der Akte im AFIS ist in der Regel nicht vorgesehen, maximal ein Hinweis auf die Existenz dieses Teils im Rahmen der Erschließung.

### **Automatisierte Gewinnung von Erschließungsinformationen**

Angesichts der schnell ansteigenden Datenmenge wird in vielen Archivverwaltungen hinsichtlich einer (zukünftigen) Erschließung von digital vorliegenden Fotos und Bildern über die Einführung neuer Verfahren, wie z. B. die Bildmustererkennung, nachgedacht. Ziel ist die automatisierte Gewinnung möglichst gut verwendbarer Erschließungsinformationen und Metadaten, die nur ein geringes Maß an Nacharbeit erfordern. In einigen Fällen bestehen schon Kontakte zu einschlägigen Anbietern, Entwicklern und Forschern auf diesem Gebiet. Der Einsatz solcher Assistenzsysteme wird, wenn diese ausgereift sind, als sehr sinnvoll im Hinblick auf die dadurch zu gewinnenden Kapazitäten für die archivarische Facherschließung angesehen.

### **Metadatenqualität und -formate**

In den meisten der befragten Archivverwaltungen wird für die Verzeichnungsdaten ein eigenes Format (je nach AFIS) und für die Weitergabe an das Archivportal-D EAD (DDB) verwendet. Als Containerformat für die Weitergabe von Digitalisaten an das Archivportal-D wird METS/MOTS eingesetzt. Die Qualität der Metadaten wird unterschiedlich eingeschätzt.

## **2. Perspektiven**

Die folgenden Gedanken sind nicht mehr nur an den Fragen aus der Umfrage orientiert, sondern sie sind das Ergebnis der anschließenden, darauf aufbauenden Diskussion im KLA-Ausschuss Archivische Fachinformationssysteme.

### **Digitalisieren und Erschließung**

Nach dem Verhältnis zwischen Digitalisierung und ‚Tiefenerschließung‘ wurde nicht explizit gefragt; dennoch lohnt der Blick auf dieses Zusammenspiel. Erschließung ist die Basis für die Recherchierbarkeit und die Bereitstellung von Archivgut. Neben dem klassischen Weg der Erschließung mittels textueller Beschreibung bzw. Erfassung treten neue Formen, wie z. B. die optische Darstellung der Inhalte durch Digitalisate, die Referenzierung durch Linked Data (Normdaten usw.) oder die Volltextrecherchierbarkeit durch OCR. Die aktuellen Überlegungen zur Erschließung reichen von einer Basiserschließung bis hin zur ‚Tiefenerschließung‘ mit umfangreichen „Enthält“-Vermerken oder Einzelblattindizierung. Das vordringliche Ziel dieser Erschließung besteht darin, Nutzer\*innen in die Lage zu versetzen, das Archivgut überhaupt zu finden. Wenn Archivgut digital bereitgestellt wird, muss jedoch eine Vernetzung mit verschiedenen Portalen und über das Semantic Web ermöglicht werden. Dazu ist es nötig, eine gewisse Menge an Erschließungsinformationen vorzuhalten; so sollten z. B. Normdaten erfasst sein. Außerdem muss die Datenqualität Mindeststandards auch für Recherchestrategien (z. B. die Facettensuche) oder für normierte Datumsangaben erfüllen. Auch ausführlichere Textinformationen in den Erschließungsdaten können sinnvoll sein, um über automatische Texterkennungsverfahren Erschließungsdaten auszuwerten und vernetzen zu können.

Anhand der Digitalisate ist auch eine nachträgliche, tiefere Erschließung möglich. Auf der Basis der Digitalisate könnte folgendermaßen erschlossen werden: a) durch einen Dienstleister mit automatisierter Texterkennung, b) durch Bereitstellung im Internet bei gleichzeitiger Anwendung von Crowdsourcing-Methoden bzw. citizen sciences (ggf. mit Forschergruppen) und c) perspektivisch durch automatisierte Verfahren (Text-/Handschriftenerkennung).

