

DFG-Praxisregeln „Digitalisierung“

AKTUALISIERTE FASSUNG 2022

Bearbeitet von

Reinhard Altenhöner Staatsbibliothek zu Berlin Preußischer Kulturbesitz

Andreas Berger Historisches Archiv der Stadt Köln

Christian Bracht Deutsches Dokumentationszentrum für Kunstgeschichte - Bildarchiv Foto Marburg

Paul Klimpel iRights.Law

Sebastian Meyer Sächsische Landesbibliothek - Staats- und Universitätsbibliothek Dresden

Andreas Neuburger Landesarchiv Baden-Württemberg

Thomas Stäcker Universitäts- und Landesbibliothek Darmstadt

Regine Stein Niedersächsische Staats- und Universitätsbibliothek Göttingen

Unter Mitwirkung von

Eva Bodenschatz Sächsische Landesbibliothek - Staats- und Universitätsbibliothek Dresden

Barbara Fichtl Niedersächsische Staats- und Universitätsbibliothek Göttingen

Oliver Götze Landesarchiv Baden-Württemberg

Melanie Gruß Institut für Theaterwissenschaft, Universität Leipzig

Angela Kailus Deutsches Dokumentationszentrum für Kunstgeschichte - Bildarchiv Foto Marburg

Celia Krause Deutsches Dokumentationszentrum für Kunstgeschichte - Bildarchiv Foto Marburg

Clemens Neudecker Staatsbibliothek zu Berlin Preußischer Kulturbesitz

Stefanie Rühle Niedersächsische Staats- und Universitätsbibliothek Göttingen

Thomas Scheidt Deutsches Dokumentationszentrum für Kunstgeschichte - Bildarchiv Foto Marburg

Zoe Schubert Staatsbibliothek zu Berlin Preußischer Kulturbesitz



Zitiervorschlag: Altenhöner, Reinhard, Berger, Andreas, Bracht, Christian, Klimpel, Paul, Meyer, Sebastian, Neuburger, Andreas, Stäcker, Thomas, & Stein, Regine. (2023). DFG-Praxisregeln „Digitalisierung“. Aktualisierte Fassung 2022. Zenodo. <https://doi.org/10.5281/zenodo.7435724>

Englische Fassung: <https://doi.org/10.5281/zenodo.7561148> (Übersetzung von Eva Bodenschatz)

Texte, Grafiken und Fotos stehen, soweit nicht anders gekennzeichnet, unter der Creative-Commons-Lizenz Namensnennung 4.0 (CC BY 4.0). Das bedeutet, dass sie vervielfältigt, verbreitet und öffentlich zugänglich gemacht werden dürfen, auch kommerziell, sofern dabei stets die Urheber, die Quelle des Textes und oben genannte Lizenz genannt wird, deren genaue Formulierung unter <https://creativecommons.org/licenses/by/4.0/legalcode> nachzulesen ist.

Inhaltsverzeichnis

| | |
|---|----|
| Vorbemerkung..... | 4 |
| 1. Prüfliste für Antragstellende und für die Begutachtung..... | 6 |
| 1.1 Allgemeine technische Abläufe/Voraussetzungen..... | 7 |
| 1.2 Technische Parameter der digitalen Reproduktion..... | 7 |
| 1.3 Metadaten..... | 8 |
| 1.4 Volltextgenerierung..... | 9 |
| 1.5 Langzeitverfügbarkeit..... | 10 |
| 1.6 Organisatorische Fragen, Zusammenarbeit mit Dienstleistern..... | 10 |
| 1.7 Zitieren, persistente Adressierung..... | 10 |
| 1.8 Bereitstellung der Digitalisate und Metadaten für die Öffentlichkeit..... | 11 |
| 2. Ziele und Auswahl..... | 12 |
| 2.1 Ziele..... | 12 |
| 2.2 Auswahl..... | 12 |
| 2.3 Publierte Texte: Dublettenprüfung und Datenabgleich..... | 13 |
| 3. Digitalisierung und Erschließung..... | 14 |
| 3.1 Bereitstellung der Materialien, konservatorische Prüfung..... | 14 |
| 3.2 Technische Parameter der digitalen Reproduktion..... | 15 |
| 3.2.1 Allgemeine Erläuterungen und Parameter (zweidimensionale Objekte)..... | 15 |
| 3.2.2 Materialspezifische Parameter (inkl. dreidimensionale Objekte)..... | 20 |
| 3.3 Metadaten..... | 26 |
| 3.3.1 Erschließung, deskriptive Metadaten..... | 27 |
| 3.3.2 Strukturelle Metadaten für digitale Faksimiles..... | 28 |
| 3.3.3 Sammlungs- und Bestandsbeschreibung..... | 29 |
| 3.3.4 Austausch und Weitergabe der Metadaten..... | 29 |
| 3.4 Volltextgenerierung..... | 30 |
| 3.4.1 Texterfassung..... | 31 |
| 3.4.2 Zeichenkodierung..... | 34 |
| 3.4.3 Markup von Volltexten..... | 34 |
| 3.4.4 Layout..... | 35 |
| 3.5 Langzeitverfügbarkeit..... | 35 |
| 4. Organisatorische Fragen – Eigendigitalisierung versus Dienstleistung..... | 36 |
| 5. Zitieren digitaler Ressourcen, persistente Adressierung..... | 37 |
| 6. Bereitstellung der Projektergebnisse für die Öffentlichkeit..... | 38 |
| 6.1 Rechte, Lizenzierung und Open Access..... | 38 |
| 6.1.1 Vorgehen vor der Veröffentlichung..... | 38 |
| 6.1.2 Umgang mit verwaisten/nicht verfügbaren Werken..... | 39 |
| 6.1.3 Zugang und Nachnutzung..... | 40 |

| | |
|--|----|
| 6.2 Mindestanforderungen für die Bereitstellungssysteme von Digitalisaten..... | 41 |
| 6.2.1 Funktionalitätsanforderungen..... | 41 |
| 6.2.2 Technische Mindestanforderungen | 42 |
| 7. Präsentationsstandards für Textwerke, DFG-Viewer..... | 42 |
| 7.1 Basisanforderungen und Architektur..... | 43 |
| 7.2 Funktionalitätsanforderungen..... | 43 |
| 7.3 DFG-Viewer und IIIF | 44 |
| Anhänge..... | 45 |
| Anhang A: METS/MODS-Profil für die Darstellung im DFG-Viewer und Übermittlung per OAI | 45 |
| Anhang B: METS/TEI-Spezifikation für die Darstellung von digitalisierten Handschriften | 53 |
| Anhang C: LIDO-Kernelemente für die Publikation | 58 |

Vorbemerkung

Die Deutsche Forschungsgemeinschaft (DFG)¹ fördert in ihrem Programm „Digitalisierung und Erschließung“ Projekte, mit denen das Angebot digitalisierter Objekte und Metadaten für die Forschung erweitert und verbessert wird. Ziel des Programms, welches sich seit der im Jahr 2020 erfolgten Öffnung auf alle wissenschaftlich relevanten Objektarten erstreckt, ist die Digitalisierung und/oder Erschließung von Beständen und Sammlungen, die für die Forschung überregional von Bedeutung sind. Zudem wird die Entwicklung von materialspezifischen Qualitätskriterien und Praxisregeln gefördert, wo diese noch nicht bestehen.²

Die Praxisregeln „Digitalisierung“ stellen eine zentrale Grundlage für DFG-geförderte Digitalisierungsprojekte dar: Sie sollen Antragstellenden die Planung von Digitalisierungsprojekten erleichtern und eine vergleichende Begutachtung von Anträgen unterstützen. Die Praxisregeln formulieren Standards und enthalten Informationen zu organisatorischen, methodischen und technischen Fragen im Kontext der Digitalisierung und Erschließung forschungsrelevanter Objekte. Sie leisten damit einen wichtigen Beitrag zur Nachhaltigkeit, Zugänglichkeit und Anschlussfähigkeit geförderter Projekte und der in diesem Zusammenhang entstehenden Infrastruktur.

Für den Bereich der handschriftlichen und gedruckten Überlieferung werden bereits seit 2007 Praxisregeln für die Digitalisierung bereitgestellt, welche in turnusmäßigen Abständen unter Koordination durch die DFG-Geschäftsstelle aktualisiert wurden. Seit 2013 sind darin auch Standards und Empfehlungen für die Digitalisierung von Archivgut und von bildbezogenen Objekten in musealen Sammlungen enthalten. Die letzte Überarbeitung erfolgte 2016. Mit der Öffnung des Förderprogramms, der Diversifizierung der zu digitalisierenden Materialien sowie der hohen Geschwindigkeit technischer Weiterentwicklungen ergibt sich die Notwendigkeit einer Revision und Ausdifferenzierung der bestehenden Praxisregeln. Diesen Bedarf hat die DFG erkannt und die Weiterentwicklung der Praxisregeln auf das breitere Fundament eines Selbstorganisationsprozesses gestellt, der den differenzierten Bedarf der jeweiligen Communitys besser abzubilden vermag.³ Zur Unterstützung stellt die DFG geeignete Förderinstrumente zur Verfügung.⁴

Das vorliegende Dokument stellt eine aktualisierte Fassung der zuletzt 2016 publizierten Praxisregeln dar. Es wurde in Absprache mit der DFG-Geschäftsstelle durch eine vom NFDI-Konsortium NFDI4Culture initiierte Autor*innen-gruppe erarbeitet, deren Mitglieder mehrheitlich seit langem an der Ausgestaltung der Praxisregeln beteiligt waren sowie aktiv in die NFDI-Konsortien NFDI4Culture, NFDI4Memory, NFDI4Objects und Text+⁵ eingebunden sind.

Die bisherige Gliederung des Dokuments wurde beibehalten, der Gesamttext wurde geprüft, wesentliche Textteile bleiben unverändert gültig. Aktualisierungen gibt es in einigen technischen Aspekten insbesondere im Bereich der 3D-Digitalisierung, in rechtlichen Aspekten vor dem Hintergrund der Urheberrechtsreform von 2021 sowie in den Anhängen zu den empfohlenen Metadaten-Anwendungsprofilen.

¹ Die Deutsche Forschungsgemeinschaft (DFG) ist die zentrale Selbstverwaltungseinrichtung der Wissenschaft zur Förderung der Forschung an Hochschulen und öffentlich finanzierten Forschungsinstitutionen in Deutschland. Die DFG dient der Wissenschaft in allen ihren Zweigen durch die finanzielle Unterstützung von Forschungsvorhaben und durch die Förderung der Zusammenarbeit unter den Forschenden (<http://www.dfg.de>). Die DFG unterstützt auch Vorhaben zum Aufbau und zur Weiterentwicklung leistungsfähiger Informationssysteme und -infrastrukturen für die Forschung. Die Ergebnisse der geförderten Projekte sollen für die Wissenschaft frei und dauerhaft zugänglich sein (<http://www.dfg.de/lis>).

Alle hier und im Folgenden angegebenen Weblinks wurden zuletzt am 15.12.2022 geprüft.

² Siehe https://www.dfg.de/foerderung/programme/infrastruktur/lis/lis_foerderangebote/digitalisierung_erschliessung/index.html.

³ Um diesen Prozess anzustoßen, fand am 26. April 2021 ein DFG-Rundgespräch zur Selbstorganisation der Praxisregeln Digitalisierung statt, siehe https://www.dfg.de/download/pdf/foerderung/programme/lis/rundgespraech_praxisregeln.pdf.

⁴ https://www.dfg.de/foerderung/programme/infrastruktur/lis/lis_foerderangebote/vigo/.

⁵ <https://nfdi4culture.de/>, <https://4memory.de/>, <https://www.nfdi4objects.net/>, <https://www.text-plus.org/>.

In Kapitel 1 werden die wichtigsten Anforderungen in Kürze zusammengefasst. In den Kapiteln 2 bis 6 findet sich eine ausführliche Einführung in die Probleme und Vorgehensweisen, die sich für alle Einrichtungen stellen, die Objekte in digitale Form überführen wollen. Diese Kapitel wenden sich auch an Einrichtungen, die Digitalisierungsprojekte neu planen, und adressieren ebenso Digitalisierungsmaßnahmen ohne Vorhandensein detaillierter Erschließungsdaten. Kapitel 7 verweist auf für Textobjekte erforderliche Präsentationsstandards und Formate.

Die jetzt überarbeiteten vorliegenden Praxisregeln „Digitalisierung“ dienen als Ausgangspunkt für eine material- und communitybezogene Ausdifferenzierung der Praxisregeln durch die Communitys. Die an diesem Dokument beteiligten Institutionen werden sich in der weiteren Entwicklung engagieren. Diese sind: Deutsches Dokumentationszentrum für Kunstgeschichte - Bildarchiv Foto Marburg, Historisches Archiv der Stadt Köln, Landesarchiv Baden-Württemberg, Niedersächsische Staats- und Universitätsbibliothek Göttingen, Sächsische Landesbibliothek - Staats- und Universitätsbibliothek Dresden, Staatsbibliothek zu Berlin Preußischer Kulturbesitz, Universitäts- und Landesbibliothek Darmstadt. Ganz im Sinne ihres gemeinsamen Memorandums of Understanding⁶ stehen die NFDI-Konsortien NFDI4Culture, NFDI4Memory, NFDI4Objects und Text+ als Kontaktpunkte⁷ und Plattform für die nächsten Schritte zur Verfügung, sind die Praxisregeln „Digitalisierung“ doch ein zentraler Baustein für die Standardbildung in ihren Forschungscommunitys. Alle mit der Digitalisierung forschungsrelevanter Objekte befassten Communitys und Einrichtungen sind dazu aufgerufen, mit ihrer Expertise am weiteren Prozess mitzuwirken.