### **Digitalisierung „on demand“**

Die Möglichkeit einer Digitalisierung „on demand“ sollte für alle Nutzer\*innen offenstehen. Digitalisate, die über die Digitalisierung „on demand“ entstanden sind, sollten nach Möglichkeit allen Nutzer\*innen zur Verfügung gestellt werden. Dabei kann das Problem entstehen, dass ein Nutzer bzw. eine Nutzerin die Kosten allein tragen muss und alle anderen die Digitalisate kostenlos nachnutzen können, also von der privat getätigten Investition profitieren. Das könnte gelöst werden, indem generell gezahlt, oder eben nicht gezahlt wird, wie z. B. im Fall des Schweizer Bundesarchivs, das generell kostenlos Digitalisate für die Benutzer\*innen erstellt. Ebenso könnte man Kontingente für die kostenfreie Digitalisierung einführen, entweder pro Nutzer\*in oder bezogen auf die Gesamtheit der herzustellenden Digitalisate. Schließlich könnte im Rahmen einer „Digitalisierungspatenschaft“ bei einer größeren Menge an Digitalisaten der Zahlende in den Metadaten (als Sponsor) namentlich genannt werden.

### **Persistent Identifier**

Die Forderung nach *eindeutigen* und *dauerhaften*<sup>3</sup> Identifikatoren (oder zumindest Permalinks) muss bekräftigt werden, da sich die verschiedenen Systeme ansonsten nicht vernetzen können. Das muss auch bei Migration von einem System in ein anderes berücksichtigt werden.<sup>4</sup>

### **Ebenen der Digitalisatverknüpfung**

Hinsichtlich der Erschließungstiefe sind die Vor- und auch Nachteile einer Digitalisatverknüpfung unterhalb der Verzeichnungsebene zu beachten. So kann man die Frage stellen, ob es sinnvoll ist, die einzelnen Blätter einer Akte separat zu verknüpfen. Es ist sicherlich vorteilhaft, die Möglichkeit einer Digitalisatverknüpfung unterhalb der Verzeichnung z. B. für einzelne Seiten einer Akte, eines Albums oder für einen Eintrag in einem Personenstandsregister in speziellen Fällen anzubieten. Dieses Verfahren generell anzuwenden, erscheint jedoch momentan noch zu ressourcenintensiv. Die Verwendung einer automatisierten Texterkennung, könnte eine Digitalisatverknüpfung unterhalb der Verzeichnungsebene ersetzen, da durch die Texterkennung einzelne Digitalisate gezielt suchbar gemacht werden könnten. Fraglich ist allerdings, ob auf diese Weise zukünftig eine automatisierte Vernetzung zwischen verschiedenen Metadaten systemen (z. B. Portalen) möglich sein wird.

---

<sup>3</sup> Dauerhaft bedeutet in diesem Zusammenhang: „über die Lebensdauer des Systems hinaus“.

<sup>4</sup> Persistent Identifier sind eindeutige, maschinenlesbare Benennungen von Ressourcen, z. B. von Archivalien oder Verzeichnungseinheiten. Sie bilden die Grundlage für die Verlinkung von Verzeichnungen oder Digitalisaten über die Lebensdauer eines Systems hinaus, indem ein unabhängiger Dienst im Internet die Adressen auflöst und die gesuchten Ressourcen aufrufbar macht, egal in welchem System diese zur jeweiligen Zeit vorgehalten werden. Vgl. [https://www.forschungsdaten.org/index.php/Persistent\\_Identifier](https://www.forschungsdaten.org/index.php/Persistent_Identifier) [letzter Zugriff: 19.08.2019].

## **Anwendung des Repräsentationenmodells**

Die Vorteile bei der Anwendung des Repräsentationenmodells, das einen Ansatz zur Beschreibung von Archivalien in ihren unterschiedlichen physischen Erscheinungsformen bietet, sind von Peter Sandner in einem früheren Empfehlungspapier des KLA-Ausschusses AFIS ausführlich dargelegt worden.<sup>5</sup> Das Repräsentationenmodell sorgt für die getrennte Abbildung von Original, Digitalisat, Mikrofilm usw., z. B. im Bereich der Bereitstellung. Dies ermöglicht, für einen virtuellen Lesesaal die verschiedenen Repräsentationen – und damit auch die Digitalisate von Archivgut – im System zu identifizieren und dadurch nutzbar zu machen. Das Repräsentationenmodell wird bereits von einigen Archivverwaltungen angewendet und dementsprechend auch in den dort verwendeten Archivinformationssystemen abgebildet.