⁶ <https://zenodo.org/record/4045000>

⁷ Kontaktadressen: contact@nfdi4culture.de, office@text-plus.org

1. Prüfliste für Antragstellende und für die Begutachtung

Wie in den folgenden Kapiteln ausgeführt wird, bestehen bei der Planung von Digitalisierungsprojekten zahlreiche Wahlmöglichkeiten. Eine rein mechanische Verpflichtung auf ganz bestimmte Standards würde daher eine unzulässige Einschränkung der zu fördernden Projekte bedeuten und deren dynamische Weiterentwicklung beschränken. Andererseits sollte klar werden, in welchen Bereichen Risiken für die Projektplanung und den Projektverlauf liegen. Die Prüfkriterien sind daher folgendermaßen zu verstehen und anzuwenden:

- Bei allen Anträgen mit Digitalisierungskomponenten werden unabhängig von ihrer inhaltlichen Begutachtung die vorgeschlagenen technischen Konzepte geprüft.
- Darüber hinaus wird geprüft, ob die notwendigen Rechtereklärungen bezogen auf das Ausgangsmaterial ebenso erfolgt und nachgewiesen sind, wie die Vorkehrungen, die eine vollumfängliche Nachnutzbarkeit der Projektergebnisse erlauben (→ 1.8).
- Antragstellende müssen in ihren Anträgen nachvollziehbar belegen, dass die Durchführung des Projekts die im Folgenden aufgelisteten Standards einhält.
- Jede Abweichung von diesen Standards muss begründet werden.
- Zur besseren Vergleichbarkeit sollen Anträge (bzw. Arbeitsberichte) Auskünfte über die geschätzten (bzw. tatsächlichen) Kosten für die technische Reproduktion inklusive Strukturdatenerfassung pro Digitalisat und Titel/Verzeichnungseinheit sowie Kosten für die Volltexterfassung enthalten.⁸
- Sieht ein Digitalisierungsprojekt vor, die im Folgenden angeführten Standards nicht einzuhalten, so ist die Notwendigkeit im Einzelnen zu begründen, soweit dadurch erhöhte Kosten entstehen.

⁸ Vgl. dazu das Datenblatt zu Neu- und Fortsetzungsanträgen im Programm „Digitalisierung und Erschließung“ https://www.dfg.de/formulare/12_152/12_152_de.xlsx. Zur Berechnung der realen Kosten gibt es zurzeit noch keine allgemein akzeptierten Standards. Die Praxisregeln empfehlen daher, die Kosten des Scannvorgangs inklusive Strukturdatenerfassung folgendermaßen zu berechnen:

(1) Bei der Vergabe an einen Dienstleister.

- a) Aus den an den Dienstleister gezahlten Kosten pro Digitalisat.
- b) Aus den anteiligen Kosten an alle vom Dienstleister verlangten Pauschalen (z.B.: Benennung und Speicherung der Dateien; Strukturdatenvergabe, Transfer der Daten vom Dienstleister zum Projektstandort).
- c) Aus den anteiligen Personalkosten aller Personen, die ausschließlich oder mit einem kalkulierbaren Anteil die Qualitätskontrolle durchgeführt haben.

(2) Bei der Digitalisierung im Hause.

- a) Aus den anteiligen Kosten an der neu angeschafften Digitalisierungshardware im engeren Sinne. Diese gilt mit Projektende als abgeschrieben. (Ganze Laufzeit: im Falle eines vier Jahre laufenden Projekts, für das zunächst zwei Jahre beantragt werden, ist also die für die volle vierjährige Laufzeit veranschlagte Zahl von Digitalisaten zu Grunde zu legen.)
- b) Aus den anteiligen Personalkosten aller Personen, die die Digitalisierungshardware bedienen und Strukturdaten vergeben.
- c) Aus den anteiligen Personalkosten aller Personen, die ausschließlich oder mit einem kalkulierbaren Anteil die Qualitätskontrolle durchgeführt haben.

In beiden Fällen (1) und (2) sind die Kosten pro Digitalisat zu errechnen. Keine Kosten des Scannens sind alle Kosten, die entstehen:

- a) Durch das Projektmanagement (z.B.: Auswählen und Ausheben der zu digitalisierenden Stücke).
- b) Durch die Erfassung von über die Strukturdaten hinausgehenden Metadaten.
- c) Durch die Langzeitarchivierung.
- d) Indirekte Kosten, wie sie in der Regel im Rahmen betriebswirtschaftlicher Kostenrechnungen erhoben werden.

- Grundsätzlich müssen Anträge technisch soweit vorbereitet sein, dass eine Gesamtbeurteilung der technischen Anforderungen und Abläufe aus dem Antrag heraus möglich ist. Es besteht also – abgesehen von der Erprobung innovativer Technik – nicht die Möglichkeit, eine erste Projektphase vorzusehen, in der beispielsweise der zeitliche Aufwand für eine Digitalisierungskampagne, die zu wählende Qualität der Digitalisate oder der allgemeine Arbeitsablauf (Workflow) erst bestimmt wird. Dazu nötige Pilotstudien sind vor der Antragstellung abzuschließen.

1.1 Allgemeine technische Abläufe/Voraussetzungen

Im Antrag ist der geplante Workflow so detailliert zu beschreiben, dass eine Bewertung der folgenden Fragen durch die Gutachterinnen und Gutachter möglich wird:

- Ist die vorgesehene Personalausstattung hinreichend und notwendig? Dazu sind Aussagen über die durchschnittlichen Ressourcen (Arbeitszeit, Speicherkapazität der am Workflow beteiligten Rechner) für jede der dem Mengengerüst zu Grunde liegenden Einheit zu machen.
- Sind die vorgesehenen Bearbeitungszeiten realistisch? Soweit die vorgesehenen Bearbeitungszeiten nicht unmittelbar plausibel sind, sind sie entweder durch die Erfahrungen in vergleichbaren Projekten, durch veröffentlichte und anerkannte Richtwerte oder durch die Ergebnisse selbst vorgenommener Vorabtests zu belegen.

1.2 Technische Parameter der digitalen Reproduktion

Ziel der Digitalisierung ist die möglichst originalgetreue Wiedergabe des Objekts nach Maßgabe der wissenschaftlichen Erfordernisse. Die Richtlinien für die Mindestanforderungen an die Digitalisierung beziehen sich ausschließlich auf den digitalen Master (→ [3.2.1](#)) und stellen übergreifende Empfehlungen insbesondere für bildhaftes und textuelles Material dar. Zu materialspezifischen Empfehlungen, die über diese Mindestanforderungen hinausgehen, vergleiche (→ [3.2.2](#)). Vergleichbare Qualitätsstandards für Digitalisate anderer Materialgruppen sind in Vorbereitung.

- Die Auflösung ist so zu wählen, dass bei den Archivkopien die kleinsten relevanten Details klar zu sehen sind, wenn die Datei auf ein Viertel der Ausgangsgröße verkleinert wird. Grundsätzlich wird eine Auflösung von mindestens 300 ppi empfohlen (→ [3.2.1.1](#)).
- Für die Sicherung des digitalen Masters ist eine Farbtiefe von 16 Bit im RGB-Farbmodus ausreichend (→ [3.2.1.2](#)).
- Die Speicherung der Master erfolgt in Form unkomprimierter Baseline-TIFF-Dateien. Die weitergehenden Optionen von extended TIFFs sollten für den digitalen Master nicht genutzt werden. Neben TIFF kann auch TIFF-LZW oder JPEG2000 in seiner verlustfreien Form als Format für den Bildmaster verwendet werden. Generell sind jedoch Formate ohne Komprimierung denen mit verlustfreier Komprimierung vorzuziehen. Für die Speicherung von Mastern im JPEG2000-Format ist darauf zu achten, dass nur die lizenzfreien Bereiche Verwendung finden. Zu den mit JPEG2000 und TIFF-LZW einhergehenden Risiken bezüglich der Langzeitarchivierung vergleiche Kapitel → [3.2.1.4](#).
- Im Sinne des angestrebten offenen und kostenfreien Zugangs zu jeglicher Form von wissenschaftlicher Information (Open Access) wird erwartet, dass digitale Reproduktionen so frei nachnutzbar wie rechtlich möglich angeboten werden.

1.3 Metadaten

Metadaten sind grundsätzlich in einer von der Software unabhängigen und standardkonformen Form bereitzustellen, in aller Regel in einer XML-Kodierung.

- Für eine optimale verteilte, auch nachhaltige Nutzbarkeit der deskriptiven Metadaten ist die Erschließung an den einschlägigen Spartenstandards und Referenzmodellen zu orientieren und wo immer möglich mit publizierten Normdaten zu verknüpfen. Für die Erfassung personenbezogener, biografischer und geografischer Information wird die Verwendung der Gemeinsamen Normdatei (GND)⁹ der Deutschen Nationalbibliothek erwartet. Darüber hinaus eingesetzte kontrollierte Vokabulare, z.B. Iconclass zur Bilderschließung, müssen national und international anschlussfähig sein (→ [3.3.1](#)).
- Die Bereitstellung der deskriptiven Metadaten zur weiteren Nutzung gemäß den materialspezifischen Standards ist verpflichtend. Für eine Minimalerschließung kann das DFG-Basisdatenset¹⁰ genutzt werden; für weitergehende Erfassung wird die Anlage eines Kernfeldkatalogs empfohlen (→ [3.3.1](#)).

Für Metadaten zur inhaltlichen Beschreibung des Objekts in Portalen in einem der folgenden Formate:

- Text Encoding Initiative (TEI¹¹) für textuelle Quellen
- Lightweight Information Describing Objects (LIDO¹²) für bildhafte und dreidimensionale Objekte
- Encoded Archival Description (EAD¹³), gängiges Anwendungsprofil: EAD(DDB)¹⁴

Für Metadaten zur Darstellung im DFG-Viewer (s.u.) in einem der folgenden Formate:

- Metadata Encoding and Transmission Standard / Metadata Object Description Schema (METS¹⁵/MODS) für gedruckte Textwerke und Archivgut
- Metadata Encoding and Transmission Standard / Text Encoding Initiative (METS/TEI) für Handschriften

Die bereitgestellten Metadaten müssen gegen das jeweilige XML-Schema valide sein und sind darüber hinaus auf semantische Korrektheit zu überprüfen.

- Digitalisate und Metadaten müssen durch persistente Identifier im Nachweissystem (Katalog, Online-Findmittel) verbunden werden. Von antragstellenden Personen wird erwartet, dass sie die Frage der Metadatenerfassung und Digitalisierung mit einer Infrastruktureinrichtung abstimmen bzw. durch diese durchführen lassen. Der Nachweis in einschlägigen Portalen und Verbundsystemen wird erwartet. Alle Materialien müssen in geeigneten fachspezifischen oder fachübergreifenden Online-Anwendungen präsentiert werden (→ [3.3](#)).
- Die Ablieferung an Portale sollte gemäß den Standardformaten möglichst per Schnittstelle über OAI-PMH¹⁶ erfolgen. Alle Einrichtungen und Zielportale, die selbst keine OAI-Schnittstelle anbieten, können auch herkömmliche http-URIs als Direktlink zu den METS-Dateien angeben. Deutsche Digitale Bibliothek (DDB¹⁷) und Archivportal-D bieten noch keine öffentlich zugängliche OAI-Schnittstelle für Datenlieferungen an (→ [3.3.4](#)).

⁹ <https://gnd.network/>.

¹⁰ https://www.dfg.de/formulare/12_155/.

¹¹ <https://tei-c.org/>.

¹² <https://cidoc.mini.icom.museum/working-groups/lido/lido-overview/>.

¹³ <https://www.loc.gov/ead/index.html>.

¹⁴ <https://wiki.deutsche-digitale-bibliothek.de/x/hBliAQ>.

¹⁵ <http://www.loc.gov/standards/mets>.

¹⁶ <http://www.openarchives.org/>.

¹⁷ <https://www.deutsche-digitale-bibliothek.de/>.

- Es obliegt dem Projektnehmer sicherzustellen, dass die im Projekt hergestellten Digitalisate (Bild und Volltext) eindeutig identifiziert und von anderen im selben System vorhandenen Einheiten getrennt durchsucht und aufgerufen werden können (→ [3.3](#)).
- Die Entscheidung über die Erstellung struktureller Metadaten betrifft nur Textdokumente. Falls strukturelle Metadaten Verwendung finden, wird die Konsultation der über die Website des DFG-Viewers¹⁸ zugänglichen Liste von Bezeichnungen erwartet. Empfohlen wird nach dem derzeitigen Stand eine Orientierung an METS oder TEI, wobei der DFG-Viewer bei textuellen Materialien derzeit noch das Mittel der Wahl ist, es sei denn, es gibt inhaltliche Gründe dagegen. Für Bildmaterial besteht die Möglichkeit einer Präsentationslösung basierend auf dem International Image Interoperability Framework (IIIF¹⁹) (→ [3.3.2](#)).
- Es wird erwartet, dass Digitalisierungsprojekte mindestens Gegenstand und Umfang der jeweiligen Objektauswahl auf einer Seite im Netz, möglichst auch in Englisch, darstellen. Wünschenswert sind standardisierte Sammlungsbeschreibungen, typischerweise in demselben Format wie die Objektbeschreibungen: MODS oder TEI-Header, EAD(DDB), LIDO (→ [3.3.3](#)).²⁰

1.4 Volltextgenerierung

Von allen Anträgen zur Digitalisierung textueller Objekte wird eine Auseinandersetzung mit Möglichkeiten der Volltextbereitstellung erwartet. Für Druckwerke ab Erscheinungsjahr 1850 gilt verpflichtend, dass Volltext hergestellt werden muss und eine bloße Bilddigitalisierung nicht ausreicht (→ [3.4](#)). Volltexte umfassen die Zeichen der Vorlage, gegebenenfalls in den Text eingetragene Auszeichnungen bzw. Markup-Daten zur Markierung von Strukturmerkmalen sowie Metadaten, die üblicherweise Teil der gleichen Datei sind (→ [3.4](#)).

- Die Herstellung von Volltext kann auf zwei Wegen erfolgen, entweder durch Optical Character Recognition/Handwritten Text Recognition (OCR/HTR) oder durch Abschreiben bzw. Transkription (→ [3.4.1](#)).
- Um eine Einheitlichkeit bei der Beurteilung der Textgenauigkeit zugrunde legen zu können, werden Antragstellende gebeten, diesbezügliche Angaben zur Buchstabengenauigkeit zu machen, d.h. mangelhafte Worttrennungen und Layoutfehler unberücksichtigt zu lassen. Stichproben sollten sich an dem in Kapitel 3.4.1 beschriebenen statistischen Verfahren orientieren (→ [3.4.1](#)).
- Bezüglich der Zeichenkodierung wird empfohlen, die Texte in Unicode abzuspeichern. Zu favorisieren ist UTF-8 (→ [3.4.2](#)).
- Wenn nicht triftige Gründe dagegensprechen, müssen Volltexte kodiert bzw. mit Markup versehen werden. Für deskriptives Markup und strukturierten Volltext wird das Modell der Text Encoding Initiative (TEI) als wesentlich empfohlen; für Layoutinformationen und Koordinaten haben sich neben TEI auch ALTO XML (ALTO V4.x²¹) und PAGE XML²² bewährt (→ [3.4.3](#)).
- In manchen Fällen ist es bei der Präsentation eines Volltextes wichtig, das Layout eines Dokuments langfristig zu sichern. Die Praxisregeln empfehlen bevorzugt für diese Fälle den Einsatz einer geeigneten Formatierungssprache (z.B. XSLT, XSL-FO, XQuery, CSS), die die Unabhängigkeit von spezieller Software

¹⁸ <http://dfg-viewer.de>. In Zukunft soll in den Praxisregeln keine Festlegung auf bestimmte Viewertechnologien mehr erfolgen.

¹⁹ <https://iiif.io/>.

²⁰ Zunehmend relevant ist außerdem das RDF-basierte Data Catalog Vocabulary (DCAT): <https://www.w3.org/TR/vocab-dcat-3/>. Vgl. auch Data Catalog Vocabulary (DCAT) - Version 2. W3C Recommendation 04 February 2020: <https://www.w3.org/TR/vocab-dcat-2/>.

²¹ <https://www.loc.gov/standards/alto/>. Seit 2018: <https://github.com/altxml/documentation/wiki/Versions>.

²² <https://ocr-d.de/de/gt-guidelines/trans/trPage>.

weitgehend sicherstellt. Falls eine Archivierung des Formats mit XML-Techniken aus nachvollziehbaren Gründen nicht möglich ist, können Layoutinformationen zu Textdokumenten auch in PDF nach der ISO-Norm 19005-1 archiviert werden. PDF-Dateien sind aber, wie in [Kapitel 3.4.3](#) dargelegt, kein Ersatz für eine Bereitstellung der mit Markup versehenen Volltexte in XML (→ [3.4.4](#)).

- Im Sinne des Open Access und der Open Source wird erwartet, dass maschinenlesbare Volltexte bzw. das den Volltexten zu Grunde liegende XML sowie XSLT-Skripte und DTDs oder XML-Schema-Dateien so frei nachnutzbar wie rechtlich möglich angeboten werden (→ [1.8](#); [6.1](#)).

1.5 Langzeitverfügbarkeit

Für die Langzeitsicherung und Archivierung digitaler Inhalte gibt es zum jetzigen Zeitpunkt keine verallgemeinerbare Lösung, die für alle Objektgattungen geeignet ist. Im Rahmen der Langzeitsicherung werden Dateien in stabilen, migrationsfähigen Formaten auf einem technisch wie organisatorisch sicheren Speichersystem gesichert.

Die Archivierung digitaler Daten setzt auf einem solchen Speicher auf, beinhaltet jedoch weitreichendere technische und organisatorische Festlegungen, die nicht nur eine physische Erhaltung der Daten, sondern Strategien zur Bereitstellung für die Nutzung (Access) einbeziehen. Als Referenz dient das OAIS-Modell²³ (Offenes Archiv-Informationssystem) (→ [3.5](#)).

Die langfristige Verfügbarkeit der Ergebnisse von Digitalisierungsprojekten ergibt sich einerseits aus der Wahl der Daten- und Metadatenformate. Andererseits ist sicherzustellen, dass die digitalen Daten auch physikalisch verfügbar bleiben. Dabei gilt: Kosten für die projektbezogene Sicherung der Daten werden in DFG-geförderten Digitalisierungsprojekten als Eigenleistung für die Laufzeit des Projekts anerkannt. Eine Förderung dieser Kosten aus DFG-Mitteln kann nicht erfolgen (→ [3.5](#)).

Es sei darauf hingewiesen, dass Digitalisierungsprojekte aus Sicht der DFG stets Projekte der gesamten Einrichtung sind: Die Unterstützung der das Projekt durchführenden Fachabteilung durch die IT-Infrastruktur des Hauses wird vorausgesetzt. Dabei wird begrüßt, wenn sich kleinere Einrichtungen der Kompetenz und der Dienstleistung größerer Einrichtungen bedienen (→ [3.5](#)).

Anträge müssen nachvollziehbare Aussagen zur institutionellen Langzeitsicherung und Archivierung enthalten. Digitalisate sind in redundanter Form zu archivieren (→ [3.5](#)).

1.6 Organisatorische Fragen, Zusammenarbeit mit Dienstleistern

Die Entscheidung darüber, ob ein Vorhaben als Inhouse-Projekt konzipiert oder ob auf das Angebot eines Dienstleisters zurückgegriffen wird, ist immer eine projektspezifische, die allein in die Verantwortung der Antragstellerinnen und Antragsteller fällt. In Verträgen müssen Leistungen exakt festgelegt werden. Die DFG erwartet, dass ein angemessener Prozentsatz des Rechnungsbetrags als Sicherheit zurückbehalten wird und erst nach einer durchgeführten Qualitätskontrolle zur Auszahlung an das Dienstleistungsunternehmen kommt. (→ [4](#))

1.7 Zitieren, persistente Adressierung

Digitalisate müssen, um von anderen Objekten oder Datenbanken aus erreichbar zu sein, eindeutig angesprochen werden können. Dazu ist über die übliche Zitierform hinaus, die durch ein Angebot in der Navigationssoftware als

²³ Siehe dazu: <http://d-nb.info/104761314X/34>.

klassische Form weiter genutzt werden kann und sollte, die Festlegung und die online zugängliche Dokumentation von Adressierungstechniken erforderlich.

Sichergestellt werden muss die Erreichbarkeit und Zitierbarkeit einer Ressource als Ganzer und ihrer einzelnen physischen Teile. Einrichtungen sollten durch geeignete Mechanismen (Persistent Uniform Resource Locator (PURL), Uniform Resource Name (URN), Digital Object Identifier (DOI), Handle, etc.) die Persistenz der Ressource und der Verknüpfung zu ihr gewährleisten, um zuverlässiges Arbeiten mit den bereitgestellten Quellen in wissenschaftlichen Kontexten zu ermöglichen. Für gedruckte Textdokumente wird die Erzeugung von URNs über die Deutsche Nationalbibliothek empfohlen (→ [5](#)).

1.8 Bereitstellung der Digitalisate und Metadaten für die Öffentlichkeit

Die Digitalisierung wissenschaftlicher Materialien wird durch die DFG gefördert, um diese Materialien der Forschung in Deutschland und weltweit zugänglich zu machen. Die folgenden Aspekte sind dabei zu beachten:

Vor Projektbeginn muss eine Rechteklärung bezüglich der zu digitalisierenden Objekte erfolgt sein. Insbesondere eventuell vorliegende Urheber-, Persönlichkeits-, Leistungsschutz- und Archivrechte müssen dabei berücksichtigt werden (→ [2.2](#) und [6.1](#)).²⁴

Metadaten, Digitalisate, Volltexte und XML-Dateien sollen so frei nachnutzbar wie rechtlich möglich angeboten werden. Sie sind entweder als gemeinfrei (Public Domain) zu markieren oder mit der freiest möglichen Creative Commons-Lizenz²⁵ zu versehen. Bei geschütztem Material sind standardisierte Rechteinweise zu verwenden (→ [6.1](#)).

Digitale Master sollten als hochaufgelöste Derivate im TIFF-Format oder vollaufgelöste JPEGs zur Verfügung gestellt werden. Für AV-Material sollte der Mindeststandard für Präsentationsderivate MP3 (Audio) bzw. MP4 (Video) sein, für 3D-Medien glTF/GLB. Die Daten sollten ohne technische Zugangsbeschränkungen (Login, IP-Schranken etc.) verfügbar sein (→ [6.1](#)).

Bei Projekten, in denen nicht nur gemeinfreie Materialien digitalisiert werden, kann eine verzögerte Open Access-Publikation (moving wall) von bis zu einem Jahr nach Projektende vereinbart werden (→ [6.1](#)).

Eine Einschränkung des Open Access kann rechtlich begründet sein, darf aber maximal 10% der zu digitalisierenden Materialien betreffen und setzt ein Konzept zur zeitlichen und technischen Verfügbarmachung voraus.²⁶

Alle Objekte sind mit langfristig zitierfähigen URLs möglichst niedriger Granularität zu versehen. Die Form, in der zitiert werden soll, ist deutlich anzuzeigen.

Die DFG erwartet, dass DFG-geförderte Projekte auf ihren im Internet angebotenen Daten eindeutige Herkunftsnachweise sowie gegebenenfalls den Hinweis auf die Förderung durch die DFG anbringen. (→ [6.1](#)). Eine Zugänglichmachung soll über unterschiedliche Wege vorgenommen werden (→ [6.2.1](#))

- über die Website der das Projekt durchführenden Einrichtung
- über den lokalen oder regionalen Katalog
- über überregionale Nachweis- und Präsentationssysteme, insbesondere die DDB beziehungsweise das Archivportal-D
- über materialspezifische Portale (sofern vorhanden)

²⁴ Eine Rechteklärung ist nicht immer im Vorfeld eines Projekts möglich. In begründeten Ausnahmefällen ist eine Rechteklärung im Rahmen des Projekts (ggf. sogar mit Förderung) erlaubt.

²⁵ Vgl. die Lizenzen in der zurzeit gültigen Version: <https://creativecommons.org/>.

²⁶ Eine Förderung der Bearbeitung nicht im Open Access zugänglichen Materials mit DFG-Mitteln ist nicht möglich.

- über Schnittstellen
- Alle jeweils gängigen Browser müssen unterstützt werden, soweit dies der Natur der Sache nach nicht unmöglich ist (→ [6.2.1](#) und [7.2](#)).

2. Ziele und Auswahl

2.1 Ziele

Digitalisierung gilt heute als ein zentrales Instrument für die geistes- und kulturwissenschaftliche Forschung und die sich entwickelnden Digital Humanities. Die direkte Forschung mit den Quellen wurde und wird durch Digitalisierungsmaßnahmen erheblich erleichtert – bei gleichzeitiger Schonung der kostbaren, bisweilen fragilen Originale. Mit der Digitalisierung von Beständen aus Bibliotheken, Archiven und Museen werden aber nicht nur leicht verfügbare Kopien ins Netz gestellt, sondern es entsteht auch eine Infrastruktur, die das Internet zu einem integralen Forschungsraum für eine zunehmend digital ausgerichtete geistes- und kulturwissenschaftliche Forschung macht. Erst die Verbindung mit anderen online verfügbaren Ressourcen ermöglicht es, das Potential des Netzes vollständig auszunutzen. Ziel der Digitalisierung ist also nicht nur das Bereitstellen und Nutzbarmachen, sondern auch und vor allem das Vernetzen der unterschiedlichen Ressourcen zu einer virtuellen Forschungsinfrastruktur.

Es gibt heute abgesicherte Kenntnisse zur Durchführung von Digitalisierungsprojekten, die freilich nicht mechanisch angewendet werden können: Was für die Digitalisierung von Handschriften des Mittelalters unverzichtbare konservatorische Sorgfalt ist, kann für die massenweise Bearbeitung von Verwaltungsschriftgut des späten 19. Jahrhunderts unnötig zeitraubend und überteuert sein.

Da sich die Mehrzahl der DFG-geförderten Projekte bislang mit Flachware (2D) befasst, werden im Folgenden nach wie vor schwerpunktmäßig die Techniken und Parameter für diesen Bereich erläutert. Neben textuellen Materialien gewinnen bildhafte Materialien beträchtlich an Bedeutung, daher integriert die vorliegende Ausgabe der Praxisregeln weiterhin besonders bei der technischen Erzeugung Entwicklungen in diesem Bereich, ohne jedoch den Anspruch auf Vollständigkeit zu erheben. Bei den Formaten und bei den Digitalisierungsverfahren geht die Richtlinie punktuell auch auf AV-Medien oder dreidimensionale Sammlungsobjekte ein, allerdings noch ohne die notwendige Tiefe.

2.2 Auswahl

Grundsätzlich ist zu beachten, dass die technischen Aspekte der Digitalisierung gut geplant werden können, während der intellektuelle Aufwand bei der Auswahl geeigneter Objekte schwer kalkulierbar ist. Es wird daher immer zu entscheiden sein, ob eine großzügigere Definition des Mengengerüsts für ein Digitalisierungsprojekt insgesamt nicht kostengünstiger und effektiver ist als ein komplexer Bewertungs- und Auswahlprozess. Angeregt wird, auf bereits erbrachte Selektionsleistungen und Vorarbeiten wie Bibliografien oder Fachdatenbanken zurückzugreifen. Als grundsätzliche Kriterien für die Auswahl gelten Forschungsrelevanz und wissenschaftliche Nachfrage.

Die Bildung eines Korpus unter den Gesichtspunkten der Forschungsrelevanz oder der wissenschaftlichen Nachfrage ist nicht immer einfach. Hilfreich für die Plausibilisierung eines Vorhabens ist in diesen Fällen die Kooperation mit einer konkretisierbaren wissenschaftlichen Community oder Institution, die einen nachvollziehbaren Eigenbedarf formulieren kann. Ideal sind Kooperationen, in denen ein wissenschaftliches Vorhaben, z.B. ein philologisches

oder rechtshistorisches Forschungs- oder Editionsprojekt, seinerseits ein Internetangebot anstrebt und auf Bibliotheks- oder Archivbestände rekurrieren will, sodass die wechselseitige Vernetzung möglich wird.²⁷

Interessant sind sicher auch Konzepte des digitization on demand oder digitization on use (Digitalisierung von in Bibliotheken und Archiven genutztem oder stark genutztem Material), wodurch in jedem Fall ein konkretes Nutzungsinteresse sichergestellt ist.

Grundsätzlich sollte bei spezialisierten, eher an einer konkreten Forschungsfrage orientierten Digitalisierungsprojekten gelten, dass sie sich in größere, eher nach formalen Gesichtspunkten vorgehende Projekte integrieren lassen. Umgekehrt sollten fachübergreifende Projekte ihre Angebote für eine vertiefende und spezialisierte Nachnutzung öffnen.