## **Hybridakten**

Eine Zielsetzung sollte sein, dass Behörden bei der Abgabe bereits vermerken müssen, wenn Hybridakten vorliegen, damit das Archiv diesen Umstand entsprechend bei der Verzeichnung berücksichtigen kann. In den Verzeichnungssystemen müssen beide Teile der Akte berücksichtigt werden. Das kann entweder in einem Datensatz für beide Teile der Akte erfolgen, oder in zwei getrennten. Es wäre von Vorteil, wenn der analoge Teil der Akten auch digitalisiert wird, um eine leichtere Nutzung zu ermöglichen.

## **Perspektiven einer automatisierten, textbasierten Erschließung**

In der deutschen Archivlandschaft ist generell ein aufgeschlossener und zukunftsorientierter Umgang mit dem Thema der automatisierten und textbasierten Erschließung festzustellen. Archive sollten entsprechende Forschungen dazu unterstützen, indem sie z. B. geeignetes Testmaterial zur Verfügung stellen. Sie sollten engen Kontakt zu Vertreter\*innen der Digital Humanities sowie zu Anbieter\*innen und Einrichtungen (z. B. kriminologische Institute) halten, die entsprechende Software entwickeln. Grundsätzlich empfiehlt es sich, möglichst alle Erschließungsdaten, ob aus Transkriptions- oder Indexierungsprogrammen oder aus der klassischen Erschließung, in einem AFIS zu integrieren, um das Angebot an die Nutzer\*innen übersichtlich zu gestalten.

## **Vor- und Nachteile des METS/MODS – Formats**

METS/MODS ist ein kompliziertes Format, das bei der Anwendung zu Mehraufwänden führt; dennoch ist es bei Verwendung des DFG-Viewers verpflichtend. Es stellt sich die berechtigte Frage, ob dieses Format einen inhaltlichen Mehrwert hat, oder ob es die Datenbereitstellung aufgrund seiner Komplexität vielmehr behindert. Angesichts des verpflichtenden Einsatzes muss bei Digitalisierungsprojekten von den Archiven der Bedarf für die Erstellung der METS/MODS-Dateien und für den Support einkalkuliert werden.<sup>6</sup>

---

<sup>5</sup> Vgl. Peter Sandner: Das Repräsentationenmodell als Konzept für die archivische Arbeit. KLA verabschiedet Thesenpapier. In: *Archivar* (2016), S. 18-21.

<sup>6</sup> U. a. Fragen zum Datenformat und zur Datenqualität sollen auf dem Workshop zur Dateninfrastruktur in der Portallandschaft, der vom KLA-Ausschuss AFIS organisiert wird, behandelt werden.

## **Fazit**

Die Umfrage unter den staatlichen Archivverwaltungen in Deutschland hat ergeben, dass einige Meilensteine bei der Einbindung von Digitalisaten in Archivinformationssysteme bereits umgesetzt sind. Dabei gilt meist der Grundsatz, dass Digitalisate eingebunden werden müssen. Nicht zuletzt aus bestandserhalterischen Überlegungen sollte die Digitalisierung forciert werden, um dem Verfall von Archivgut (z. B. angesichts des zerfallsbedrohten Papiers ab der zweiten Hälfte des 19. Jahrhunderts) vorzubeugen. Im Prinzip muss die Digitalisierung schneller voranschreiten als der Verfall von Papier, Filmmaterial u. a. Eine verstärkte Digitalisierung ist erforderlich, damit die Archive in der digitalen Welt die ihnen zustehende Stellung behaupten können. Archivgut, das künftig nicht über das Internet bereitgestellt wird, wird weniger wahrgenommen, und noch weniger von Nutzer\*innen tatsächlich benutzt werden. Schließlich bietet die Digitalisierung neue Möglichkeiten, Wissen zu vernetzen und breiteren Kreisen über die Forschung hinaus zugänglich zu machen.

Für den Ausschuss: Dr. Kathrin Pilger und Dr. Bettina Joergens