Die Digitalisierung wissenschaftsrelevanter Materialien wird durch die DFG gefördert, um diese Materialien der Forschung in Deutschland und weltweit zugänglich zu machen. Bereits mit Beginn der Projektplanung, spätestens aber zum Zeitpunkt der Antragstellung, muss die Rechtklärung bezüglich der zu digitalisierenden Materialien erfolgt sein und im Antrag nachgewiesen werden. Insbesondere eventuell vorliegende Urheber-, Persönlichkeits- sowie Leistungsschutz- und Archivrechte sind dabei zu berücksichtigen. Die Rechtklärung gestaltet sich häufig aufwendiger als zunächst erwartet. Für die entsprechenden Aufwände ist daher ausreichend Zeit und personeller Einsatz einzuplanen. Stellt sich die Rechtklärung als schwierig dar, wird empfohlen, professionelle Unterstützung, beispielsweise durch das Justitiariat der bestandshaltenden Einrichtung zu suchen. Auch ist schon bei der Auswahl zu beachten, dass die Ergebnisse der durch die DFG geförderten Digitalisierungsprojekte der Forschung im Sinne des Open Access so frei wie möglich zur Verfügung gestellt werden müssen. Wenn eine Markierung als Public Domain rechtlich nicht möglich ist, ist eine freie Creative Commons-Lizenz zu wählen (CC0, CC-BY oder CC-BY-SA) (→ [6.1](#)). Nur in begründeten Ausnahmefällen ist auch eine Online-Stellung nach den gesetzlichen Erlaubnissen (Schranken) für verwaiste (§§ 61 UrhG ff) oder nicht verfügbare Werke (§§ 61d UrhG ff) zulässig, da diese die Anforderung an eine Open Access-Publikation nicht erfüllen.

2.3 Publierte Texte: Dublettenprüfung und Datenabgleich

Zur Vermeidung von Doppeldigitalisierungen gedruckter textueller Quellen ist es sinnvoll, vor Antragstellung zu prüfen, ob die zur Digitalisierung vorgesehenen Werke in Deutschland oder im Ausland bereits digital vorliegen. Hierfür gelten folgende Anforderungen:

In Anträgen und Berichten werden Aussagen zu bereits vorliegenden oder laufenden nationalen und internationalen Digitalisierungsvorhaben erwartet, soweit sich diese Vorhaben auf das beantragte oder laufende DFG-Projekt und die dort bearbeiteten Bestände beziehen.

Projekte mit größeren Mengengerüsten (z.B. über 1.000 Drucke) sollten in Anträgen und Berichten erläutern, wie sich ihr Vorhaben zu kostenfrei zugänglichen kommerziellen Digitalisierungsangeboten verhält. Ziel sollte es sein, die Zahl der gegebenenfalls auftretenden Doppeldigitalisierungen mit pragmatischem Aufwand möglichst gering zu halten.

Bei Drucken der Erscheinungsjahre von 1501 bis 1800 sind das VD16, VD17 und VD18 als Referenzinstrumente für Prüfungen auf Doppeldigitalisierungen heranzuziehen. URN und PURL der Digitalisate sind an diese Verzeichnisse zu melden. Bei der Digitalisierung von Inkunabeln ist der deutsche Inkunabelzensus (ISTC) heranzuziehen.

²⁷ In der Regel bauen Forschungsprojekte auf geförderten Infrastrukturen auf. Es ist jedoch ebenfalls möglich, Projekte, die Forschung und Infrastrukturelemente kombinieren, zu beantragen. Alternativ kann man auf einschlägige Fachbibliografien zurückgreifen, die einen Kanon formulieren, oder aber auch eine Mischung aus beidem anstreben.

3. Digitalisierung und Erschließung

Die Retrodigitalisierung umfasst im 2D-Bereich immer die Herstellung digitaler Derivate (vom Master → [3.2.1](#)) und die Erzeugung von Metadaten, im Falle von Textwerken gegebenenfalls zusätzlich auch die Volltexterfassung und die Erzeugung von Strukturdaten und Markup. Mit digitalen Images sind auch genuin digital erstellte (born digital) Reproduktionen materiell vorliegender Objekte gemeint (z.B. digitale Fotografien von Bauwerken). Wenn im Folgenden von Digitalisierung gesprochen wird, so ist der gesamte Arbeitsgang gemeint (Vorbereitung, Digitalisierung im engeren Sinne, Erzeugung von bibliografischen Metadaten, Strukturdaten, Volltexten sowie Langzeitsicherung/digitale Bestandserhaltung). Auf eine detaillierte Differenzierung nach Materialarten (u.a. Textwerke, bildmediale Werke wie z.B. Malerei, Zeichnung, Druckgrafik, Film- und Tonträger wie z.B. Video, DVD, CD, Schallplatte, Musikkassette sowie Gebrauchsgegenstände, ortsfeste Objekte/Bauwerke etc.) wird im Folgenden noch verzichtet.

3.1 Bereitstellung der Materialien, konservatorische Prüfung

Vorbereitende Tätigkeiten werden bei Digitalisierungsprojekten oft unterschätzt und sollten vor einem Projekt genauestens geprüft werden. Sind die Objekte überhaupt verfügbar? Gibt es möglicherweise konservatorische Bedenken gegenüber der Digitalisierung der Originale? Stehen genügend Mitarbeiterinnen und Mitarbeiter zur Aushebung und Bereitstellung der Objekte zur Verfügung? Ist wissenschaftlich bzw. bibliografisch geschultes Personal zur Hand, das Vollständigkeitskontrollen bzw. Kollationierungen vornimmt, sofern dies aus vorhandenen Katalogeinträgen nicht hervorgeht? Bei gedruckten Textwerken sollte eine Digitalisierung von unvollständigen oder defekten Exemplaren nach Möglichkeit vermieden und eine Reproduktion eines vollständigen Exemplars angestrebt werden. Für andere Materialgattungen gilt dies gleichermaßen.

Die konservatorische Prüfung kann sehr viel Zeit in Anspruch nehmen, sollte aber zum Schutz der Objekte keinesfalls unterbleiben. Es wird empfohlen „Checklisten“ für die Prüfung der Digitalisierungsfähigkeit zu verwenden und daraus die Vorgaben für die Digitalisierung zu entwickeln.²⁸ Stellt die Reproduktion mit den vorhandenen Reproduktionstechniken ein Risiko bzw. eine besondere Belastung für die Originale dar²⁹, so sollte sie – sofern möglich – von einem bestehenden Mikrofilm³⁰ erfolgen oder ganz unterbleiben. In jedem Fall sind wertvolle historische Materialien mit der konservatorisch nötigen Sorgfalt zu behandeln, auch wenn dies den Durchsatz bei der Digitalisierung reduziert und einen erhöhten Zeitaufwand erfordert. Die Scantechnik und der Einsatz von Hilfsmitteln, wie Buchkeile oder Fixierhilfen, sind entsprechend den Vorgaben der konservatorischen Prüfung zu wählen.

²⁸ Kriterien einer konservatorischen Prüfung sollten sein: Druckfarbe/Tinte/Tusche gefährdet; Farbauftrag/Grundierung gefährdet; Tintenfraß/Farbfraß; Heftung lose; Bünde oder Überzugmaterial zu steif/unflexibel; Bünde (an-)gebrochen; Überzug im Gelenk (an-)gebrochen; Rückeneinlage zu steif; Einbandrücken beschädigt; Brüche, Risse oder Fehlstellen; Ledernarben im Rücken empfindlich; Rückenvergoldung gefährdet; sehr dicker Pergamentband mit hohlem Rücken; Buchdeckel (an-)gebrochen; Buchdeckel gelöst (vorne/hinten); Schließenriemen zu steif/angebrochen; mikrobieller Befall; Mikroform vorhanden. Darüber hinaus bestehen mitunter auch technische Probleme, die eine Digitalisierung erschweren oder verhindern: Bundsteg zu schmal; Buchblock wellig; Blätter/Lagen sehr steif; Textverlust unvermeidbar; Planlage der Vorlage unmöglich; extremes Format.

²⁹ Corbach, Almuth: Bestandsschonendes Digitalisieren von schriftlichem Kulturgut. In: Digital und analog. Die beiden Archiwelten. 46. Rheinischer Archivtag. Ratingen 21.-22. Juni 2012. Beiträge. Hrsg.: LVR-Archivberatungs- und Fortbildungszentrum. Bonn 2013 (Archivhefte 43), S. 90-102;

³⁰ In Archiven ist die Digitalisierung vom Mikrofilm unabhängig von konservatorischen Erwägungen durchaus Praxis. Zu den materialspezifischen Parametern für eine Digitalisierung vom Mikrofilm vgl. (→ [3.2.2.4](#)).

3.2 Technische Parameter der digitalen Reproduktion

Ziel der Digitalisierung ist die möglichst originalgetreue Wiedergabe des Objektes nach Maßgabe der wissenschaftlichen Erfordernisse. Die anzuwendenden Parameter für die Digitalisierung sind mit Blick auf die Qualität der gewünschten Reproduktion, seiner Langzeitverfügbarkeit und Interoperabilität zu wählen.

3.2.1 Allgemeine Erläuterungen und Parameter (zweidimensionale Objekte)

Bei der Herstellung von Digitalisaten sind grundsätzlich zwei Formen zu berücksichtigen. Zum einen der sogenannte digitale Master, also das Ausgangs- oder Archivformat, und für die Nutzung hergestellte Derivate. Der Master ist die Grundlage für alle weiteren Prozesse. Daher sollte man seiner Herstellung besondere Aufmerksamkeit schenken und bei der Langzeitsicherung entsprechende Empfehlungen berücksichtigen. Die folgenden Richtlinien für die Mindestanforderungen an die Digitalisierung beziehen sich ausschließlich auf den digitalen Master.

3.2.1.1 Auflösung

Das gebräuchliche Maß für das Auflösungsvermögen einer Datei ist ppi (pixel per inch)³¹, gleich Bildpunkte pro inch. ppi ist als Verhältnismaß immer bezogen auf ein Seitenmaß. Das totale Maß einer Bilddatei ist das Pixelmaß.

Als Untergrenze sollte daher eine Scanauflösung für die Digitalisate gewählt werden, bei der die Details einer Vorlage vollständig in einer gleichgroßen Reproduktion wiedergegeben werden können.³²

Daher gilt die grundsätzliche Empfehlung von mindestens 300 ppi bezogen auf das Vorlagenformat als Zielauflösung (Vorlagenformat = Ausgabeformat des Digitalisats bei 300 ppi). 300 ppi beziehen sich allerdings nur auf Vorlagen, die für die Betrachtung mit bloßem Auge gedacht sind, wie etwa Textwerke, Grafiken und fotografische Aufsichtsvorlagen. Anders verhält es sich bei Medien, deren vollständige Bildinformation nur in Vergrößerung erfassbar wird. Das sind Miniaturen jeglicher Art und ganz besonders fotografische Durchsichtmedien (z.B. Negative oder Dias).³³

Weiterhin gibt es Objekte, insbesondere Werke der Architektur, aber auch überformatige Malereien, die sich auf Grund ihrer Dimensionen den oben beschriebenen Verfahren zur Bestimmung der Auflösung entziehen. So wird zum Beispiel mit einer 50-Megapixel Digitalkamera eine Auflösung von 300 ppi bei DIN A2 Ausgabegröße erreicht.³⁴

³¹ Bei digitalen Bilddaten hat sich im Unterschied zum Druck inzwischen ppi (pixel-per-inch) durchgesetzt.

³² Ein Ansatz zur Ermittlung der notwendigen Mindestauflösung eines Digitalisats ist das Auflösungsvermögen des menschlichen Auges bei deutlicher Sehweite. Die kürzeste Sehweite, bei der man einen Gegenstand ermüdungsfrei über einen längeren Zeitraum betrachten kann, liegt bei ca. 25 cm. Das menschliche Auge ist in der Lage, zwei Linien noch zu unterscheiden, wenn das Licht auf mindestens zwei nicht benachbarte Sehzellen trifft und mindestens eine Sehzelle dazwischenliegt. Eine Rasterfrequenz von 60 Linien/cm bei 20 cm Betrachtungsabstand kann nicht mehr als getrennte Linien wahrgenommen werden. Daraus resultiert folgende Auflösungsanforderung, die auch in der Druckindustrie gebräuchlich ist:

60 Linien/cm (60-er Raster) benötigen mindestens 120 Pixel zur Darstellung.

120 px/cm x 2,54 cm/inch = 304,8 ppi, rund 300 ppi

³³ Beispiel: Kleinbildnegative mit der Größe von 24 x 36 mm sind für die anschließende Vergrößerung auf Fotopapier hergestellt worden. Sie eignen sich nicht für die Betrachtung mit bloßem Auge. Als fotografische Praxisregel gilt, dass sich Kleinbildnegative (abhängig von Objektiv, Belichtung, Film und Entwickler) ca. 10-fach vergrößern lassen, also auf einen Fotoabzug der Größe von 24 x 36 cm. Um diese Zielgröße von 24 x 36 cm nach obiger Maßgabe im Digitalisat des KB-Negativs zu erreichen, müsste es mit mindestens 2835 x 4252 Pixel digitalisiert werden.

³⁴ Bei einem Flächensensor mit z.B. 5.700 x 8.600 Pixeln ergibt sich also nach der Beziehung Kamerapixel/Zielauflösung in ppi = Zielgröße in inch folgende Ausgabegröße bei 300 ppi:

8600/300 = 28,7 inch x 2,54 cm/inch = 72,81 cm

5700/300 = 19 inch x 2,54 cm/inch = 48,26 cm

Hinweise für eine gegebenenfalls sinnvolle Abweichung von der grundsätzlichen Auflösungsempfehlung von 300 ppi bezogen auf das Vorlagenformat werden bei den materialspezifischen Parametern genannt (→ [3.2.2](#)).

Die erreichbare Auflösung eines digitalen Aufnahmesystems ist dabei nicht nur von der Anzahl der zur Verfügung stehenden Pixel einer Scanzeile oder eines Aufnahmesensors abhängig. Sie ist ein Zusammenspiel aus unterschiedlichen Faktoren. So spielen etwa die technische Qualität des digitalen Aufzeichnungsgeräts (Kamera, Scanner) und die Abbildungsleistung der verwendeten Objektive die entscheidende Rolle.

3.2.1.2 Farbtiefe

Die Farbtiefe bestimmt die Differenzierung der Helligkeits- und Farbwerte in einem Digitalisat. Da in der digitalen Technik nur diskrete Zustände (ja/nein) möglich sind, können Helligkeits- und Farbunterschiede – im Gegensatz zur analogen Fotografie – nicht kontinuierlich, d.h. mit fließenden Übergängen dargestellt werden.

Der Vorteil der höheren Farbtiefe von 16 Bit pro Kanal liegt also in der größeren Farbdifferenzierung.³⁵ Dadurch gehen bei einer späteren Bildbearbeitung weniger Tonwerte verloren. Dieser Bearbeitungsspielraum ist wichtig bei bearbeitungsintensiven Digitalisaten, z.B. von SW-Negativen.

Digitale Aufnahmesysteme zeichnen grundsätzlich mit einer hohen Farbtiefe von 14 Bit bzw. 16 Bit pro Kanal auf.

3.2.1.3 Digitaler Aufnahmeablauf

Aufnahmetechnik

In der fotografischen digitalen Aufnahmetechnik unterscheidet man zwischen der Zeilenscantechnik, bei der eine trilineare Scanzeile mit fester Breite (je eine Zeile ist für eine Farbe – Rot, Grün, Blau – empfindlich) eine bestimmte Strecke abtastet. Flachbettscanner und Zeilenscanrückteile für Fach- oder Mittelformatkameras arbeiten mit dieser Technik. Hierbei wird für jeden Bildpunkt die Farbinformation physikalisch erzeugt.

Das zweite gebräuchliche Verfahren ist der Flächensensor, bei dem auf einer definierten Fläche eine bestimmte Anzahl von Fotodioden angebracht ist.

Zeilenscansysteme und multishotfähige Flächensensorkameras bieten Vorteile bei filigranen Motiven wie bildhaften Tiefdruckwerken, Kartenwerken oder Textilien.

Mit gebräuchlichen Flachbettscannern lassen sich zweidimensionale Medien bis zu einer Vorlagengröße von DIN A3 digitalisieren. Das Verfahren ist allerdings nicht berührungsfrei, da die Medien mit dem Vorlagenglas des Scanners in Kontakt kommen, sie eignen sich daher nur für unbedenkliche Vorlagen in Bezug auf die Bestandhaltung.

Bildrauschen

Um die Details einer Vorlage möglichst vollständig zu reproduzieren, sollten bei der eingesetzten Kameratechnik die herstellerspezifischen Empfehlungen zur Verminderung von Bildrauschen beachtet werden. Grundsätzlich gilt dabei die Faustregel: Je geringer die verwendete ISO-Einstellung ist, desto geringer ist das Rauschen.³⁶

³⁵ Aus technischer Sicht ist hier ggf. zu beachten, dass gängige Programme für die Bildverarbeitung (bspw. OpenCV, Pillow, ImageMagick) keine vollständige Unterstützung für 48-Bit Farbe bieten und damit entsprechende Konvertierungen (potentielle Fehlerquelle) für die Verarbeitung notwendig werden.

³⁶ ISO 19264-12021, Seite 19.

Objektive

Weiterhin empfiehlt sich die Verwendung moderner, hochwertiger Objektive, die für das hohe Auflösungspotential digitaler Kameras entwickelt worden sind. Festbrennweiten sind Zoomobjektiven vorzuziehen, da bei ihnen Abbildungsfehler gezielt korrigiert werden und der „Überkopf-Einsatz“ auf Dauer die Mechanik beschädigt. Zoomobjektive bieten immer nur einen Qualitätskompromiss über den vorhandenen Brennweitenbereich. Shiftoptiken vermindern Verluste in der Bildqualität, die durch nachträgliches Entzerren stürzender Perspektivlinien entstehen. Ist der verwendete Abbildungsmaßstab unter 1:10, sollten spezielle Makroobjektive verwendet werden, die für diesen Abbildungsbereich konstruiert wurden. Abhängig vom verwendeten Kamerasystem können für Reproduktionen in diesem Maßstabsbereich auch hochwertige Vergrößerungsobjektive eingesetzt werden.³⁷

Arbeitsplatz

Zur Vermeidung von Unschärfen durch Verwacklungen während des Digitalisierens sollte darauf geachtet werden, dass bei der Wahl des Standorts für eine Digitalisierungseinheit ein erschütterungsfreier Platz gesucht wird. Holzfußböden z.B. sind ungeeignet, da sie schwingen und somit Vibrationen leicht übertragen. Werden Spiegelreflexkameras benutzt, sollten sie mit Spiegelvorauslösung ausgelöst werden, um Vibrationen des Spiegelschlags auszuschließen. Stative und Reprosäulen sollten so dimensioniert sein, dass sie auch das Gewicht der eingesetzten Kameras problemlos halten können.

Moiré

Bei Objekten mit feinsten, gleichmäßig verteilten Details kann das Oneshot-Verfahren bei Digitalkameras mit Flächensensor zu Farbverfälschungen und Moiré-Effekten führen.³⁸

Ein Moiré-Effekt tritt in der digitalen Bilderzeugung auf, wenn in Abhängigkeit von der Pixelgröße der Scanzeile oder des Flächensensors, des Abbildungsmaßstabes und der Linienfrequenz der Vorlage die gleichmäßige Pixelmatrix eines digital erzeugten Bildes mit einer regelmäßigen Linienstruktur der Vorlage interferiert. Das Auftreten dieses Effekts ist schwer vorhersehbar, da die entsprechenden Parameter selten konstant bleiben.

Farbverfälschungen treten bei Oneshot-Flächensensoren unter den gleichen Bedingungen auf, sind aber ein Interpolationsfehler der Kamerasoftware. Bei Oneshot-Kamerasystemen ist ein Pixel jeweils nur für eine der Grundfarben Rot, Grün, Blau empfindlich. Die restlichen Farbinformationen für den jeweiligen Bildpunkt werden aus den Farbinformationen der umliegenden Pixel interpoliert. Sind in Pixelabstand die Farb- und Helligkeitsunterschiede der Vorlage zu extrem, kommt es zu Farbverfälschungen. In solchen Fällen können Zeilenscanner oder multishotfähige Digitalrückteile bessere Ergebnisse liefern, da hier die Farbinformation physikalisch erzeugt wird und nicht rechnerisch. Im Gegenzug erhöht sich jedoch der Digitalisierungsaufwand etwas, da die Scanzzeiten länger sind und eine absolut konstante Beleuchtung während des Aufnahmevorgangs unabdingbar ist.

³⁷ Die Abbildungsleistung eines Objektivs lässt sich auch an den dazugehörigen Modulationstransferfunktions-Diagrammen (MTF-Diagrammen) ablesen. Diese Diagramme werden von diversen Objektivherstellern veröffentlicht. Die Modulationstransferfunktion beschreibt, wie gut ein Objektiv den Kantenkontrast in einer Vorlage auf das fotografische Abbild übertragen kann. Sie bietet somit einen Hinweis auf das Auflösungsvermögen einer Optik. Die technischen Datenblätter der Objektive geben auch Auskunft über den Helligkeits- und Schärfefall zum Bildrand. Optische Abbildungsfehler machen sich hier besonders bemerkbar. Zur Reduzierung der Bildfehler sollte ein Objektiv zweimal abgeblendet werden, um hauptsächlich die bessere Abbildungsleistung der Objektivmitte zu nutzen. Wenn keine große Tiefenschärfe benötigt wird, wie bei Reproduktionen planer Vorlagen, sollte mit dieser optimalen Arbeitsblende fotografiert werden.

³⁸ Das hier behandelte Moiré ist nicht mit dem Phänomen zu verwechseln, das aus der Interferenz von Druckpunkten gerasterter Vorlagen mit den Bildpunkten des Monitors entsteht.

Zur Vermeidung des Moiré-Effekts sollte das notwendige Verfahren und die notwendige Auflösung im Vorfeld eines Digitalisierungsvorhabens an Hand einer Testdigitalisierung ermittelt werden.

Beispielhaft zu nennende Objekte, bei denen Moiré auftreten kann, sind: gedruckte Halbtonvorlagen, Kupferstiche, Textilien.

Beleuchtungstechnik

Die Wahl der richtigen Beleuchtungstechnik ist im Vorfeld eines Digitalisierungsvorhabens unter bestandserhaltenden Gesichtspunkten zu prüfen. Zeilenscansysteme benötigen flickerfreies Dauerlicht, da hier über einen bestimmten Zeitraum die Vorlage kontinuierlich abgetastet wird. Lichtleistung und -farbe müssen während des Scanvorgangs konstant sein, um gleichmäßige Ergebnisse zu erzielen. Flächensensoren können auch mit Blitzlicht benutzt werden. Bei Multishotssystemen ist darauf zu achten, dass die Blitzleistung und -farbe während der einzelnen Belichtungen konstant bleiben. Gleiches gilt für Dauerlicht.

Unter konservatorischen Gesichtspunkten lässt sich die Meinung vertreten, dass eine kurzfristige hohe Lichtbelastung weniger schädlich ist als eine hohe Wärmebelastung, zumal bezüglich der Schädigung des Lichts die Lichtbilanz zählt, d.h. es ist unerheblich, ob die gleiche Lichtmenge über einen längeren Zeitraum verteilt oder kurzfristig auf das Objekt trifft. Kurzfristige hohe Helligkeitsunterschiede sind jedoch aus arbeitsmedizinischer Sicht für die Mitarbeitenden problematisch.

Bildverarbeitung

Fotografische Aufnahmen werden grundsätzlich im herstellerbedingten Rohdatenformat (RAW-Format) in der maximalen Größe erzeugt, mit einer Farbtiefe von 14 oder 16 Bit pro Farbkanal. RAW-Bilder repräsentieren die originalen Kameradaten. Bildkorrekturen sollten möglichst in RAW-Software vorgenommen werden. Da diese Datei nicht verändert wird, kann sie bei Bedarf verlustfrei neu korrigiert werden. Bearbeitungsvorschriften können überdies als Preset auf beliebig viele Bilder angewendet werden, wodurch eine hohe Arbeitseffizienz erreicht wird.

Das RAW-Format ist als proprietäres Dateiformat noch nicht archivierfähig. Dennoch sollte es dringend gesichert werden, da es immer die ursprüngliche Bildinformation behält, anders als das bearbeitete Master-TIFF. Nach aktuellem Stand empfiehlt sich eine Sicherung im Containerformat DNG.

Ob ein digitales Aufnahmesystem für ein Digitalisierungsvorhaben geeignet ist, lässt sich vorab im Rahmen einer Testdigitalisierung mit genormten Testcharts (zum Beispiel ISO-12233-Testchart, USAF-1951-Testchart) ermitteln. Die effektive Auflösung der Testbilder lässt sich dann mit einer Analysesoftware oder mit entsprechenden, in der Regel beigefügten Auflösungstabellen ermitteln. Es empfiehlt sich, beim Aufnahmegerät mit der Hinterlegung von Farbprofilen (Eingabeprofile, Ausgabeprofile) zu arbeiten.

Als Ausgabeprofil für Farbaufnahmen sind entweder Adobe RGB, der von der European Color Initiative (ECI)³⁹ empfohlene Arbeitsfarbraum ECI-RGB v2 oder der dazu identische Farbraum L-Star RGB⁴⁰ zu wählen. Für eine nachvollziehbare Farbproduktion ist es notwendig, eine geeignete Farbtafel bzw. einen Farbkeil mit aufzunehmen. Bei gleichbleibenden Reproduktionsbedingungen kann die Farbreferenzaufnahme einer Serie repräsentativ vorgelegt werden.

Bei der Helligkeits- und Tonwertsteuerung ist darauf zu achten, dass das Tonwerthistogramm (für Schwarz und Weiß) seitlich nicht angeschnitten wird, damit der Tonwertumfang erhalten bleibt. Als Grenzwerte für die Wiedergabe dunkler und heller Tonwerte empfehlen sich nach dem Lab-Farbmodell:

³⁹ European Color Initiative: <http://www.eci.org>.

⁴⁰ <http://www.colormangement.org/de/workingspaces.html>.

Schwarz: L*5

Weiß: L*95

Das Nachschärfen sollte für den digitalen Master nur moderat erfolgen. Anwendungsgebunden kann für unterschiedliche Derivate eine zusätzliche Schärfung erfolgen.

Für einen verlässlichen Abgleich originaler Vorlagen mit der Monitordarstellung des Digitalisats sollten die Vorlagen unter D50-Normlichtumgebung für grafische Arbeitsplätze nach ISO 3664:2000 betrachtet werden (so zumindest unsere Idealempfehlung). Monitormodelle, die in der grafischen Industrie Verwendung finden, liefern in diesem Zusammenhang die verbindlichsten Darstellungen.

Digitale Bildnachbearbeitung

Zur Optimierung der Bildqualität digitaler Aufnahmen kann eine zusätzliche Nachbearbeitung notwendig sein. Dabei sollte sich diese im Master auf notwendige Farb- und Tonwertkorrekturen beschränken. Auf jeden Fall zu vermeiden sind Objektdeformationen, das Hinzufügen oder Löschen von Objektteilen, sowie Spezialeffekte, wie z.B. der Einsatz von Verfremdungsfiltren. Zur Steigerung der Bildintegrität können in Derivaten Positionierungshilfen nachträglich aus dem Bild entfernt oder Hintergründe bereinigt werden.

3.2.1.4 Dateiformate

Bildmaster sollten nach dem derzeitigen Kenntnisstand in „TIFF uncompressed“⁴¹ langzeitgesichert werden. Das Format TIFF gibt es schon seit den 1980er Jahren. Es hat sich als einer der wichtigsten De-Facto-Standards etabliert und es ist damit zu rechnen, dass es auch in Zukunft von allen Standardprogrammen unterstützt wird. Allerdings gilt dies nur für sogenannte „Baseline-TIFFs“. Die weitgehenden Optionen von Extended TIFFs sollten für den digitalen Master nicht genutzt werden.

Neben TIFF kann auch TIFF-LZW oder JPEG2000⁴² in seiner verlustfreien Form als Format für den Bildmaster verwendet werden. Für die Speicherung von Mastern im JPEG2000-Format ist allerdings darauf zu achten, dass nur die lizenzfreien Bereiche von JPEG2000 Verwendung finden.

In den letzten Jahren gelangten gerade angesichts der zunehmenden Speicherproblematik der ISO-Standard JPEG2000⁴³ und nach Ablauf der Lizenzen TIFF-LZW als effiziente Kompressionsformate ins Blickfeld der Gedächtnisinstitutionen. Mit Blick auf die Langzeitarchivierung sollten die Betreibenden von Repositorien die Vor- und Nachteile des Einsatzes beider Formate (TIFF-LZW und JPEG2000) sorgfältig abwägen. Kompressionsformate sind grundsätzlich anfälliger für Bildverluste und über ihren Einsatz sollte nach einer Nutzen-Risiko-Abwägung entschieden werden. Bei der Auswahl eines Formates ist auch seine Verbreitung und Marktdurchdringung zu berücksichtigen. Unabhängig davon, dass international große und einflussreiche Bibliotheken wie die Library of Congress und die British Library auf JPEG2000 setzen, ist seine Verbreitung bisher geringer als die des unkomprimierten TIFF-Formates.

Die Lizenzierungssituation ist noch nicht abschließend geklärt, einige Bereiche von JPEG2000 wurden jedoch als frei deklariert.

Auch die unterschiedlichen proprietären RAW-Formate sind nicht zur Sicherung der Bildaufnahme als Master geeignet, zumal sie oftmals nur mit der dazugehörigen RAW-Software angezeigt werden können. Ebenso hat sich das

⁴¹ <https://www.nationalarchives.gov.uk/PRONOM/Format/proFormatSearch.aspx?status=detailReport&id=686;>
<http://www.nationalarchives.gov.uk/PRONOM/Format/proFormatSearch.aspx?status=detailReport&id=1099>

⁴² <http://www.jpeg.org/jpeg2000/>.

⁴³ ISO/IEC 15444-6:2013.

plattformübergreifende, kameraunabhängige Adobe-RAW-Format DNG nicht durchgesetzt und ist somit als Archivformat ungeeignet.

Für die Publikation im Internet empfehlen sich wegen ihrer großen Verbreitung grundsätzlich JPEG und PNG. Jedoch können auch andere Formate oder Techniken zum Einsatz gelangen, wenn diese für die Präsentation z.B. von sehr großen Bilddateien praktikabler sind. Hierbei ist darauf zu achten, dass nur Formate und Techniken eingesetzt werden, die ohne spezielle Technik oder Software genutzt werden können. Für die Präsentation im Internet ist insbesondere wichtig, dass das gewählte Format nativ von allen gängigen Browsern dargestellt werden kann.

Für AV-Medien muss ebenso zwischen Archivformaten und Nutzungsformaten unterschieden werden. Im Bereich Audio hat sich das Waveform Audio File-Format (WAV) in Verbindung mit der Pulse Code Modulation (PCM)⁴⁴ als Archivierungsformat etabliert. Als Nutzungsformat hat sich MP3 (.mp3, MPEG-2 Audiolayer III) durchgesetzt. Für Videoformate kann zurzeit noch keine eindeutige Empfehlung ausgesprochen werden. Jedoch sollte unbedingt darauf geachtet werden, nur lizenzfreie und offen beschriebene Formate zu verwenden. Gerade bei Videodateien liegt das Problem in der enormen Größe, so ist die Anforderung einer unkomprimierten Speicherung bei größeren Projekten kaum umsetzbar.⁴⁵

Digitale 3D-Dokumentation von digitalisierten bzw. digital rekonstruierten 3D-Objekten werden als 3D-Modelle abgelegt. Diese Modelle ermöglichen es, die Form, die Textur und die optischen Materialeigenschaften des Ursprungsobjekts originalgetreu abzubilden, bzw. nicht mehr vorhandene (u.a. zerstörte) Objekte zu dokumentieren und zu visualisieren.

Infrage kommende Datenformate sollten möglichst robust gegenüber Beschädigungen auf den Datenträgern sein, Speicherplatz möglichst effizient nutzen, das Modell logisch strukturiert nachbilden, eine schnelle Datenverarbeitung ermöglichen und entsprechend verbreitet sein.

Als offenes Dateiformat im Bereich der 3D-Modelle hat sich OBJ (.obj)⁴⁶ bewährt. Das Format wird von vielen 3D-Grafikprogrammen unterstützt und ist daher geeignet für die programm- und plattformübergreifende Weitergabe von 3D-Modellen. Optische Materialeigenschaften (z.B. Spiegelung, Transparenz, Glanzlicht usw.) werden in einer separaten Materialdatei (.mtl) definiert, die auch Angaben zu Texturierungen enthalten kann. Daneben findet vor allem im wissenschaftlichen Kontext glTF⁴⁷ Verwendung (auch als Webformat). Für die web-basierte Dokumentation und Vermittlung bietet sich das speziell zur Visualisierung von 3D-Modellen innerhalb der WebGL-Technologie entwickelte Format X3D (.x3d)⁴⁸ an. Allerdings spielt es wegen seines geringeren Leistungsumfangs eine nachgeordnete Rolle.

3.2.2 Materialspezifische Parameter (inkl. dreidimensionale Objekte)

3.2.2.1 Textwerke

Zu den Textwerken zählen in diesem Zusammenhang sowohl gedruckte Werke als auch unikale Dokumente wie Handschriften und Archivgut.

⁴⁴ Mindestqualität: 44,1 khz Abtastrate und 16 bit Abtasttiefe.

⁴⁵ Folgende Formate werden für die Langzeitarchivierung empfohlen, wobei derzeit noch keine befriedigende Lösung vorliegt: JPEG2000/MPEG-4 (ISO/IEC 14496-12:2015, ISO/IEC 15444-12:2015), DPX (Standardisiert als SMPTE 268M-2003, v 2.0), MXF/AAF (Standardisiert als SMPTE 377M und als ISO-Standard vorgeschlagen). Die FADGI arbeitet zurzeit an einer Richtlinie zur Archivierung von Filmen: http://www.digitizationguidelines.gov/guidelines/MXF_app_spec.html.

⁴⁶ <http://www.fileformat.info/format/wavefrontobj/egff.htm> und https://de.wikipedia.org/wiki/Wavefront_OBJ.

⁴⁷ <https://www.khronos.org/glTF/>.

⁴⁸ <http://www.web3d.org/x3d/what-x3d> und <https://de.wikipedia.org/wiki/X3D>.

Bei der Altbestandsdigitalisierung ist auf jeden Fall die Imagedigitalisierung vorzunehmen. Auch bei Vorliegen eines maschinenlesbaren Volltextes sollte auf die Imagedigitalisierung bzw. auf die Präsentation des digitalen Faksimiles nicht verzichtet werden, da eine Fülle von Informationen nur über das visuelle Abbild transportiert werden.

Die Lesbarkeit der Texte ist ausschlaggebend für die Wahl der Auflösung und von der im Text verwendeten Schriftgröße abhängig. Die benötigte Scanauflösung richtet sich in diesem Fall weniger nach den Maßen der Vorlage als nach der Lesbarkeit der einzelnen Lettern. Für Textdokumente mit dem kleinsten signifikanten Zeichen ab 1,0 mm ist eine Auflösung von 400 ppi und erst ab 1,5 mm Mindestzeichengröße eine Auflösung von 300 ppi zu empfehlen.⁴⁹

Die Blätter/Seiten der Textwerke werden immer vollständig mit leichtem, umlaufendem Rand digitalisiert, um deutlich zu machen, dass nichts von der Vorlage abgeschnitten wurde.

3.2.2.2 Grafische Darstellungen

Wie unter (→ [3.2.1.1](#)) beschrieben liegt die empfohlene Mindestauflösung für die Digitalisierung bei 300 ppi unter Berücksichtigung der Ergebnisse einer Testdigitalisierung mit genormter Testtafel.

Häufig wird jedoch bei kleinen Vorlagen eine Auflösung von 300 ppi bezogen auf das Vorlagenformat nicht ausreichen, um die Eigenschaften der Vorlage umfassend erkennbar wiederzugeben. Ist z.B. die künstlerische Technik der Ausführung nicht identifizierbar, sollte unbedingt eine höhere Auflösung gewählt werden. Beispielhaft seien hier erwähnt: Kupferstiche, Briefmarken, Medaillonportraits, Miniaturmalerei.

Die erreichbare Maximalauflösung ist dabei von der eingesetzten Digitalisierungstechnik abhängig. Hierzu ein Rechenbeispiel ohne Berücksichtigung der Objektivqualität: Eine Briefmarke ließe sich mit einer Digitalkamera mit einem Sensor im Kleinbildformat (24 x 36 mm) im Maßstab 1:1 reproduzieren.

Umgekehrt kann bei überformatigen Vorlagen (DIN A0 und größer) die Auflösung bezogen auf das Vorlagenformat reduziert werden, wenn die Objekte für einen großen Betrachtungsabstand entworfen wurden. Plakate sind hier als Beispiel zu nennen. In diesem Fall kann die Auflösung bis auf 150 ppi reduziert werden.

Sind Überformate jedoch detailreich wie beispielsweise Generalkarten oder Kupferstiche, so sollte mit mindestens 300 ppi digitalisiert werden. Alle bildwichtigen Details müssen erkennbar und lesbar sein.

Aufgrund des Mehrwerts für die Benutzung sollte in einem ökonomisch vertretbaren Rahmen die volle physikalische Auflösung der Aufnahmetechnik bei den grafischen Darstellungen genutzt werden.

Die Grafiken werden immer vollständig mit leichtem, umlaufendem Rand gesichert, um deutlich zu machen, dass nichts von der Vorlage abgeschnitten wurde.

3.2.2.3 Fotografien

Bei Fotografien ist zunächst zwischen Durchsichtmedien, also fotografischen Negativen oder Diapositiven, und Aufsichtmedien, also Fotopositiven z.B. auf Papier, zu unterscheiden. In der Regel handelt es sich bei Letzteren um Kontaktkopien oder Vergrößerungen von Negativen auf Fotopapier.

Durchsichtmedien

Das fotografische Negativ oder Diapositiv ist das Ergebnis einer fotografischen Aufnahme und damit die originale Quelle. Negative sind die Vorlagen für die Erstellung von Reproduktionen zur Verbreitung der Aufnahme als Fotopositiv oder Druck.

⁴⁹ Vgl. Federal Agencies Digitization Guidelines Initiative (FADGI): Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files, August 2010: http://www.digitizationguidelines.gov/guidelines/FADGI_Still_Image-Tech_Guidelines_2010-08-24.pdf, S. 59. <https://www.digitizationguidelines.gov/>.

Die negative Darstellung ist für die Betrachtung des Motivs ungeeignet. Um das Bildmotiv zu nutzen, muss ein positives Derivat aus dem digitalen Master erzeugt werden. Dieses dient als digitaler Fotoabzug vom Negativ und gibt den Bildinhalt wieder. Die positive Darstellung eines Masters ist aber im Regelfall zu flau und bei Farbnegativen zusätzlich durch deren Farbmaske nicht farbrichtig. Um das Motiv im digitalen Fotoabzug klar erkennbar wiederzugeben, ist daher eine beträchtliche Bildkorrektur nötig.

Gut ausgearbeitete Negative geben Bilder mit einem Dichteumfang (= Tonwertumfang) von bis zu 12 Blendenstufen wieder, die so auch von der Digitalisierungsvorrichtung reproduziert werden müssen.

Anders als die grafischen Darstellungen wird das fotografische Abbild technisch erzeugt. Die Qualität einer fotografischen Aufnahme ist somit vom Zusammenspiel unterschiedlicher technischer Faktoren abhängig: Aufnahmeformat, Objektiv, Filmtyp, Körnigkeit der Emulsion, Entwicklung, Belichtung und Fokussierung etc.

Um diesen Aspekten in die Bestimmung einer sinnvollen Scanauflösung einfließen zu lassen, bietet sich folgende Alternative zu dem in Kapitel (→ [3.2.1.1](#)) beschriebenen Verfahren an:

Um den individuellen Qualitäten eines durchmischten Bestands umfassend gerecht zu werden, kann eine Auflösung von 80 Linien/mm als Ausgangspunkt zur Festlegung der Scangrößen gewählt werden. Das entspricht der Auflösung eines modernen, feinkörnigen Diafilms. Dadurch wird gewährleistet, dass alle Details sowohl historischer Aufnahmen als auch Aufnahmen neueren Datums im Digitalisat wiedergegeben werden. Die Auflösungswerte moderner Filme sind größtenteils in den Datenblättern der Hersteller dokumentiert. Eine Scanauflösung von 4.000 ppi bildet die Details also ausreichend ab.⁵⁰ Auf das Filmmaterial bezogen wäre die theoretisch erzielbare Auflösung für alle Aufnahmeformate konstant.

Für Kleinbildaufnahmen kann eine Scanauflösung von 4.000 ppi gewählt werden. Für größere Aufnahmeformate kann die Scanauflösung daher verringert werden. Einen entsprechenden Hinweis liefert der maximale Zerstreuungskreisdurchmesser eines Bildpunkts, der aufnahmeformatabhängig vom menschlichen Auge noch als scharf wahrgenommen wird.

Fotografische Praxisregel ist hierbei, dass die Formatdiagonale in mm (entspricht Normalbrennweite) mit 1/1500 multipliziert wird. So ergeben sich folgende maximal zulässige Zerstreuungskreise je Aufnahmeformat:

Kleinbild: 0,03 mm; Mittelformat: 0,05 mm; 9 x 12 cm: 0,1 mm; 18 x 24 cm: 0,2 mm

Daraus abgeleitet ergeben sich folgende Mindestzielgrößen, wenn möglich sollte die physikalische Auflösung der Scanner/Kameras voll ausgenutzt werden:⁵¹

Mittelformat: $4.000 \text{ ppi} \times 0,03/0,05 = 2.400 \text{ ppi}$

9 x 12: $4.000 \text{ ppi} \times 0,03/0,1 = 1.200 \text{ ppi}$

18 x 24: $4.000 \text{ ppi} \times 0,03/0,2 = 600 \text{ ppi}$

Die Fotografien werden immer vollständig mit leichtem, umlaufendem Rand gesichert, um deutlich zu machen, dass nichts von der Vorlage abgeschnitten wurde.

⁵⁰ Um die Filmauflösung in Linien/mm digital darstellen zu können, sind pro Linie mindestens zwei Pixel notwendig. Damit entsprechen 80 Linien/mm 160 Pixel/mm oder 1.600 Pixel/cm. Dieser Wert multipliziert mit 2,54 (1 inch = 2,54 cm) ergibt eine Scanauflösung von 4.064 ppi, gerundet 4.000 ppi.

⁵¹ Vergleiche auch: <https://www.digitalizationguidelines.gov/guidelines/digitize-technical.html>. Die FADGI-Guidelines allgemein unter http://www.digitalizationguidelines.gov/guidelines/FADGI_Still_Image-Tech_Guidelines_2010-08-24.pdf (Federal Agencies Digitization Guidelines Initiative (FADGI): Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files, September 2016).

Aufsichtmedien

Das Fotopositiv ist ein Endprodukt. Es dient anders als das fotografische Negativ nicht als Reproduktionsvorlage oder Ausgangspunkt für eine Vergrößerung. Die Bildqualität ist demnach von der Negativqualität abhängig.

Wie in Kapitel (→ [3.2.1.1](#)) beschrieben, liegt die empfohlene Mindestauflösung für die Digitalisierung bei 300 ppi unter Berücksichtigung der Ergebnisse einer Testdigitalisierung mit genormter Testtafel.

Aufgrund des großen Mehrwerts für die Benutzung sollte in einem ökonomisch vertretbaren Rahmen die volle Auflösung der eingesetzten Aufnahmetechnik bei den Fotografien – Durchsicht und Aufsicht – genutzt werden.

3.2.2.4 Mikroformen

Für Mikroformen (Mikrofilme 16 und 35 mm, Mikrofiche 105 x 148 mm, positiv und negativ, S/W und Farbe) gelten bezüglich der Parameter die Hinweise, die im Abschnitt zu Durchsichtmedien (→ [3.2.2.3](#)) gegeben wurden. Hierbei ist aber für die Auflösung abweichend zu berücksichtigen, dass es sich um Sicherheits- oder Schutzreprografien von Originalen handelt. Die Verfilmung kann mit einem Verkleinerungsfaktor zwischen 1:7,5 und 1:96 stattfinden. Die Auflösung bei der Digitalisierung muss sich also soweit möglich am Original orientieren und nicht am Film.

Die Digitalisierung von Mikrofilmen wird häufig als Massendigitalisierung, die mit niedrigen Kosten realisierbar ist, eingesetzt. Dabei lassen sich die angestrebten 300 ppi in Bezug zur Vorlagengröße nur in den seltensten Fällen realisieren, auch wenn die Auflösung des Films dafür theoretisch ausreicht. Der limitierende Faktor sind hierbei die derzeit marktüblichen Scanner, die in der Lage sind, teilautomatisch in kurzer Zeit ganze Filme zu digitalisieren. Für S/W-Mikrofilme und Mikrofiche können Auflösungen bis 600 ppi bezogen auf den Film erreicht werden. Die Wahl der Auflösung muss sich also am technisch Machbaren der Massenverfahren orientieren. Nur in Ausnahmefällen sollte auf Reproduktionsverfahren zurückgegriffen werden, die Einzelscans mit höherer Auflösung ermöglichen.

Liegen filmische Mikroformen vor, die nach den Standards der Bundessicherungsverfilmung erstellt wurden oder diesen entsprechen (→ [2.3](#)), und dienen diese als Grundlage für eine Digitalisierung, ist zu prüfen, ob die Erstellung und Erhaltung eines Masters sinnvoll ist. Die Mikroverfilmung selbst ist in diesem Fall Langzeitsicherungsmedium und Kopiermaster. Jedoch muss immer geprüft werden, ob die Erstellung von Digitalisaten (Master und Nutzungsversionen) vom Original sinnvoller ist als die Erzeugung von Nutzungsversionen vom Mikrofilm.

3.2.2.5 Dreidimensionale Objekte

Die digitale Erfassung und Vermittlung von 3D-Objekten im Kulturerbe-Bereich hat für eine Reihe von Disziplinen erhebliche Bedeutung und ermöglicht neue methodische Analysen und Erkenntnisse. Die digitale 3D-Dokumentation erfasst dabei alle forschungsrelevanten Objekte wie zum Beispiel durch die 3D-Digitalisierung von Objekten des kulturellen Erbes (wie museale Sammlungsbestände, Architektur und ihre Ausstattung etc.).

Erfolgte diese früher hauptsächlich durch fotografische Abbildung aus verschiedenen Aufnahmestandpunkten, so kann heute auf eine bewährte Auswahl an 3D-Erfassungs- und Rekonstruktionsmethoden sowie diesen Prozessen bestehende Technologien zurückgegriffen werden.

Im Gegensatz zur Fotografie werden bei der 3D-Dokumentation nach Möglichkeit die gesamte Geometrie eines Objekts, immer aber seine Oberflächentextur und nach Möglichkeit seine optischen und physischen Materialeigenschaften erfasst, integriert und im digitalen 3D-Modell zusammengeführt.

Eine umfangreichere Erfassung gegenüber einer Fotografie ist bei der Digitalisierung durch 3D unter anderem durch das Festhalten von Form und Materialeigenschaften gegeben. Durch diese kann die Oberflächen-Licht-Interaktion des Objekts per manueller Nachbearbeitung (mit einem passenden Renderer) originalgetreu erfasst und wiedergegeben werden. Das 3D-Modell kann mit der entsprechenden Software aus beliebigen Perspektiven in der bei der

Aufnahme vorherrschenden Lichtsituation sowie in neuen Lichtsituationen und Umgebungen visualisiert und simuliert werden.

Zur interaktiven Web-Präsentation von 3D-Modellen existieren viele Tools und Game Engines, die als 3D Webviewer durch kommerzielle Anbieter (z.B. Sketchfab) oder als Open Source Entwicklung (z.B. Babylon.js) verfügbar sind. Speziell für die Wissenschaft wurden auch frei verfügbare Lösungen geschaffen. Dazu gehört der sog. 3D Heritage Online Presenter (3DHOP) oder auch das Tool Kompakkt (das im Rahmen der NFDI-Förderung derzeit weiterentwickelt wird und in diesem Zuge auch Punktwolken anzeigbar macht). Sie bieten teilweise spezifische Lösungen und auch ein Repositorium an. Außerdem soll der derzeit in der Entstehung befindliche DFG-3D-Viewer eine weitere Lösung bieten.

Für die Online-Repräsentation sollte das 3D-Modell in entsprechend geeigneter Form vorliegen (konkrete Angaben hinsichtlich Polygon, Anzahl oder Dateigröße sind nicht pauschalisierbar. Formate ergeben sich aus dem verwendeten Tool, meist .glTF oder .obj). Bei der digitalen 3D-Dokumentation geht es im Rahmen dieses Dokuments um die Retrodigitalisierung (Überführung eines physisch vorhandenen Objekts in ein Digitalisat und daraus abgeleiteten 3D-Modellen). Davon ausgehend werden Rekonstruktionen eines physisch nicht (mehr) vorhandenen Objekts in Form eines nativ digitalen 3D-Datensatzes möglich, die hier nicht weiter betrachtet werden.

Digitale 3D-Modelle zeichnen sich durch eine Reihe von Vorteilen aus, die hier beispielhaft genannt sind:

- praktisch beliebige und parallele Verfügbarkeit (auch in unterschiedlicher Qualität) digitaler 3D-Modelle von Kulturobjekten für Wissenschaftlerinnen und Wissenschaftler
- Ergänzung fehlender (u.a. zerstörter) Bereiche infolge der Rekonstruktion
- Simulation verschiedener hypothetischer Versionen und/oder Varianten eines 3D-Objekts
- Einsatz von digitalen 3D-Modellen im Museumsbetrieb, z.B. zur Ausstellungsplanung, Dokumentation, Beschaffungsplanung, etc.
- virtuelle Präsentation und Ausstellung (in Kombination mit neuen Präsentationstechniken, wie z.B. hybriden Exponaten) als Mittel zum Wissenstransfer und Attraktivitätssteigerung
- Referenz für die Restauration beschädigter Originale bzw. Generierung physischer Repliken auf Basis des digitalen 3D-Modells
- Ersatz für Ausleihe (Vermeidung von Beschädigung, Versicherungskosten, Rechtsunsicherheit zu Eigentum)

Datenformate für 3D-Modelle richten sich nach der mathematischen Repräsentationsform, welche auch vom Aufnahme- oder Entstehungsprozess beeinflusst wird. Konkret geht es hier um den Unterschied zwischen Meshes und Punktwolken. Außerdem sind Modelle, welche zur langfristigen Sicherung archiviert werden zu unterscheiden von Webrepräsentationen. Grundsätzlich sollten proprietäre Dateiformate vermieden werden, sodass die Modelle softwareunabhängig nutzbar sind. Die Auflösung der Texturen, die Anzahl der Polygone oder Punkte usw. ergibt sich aus dem Verwendungszweck und der beabsichtigten Repräsentation.

Mit den 3D-Repositorien der HS Mainz und der UB Heidelberg existieren freie Infrastrukturen für die Publikation von 3D-Modellen. Daneben gibt es mit Sketchfab eine nicht freie Plattform (Kosten, Rechte) zum Veröffentlichen, Teilen, Kaufen und Verkaufen von 3D-, VR- und AR-Inhalten. An einem niedrigschwelligen Dokumentations- und Austauschformat fehlt es bislang. Hier sind in absehbarer Zeit Ergebnisse in Form eines Anwendungsprofils für die Kodierung und in MODS abgelegte deskriptive Metadaten zu erwarten.

Es gibt verschiedene optische Digitalisierungsverfahren für die Retrodigitalisierung von 3D-Objekten:

- Laserscanverfahren werden aufgeteilt in Lauflängenmessung und Triangulierung. Im ersten Fall wird ermittelt, wie lange ein Laserpuls zum Objekt und wieder zurück benötigt, um die Entfernung zu diesem

Punkt zu messen. Im zweiten Fall wird mittels einer Kamera aus der Verzerrung der projizierten Laserlinie auf der Oberfläche eines Objekts auf ihren Verlauf in 3D geschlossen.

- Streifenlichtverfahren setzen auf die Projektion definierter Muster (meist parallele Linien) auf Objekt-oberflächen und schließen ebenfalls aus ihrer Verzerrung auf die zugrundeliegende 3D-Oberflächen-geometrie. Zur Erhöhung der Auflösung können diese Muster über die Oberflächen phasenverschoben werden.
- Bei photogrammetrischen Verfahren werden aus verschiedenen Perspektiven Bilder des zu digitalisierenden Objekts angefertigt und nach Merkmalspaaren auf mindestens zwei sich überlagernden Aufnahmen durchsucht, so dass sich für diese mittels Triangulierung die Tiefe des zugehörigen Bildpunktes ergibt. Charakteristische Merkmale können dabei beispielsweise Intensitäts- oder Kontrastunterschiede sein.

Verschiedene Technologien stehen zur Auswahl, die je nach Objektart und Eigenschaften (Größe, Material, Zustand, etc.) zu wählen sind. Gegenwärtig wird in dem Projekt Colour and Space in Cultural Heritage⁵² (COSCH) versucht, einen Online-Ratgeber für die Wahl einer adäquaten Methodik und Technologie je nach Problemstellung zu entwickeln. Zielführend ist jedoch weiterhin der frühe Kontakt mit führenden Technologie-Entwicklern und Anwendern im Bereich der Massendigitalisierung.

Das Prinzip bei Time-of-Flight-Kameras entspricht der Lauflängenmessung bei Laserscannern, mit dem Unterschied, dass eine ganze Szene auf einmal beleuchtet, aufgenommen und für jedes Pixel des Kamerasensors eine Tiefe berechnet wird.

Ab bestimmten Objektgrößen, wie etwa größere Statuen oder gar Gebäude, ist eine punktweise Vermessung wie etwa mit Laserverfahren oder mittels Oberflächenkodierung nicht mehr effizient und photogrammetrische Verfahren zeigen ihre Stärken.

Bei Retrodigitalisierung existiert ein generelles Spannungsverhältnis zwischen Objektgröße und Auflösung, bzw. genauer zwischen dem Messbereich eines Sensors und der Genauigkeit. Je größer das Objekt bzw. je größer die Entfernung, desto größer werden Details erfasst.

Die Geometrie und Textur werden üblicherweise im selben Arbeitsschritt, der Geometrieakquise, erfasst, weil eine direkte Korrespondenz zwischen Messpunkten und Punkten des Bildgebers vorliegt und ein Großteil der Akquise-Methoden einen Kamerasensor als integralen Bestandteil der Erfassung verwenden, um Tiefeninformationen zu extrahieren. Dieser hat pro erfasstem Ausschnitt des Objekts meist bereits die ideale Ausrichtung auf das Objekt, so dass dieselbe Pose des Kamerasensors für die Texturgewinnung genutzt werden kann.

Für die Erfassung der Textur ist eine ausreichende, diffuse und gleichmäßige Beleuchtung erforderlich. Wichtig ist auch ein passender Weißabgleich, da die meist künstliche diffuse Beleuchtung der Objekt-oberfläche kompensiert werden muss. Eine Farb-Kalibrierung ist hierbei wichtig, da Textur einen wesentlichen Einfluss auf die Realitätsnähe hat.

Zusätzlich zu Geometrie und Textur lassen sich auch optische Materialeigenschaften erfassen. Darunter versteht man beispielsweise Glanz- und Reflexionseigenschaften. Typischerweise geschieht dies indem man Kombinationen aus Lichteinfall und Perspektive auf ein Objekt aufnimmt, um im Anschluss die für jeden technisch definierten Oberflächenpunkt (Sichtbarkeitsprüfung) relevante Texturen abzuspeichern, so dass bei späterer interaktiver Betrachtung des virtuellen Objektes möglichst gut unter verschiedenen Beleuchtungsverhältnissen die physikalisch korrekten Texturen appliziert werden und damit eine möglichst gute Annäherung an die originale Materialanmutung und Licht-Oberflächeninteraktion erreicht wird. In manchen Fällen ist eine komplette Materialakquise eines Objekts nicht möglich, entweder bedingt durch seine Größe, oder weil nur Materialreste erhalten sind. In diesen Fällen

⁵² <http://cosch.info/>.

bietet sich eine punktuelle Erfassung von Materialeigenschaften und eine anschließende Extrapolation auf Regionen gleichen oder ähnlichen Materials an.

Einen wichtigen Beitrag zur Integration und Konvergenz zwischen digitalisierten und nativ digitalen Daten leisten die strukturierten Daten, welche sich an einschlägigen Metadata-Schemata (z.B. LIDO und CARARE 2.0⁵³), kontrollierten Vokabularen (z.B. Getty Vocabularies⁵⁴) und Normdateien (z.B. GND) sowie Referenzontologien (allen voran CIDOC CRM⁵⁵) richten. Die Erweiterung von CIDOC CRM zur besseren Erfassung der Retrodigitalisierung fand innerhalb von CIDOC CRM_{dig}⁵⁶ ihren Ausdruck. Die Vorstellung von Applikationsontologien zur Wissensrepräsentation einer computergestützten Rekonstruktion werden gerade in der Fachgemeinschaft vorgestellt und diskutiert.⁵⁷

3.3 Metadaten

Die Erzeugung von Metadaten, welche erst die Auffindbarkeit der Objekte gewährleisten und eine kontextualisierende Präsentation ihrer digitalen Repräsentationen erlauben, ist zentraler Bestandteil der Digitalisierung. Die DFG geht davon aus, dass die der Digitalisierung zu Grunde liegenden analogen Objekte bereits primär in anerkannten digitalen Nachweissystemen erschlossen sind bzw. mit der Digitalisierung einhergehend erschlossen werden. Metadaten, die im Rahmen des Digitalisierungsprojekts entstehen, sind grundsätzlich in einer software-unabhängigen und standardkonformen Form bereitzustellen, in aller Regel in einer XML-Kodierung. Dies ist in den Workflow des Projekts so einzubetten, dass ein vollständiger Metadatenatz in software-unabhängiger Form auch dann bereitsteht, wenn das Projekt – aus welchen Gründen auch immer – zu einem vorzeitigen Ende kommt.

Wenn sich die im Rahmen eines geförderten Digitalisierungsprojekts ausgewählten Objekte sachlich für die Einbindung in ein material- oder fachspezifisches Portal eignen, wird erwartet, dass ein Projektantrag entweder erläutert, welche Vorkehrungen projektseitig getroffen werden, um die Anbindung an dieses Portal während und nach der Projektlaufzeit sicherzustellen, oder plausibel macht, warum eine Anbindung aus inhaltlichen Gründen oder aus Gründen des damit verbundenen Aufwands nicht notwendig bzw. sinnvoll ist.

Allgemein werden deskriptive (bibliografische Beschreibung, archivische Erschließung, Beschreibung von unikalenen, häufig nicht-textuellen Objekten), strukturelle (Text-, Dokumentstruktur), administrative (z.B. Rechteverwaltung) und technische Metadaten (z.B. Dateitypen oder technische Bildmetadaten im Exif-Format) unterschieden. Die folgenden Überlegungen beziehen sich ausschließlich auf deskriptive und strukturelle Metadaten.

Die Verknüpfung zwischen den Metadaten einerseits und den digitalen Repräsentationen andererseits zu einem Objekt muss dabei immer auf der Ebene der Metadaten gewährleistet sein. Zusätzlich können Metadaten auch in den Header der Masterdateien eingebettet werden, jedoch werden diese von den Software-Produkten unterschiedlich dargestellt und im schlimmsten Fall sogar korrumpiert, so dass die Einbettung in jedem Fall nur eine ergänzende Option ist.

Die Metadatenformate realisieren die Referenzierung auf unterschiedliche Weise. Mit dem Containerformat METS können sowohl deskriptive Metadaten über Mappings in Standardformaten wie z.B. MODS oder TEI als auch strukturelle Metadaten inklusive der Referenzen auf die digitalen Derivate transportiert werden. Dies bietet sich insbesondere für Volldigitalisate von Textwerken an, die sowohl mit deskriptiven als auch strukturellen Metadaten erschlossen sind. METS ist aber keineswegs auf Bilddigitalisate beschränkt, sondern kann beliebige digitale Ressourcen referenzieren. Formate wie LIDO enthalten zusätzlich zu den semantischen Elementen zur Objektbeschreibung

⁵³ <https://pro.carare.eu/en/introduction-carare-aggregation-services/carare-metadata-schema/>.

⁵⁴ <https://www.getty.edu/research/tools/vocabularies/>.

⁵⁵ <https://cidoc-crm.org/>.

⁵⁶ <https://cidoc-crm.org/crmdig/>.

⁵⁷ <http://www.digitale-rekonstruktion.info/>.

eigene Elemente zur Referenzierung beliebiger digitaler Ressourcen zum Objekt (neben Bilddigitalisaten auch Audio- oder Videodaten). Auch hier können beliebig viele Ressourcen mit einem Objekt verknüpft und noch mit eigenen Beschreibungselementen versehen sein. In den gängigen EAD-Profilen ist eine Verknüpfung zu beliebig vielen digitalen Ressourcen einschließlich der Referenzierung von METS-Containern sowie weiteren Formaten zur Referenzierung von Digitalisaten oder auch Volltexten möglich. Wichtig ist, dass die digitalen Ressourcen innerhalb der Metadaten über global eindeutige, persistente Adressen, in der Regel URLs, referenziert werden (→ 5).

Der Nachweis der Digitalisate und Metadaten muss bei Bibliotheks- und Archivgut entweder durch Katalogisierung der elektronischen Ausgabe oder durch Angabe der PURL der Bilddateien bzw. der Angabe einer persistenten Verknüpfung im Katalog (OPAC, Verbundsystem, Informationssystem) erreicht werden. Von Antragstellerinnen und Antragstellern aus universitären Einrichtungen wird erwartet, dass sie die Frage der Katalogisierung mit ihren örtlichen Einrichtungen abstimmen bzw. durch diese durchführen lassen.

Die Digitalisate sind in einschlägigen materialspezifischen Portalen nachzuweisen (z.B. das Handschriftenportal⁵⁸, Archivportal-D oder Kalliope⁵⁹ für Nachlässe). Materialien, die nicht in bibliothekarische oder archivarische Nachweisinstrumente eingebracht werden können, sollten in geeigneten fachlichen oder fachübergreifenden Online-Anwendungen präsentiert werden. Der Nachweis der Digitalisate und Metadaten sollte in der DDB erfolgen. Die Einbringung der Daten in die DDB und von dort in die Europeana wird von allen Projekten erwartet. Die Ablieferung an diese Portale sollte gemäß den Standardformaten wenn möglich über OAI erfolgen (→ 3.3.4).

Es obliegt dem Projektnehmer sicherzustellen, dass die im Projekt hergestellten Digitalisierungsdateien (Bild und Text) eindeutig identifiziert und von anderen im selben System vorhandenen Einheiten getrennt durchsucht und aufgerufen werden können.

3.3.1 Erschließung, deskriptive Metadaten

Die Erschließung eines Objekts mit deskriptiven Metadaten erfüllt je nach Material unterschiedliche Funktionen für die Bereitstellung. Dabei wird unterschieden zwischen deskriptiven Metadaten zur Darstellung im DFG-Viewer und denjenigen zur inhaltlichen Beschreibung des Objekts im Portal. Die deskriptiven Metadaten ermöglichen es überhaupt erst, die Inhalte gezielt durch eine Recherche aufzufinden. Das Objekt wird klassifiziert, historisch kontextualisiert und gegebenenfalls in seiner Materialität beschrieben. Im Idealfall bieten die deskriptiven Metadaten Anknüpfungspunkte für unterschiedliche Fragestellungen und Fachdisziplinen. Digitalisierung ohne Nachweis von Metadaten nach den gängigen Community-Standards ist nicht förderfähig.

Für eine Minimalerschließung kann das von der DFG bereitgestellte Basisdatenset⁶⁰ herangezogen werden. Es enthält bereits alle notwendigen verpflichtenden und empfohlenen Datenfelder für die Beschreibung von Metadaten, analogen Ausgangsobjekten und ihre digitalen Reproduktionen. Sofern in einem Digitalisierungsprojekt über deskriptive Metadaten hinaus eine projektspezifische Erschließung und/oder Informationsbereitstellung geplant ist – und dies wird zumindest im Bereich unikalener Objekte der Regelfall sein – sind die wichtigsten Fragestellungen an das Material zu antizipieren und in einem Kernfeldkatalog mit den zu verwendenden Pflichtfeldern und Standards (und ggf. empfohlenen bzw. optionalen Datenfeldern) zu formulieren. Die systematische Bearbeitung der Metadaten auf der Grundlage eines Kernfeldkatalogs ist Voraussetzung für eine optimale Bereitstellung der digitalisierten Bestände (→ 6.2).

⁵⁸ <https://handschriftenportal.de/>.

⁵⁹ <http://kalliope.staatsbibliothek-berlin.de/de/index.html>.

⁶⁰ https://www.dfg.de/formulare/12_155/.

Für eine optimale verteilte, auch nachhaltige Nutzbarkeit der Metadaten ist die Erschließung an den einschlägigen Spartenstandards (z.B. RDA⁶¹, ISAD(G)⁶², RNAB⁶³) und Referenzmodellen (CIDOC-CRM⁶⁴, perspektivisch auch IFLA LRM⁶⁵/FRBRoo⁶⁶) zu orientieren und wo immer möglich mit publizierten Normdaten zu verknüpfen. Ausdrücklich wird für die Erfassung personenbezogener, biografischer Information und für den geografischen Nachweis die Verwendung der Gemeinsamen Normdatei (GND) der Deutschen Nationalbibliothek empfohlen.

Darüber hinaus eingesetzte kontrollierte Vokabulare, z.B. Iconclass⁶⁷ zur Bilderschließung, der Getty Art and Architecture Thesaurus (AAT) oder spezielle Vokabulare wie GeoNames, Getty Union List of Artist Names (ULAN) usw., sollten national und international anschlussfähig sein.⁶⁸

Die Bereitstellung der Metadaten zur weiteren Nutzung gemäß den materialspezifischen Standards ist verpflichtend: METS/MODS für Handschriften und gedruckte Textwerke (→ [Anhang A](#)), METS/TEI als zusätzliche Option für Handschriften (→ [Anhang B](#)), ein gängiges EAD-Profil⁶⁹ in Verbindung mit einem METS/MODS-Mapping für Archivmaterial (→ [Anhang A](#)), LIDO für (i.d.R. unikale) bildhafte und dreidimensionale Objekte (→ [Anhang C](#)). In Entwicklung befindet sich derzeit ein METS/MODS-Profil auch für 3D-Modelle (das in Verbindung mit LIDO genutzt werden kann), multimediale AV-Medien sowie Musikalien. Die bereitgestellten Metadaten müssen gegen das jeweilige XML-Schema valide sein und sind darüber hinaus auf semantische Korrektheit zu überprüfen.

3.3.2 Strukturelle Metadaten für digitale Faksimiles

Wohl zu erwägen ist die Frage der Anwendung von strukturellen Metadaten zur Erschließung von digitalisierten Textdokumenten, also der Kodierung der strukturellen Elemente eines Dokumentes, wie z.B. Widmung, Vorwort, Kapitel oder Illustration (→ [3.4.3](#)). In manchen Fällen ist die Erstellung von strukturellen Metadaten von eher nachgeordneter Bedeutung, bei umfangreichen Kompendien, Wörterbüchern o.ä. ist die Erstellung eines künstlichen Inhaltsverzeichnisses für die Navigation im Digitalisat unverzichtbar. Bei manchen Fragestellungen sind strukturelle Metadaten auch für die Recherche interessant. Die Entscheidung über die Erstellung struktureller Metadaten ist daher immer eine material- und projektspezifische.

Falls strukturelle Metadaten Verwendung finden, wird die Konsultation des über die Website des DFG-Viewers zugänglichen Strukturdatenset empfohlen, eine Liste von Bezeichnungen, die vom DFG-Viewer unterstützt wird.⁷⁰ Sollten darüber hinaus noch andere Bezeichnungen erforderlich sein, sollte man sich im Rahmen konkreter Digitalisierungsprojekte über standardisierte Bezeichnungen verständigen und diese meist fach- oder sachbezogenen Vokabularien über die Projektwebsite und gegebenenfalls auch die Website des DFG-Viewers publizieren, um eine umfassende Nachnutzung zu ermöglichen.

⁶¹ <https://wiki.dnb.de/display/RDAINFO/RDA-Info>.

⁶² http://www.ica.org/sites/default/files/CBPS_2000_Guidelines_ISAD%28G%29_Second-edition_DE.pdf.

⁶³ <https://d-nb.info/1186104252/34>.

⁶⁴ CIDOC Conceptual Reference Model: <http://cidoc-crm.org/>.

⁶⁵ IFLA Library Reference Model: <https://repository.ifla.org/handle/123456789/40>.

⁶⁶ IFLA Functional Requirements for Bibliographic Records - object-oriented: <https://repository.ifla.org/handle/123456789/659>.

⁶⁷ <https://www.iconclass.org/>.

⁶⁸ <https://www.getty.edu/research/tools/vocabularies/aat/>; <https://www.geonames.org/>; <https://www.getty.edu/research/tools/vocabularies/ulan/>.

⁶⁹ <https://wiki.deutsche-digitale-bibliothek.de/pages/viewpage.action?pageId=19010180>. Bei der DDB ist in den nächsten Jahren eine zunehmende Annäherung an den internationalen Standard EAD (3.0) geplant.

⁷⁰ <https://dfg-viewer.de/strukturdatenset>.

Bei der Vergabe von strukturellen Metadaten stellt sich die Frage, ob man sich bei der Erschließung der Dokumente eher am digitalen Faksimile, der physikalischen Seitenfolge, oder an der Text- bzw. Kapitelstruktur des Werks orientiert.

Geht man davon aus, dass dem digitalen Faksimile eines alten Drucks oder einer Handschrift die Transkription oder Edition beigegeben werden soll, ist eine Orientierung am TEI-Kodierungsstandard zu empfehlen. Bei einer Seitenbeschreibung mit einigen qualifizierenden Merkmalen (z.B. Illustrationen oder Annotationen) ist eher METS gemäß dem „Anwendungsprofil für die Verwendung von METS/MODS in der DDB“⁷¹ zu empfehlen. Für beide Modelle, die seitenorientierte und die dokumentenorientierte, gibt es gute Argumente. Meist lassen sich beide Ansätze auch ineinander überführen. Tendenziell ist eine Struktur, die der Logik des Textes folgt, leistungsfähiger, was spätere Abfragemöglichkeiten und Repräsentation der Vorlage betrifft. Allerdings wird dieser Vorteil durch einen höheren technischen Aufwand bei der Bearbeitung und Präsentation der Dokumente erkauft.

Es sei darauf hingewiesen, dass auch eine auf die physikalische Seitenfolge bezogene Kodierung, wie sie in Bibliotheken eher anzutreffen ist, die Verwendung von TEI nicht ausschließt. Für eine Synchronisierung von Volltext/Transkription und Digitalisat bis auf Seitenebene sollte eine Strukturdatenerfassung aber immer in METS erfolgen, unabhängig davon, ob zusätzlich eine Strukturierung in TEI erfolgt oder nicht.

Empfohlen wird nach dem derzeitigen Stand bei alten Drucken eine Orientierung an METS oder TEI. Wenn TEI für Strukturdaten zum Einsatz kommt, ist für die Nutzung des DFG-Viewers deren Konversion zu METS im Rahmen des Projekts notwendig.

3.3.3 Sammlungs- und Bestandsbeschreibung

Sammlungs- und Bestandsbeschreibungen kontextualisieren und verorten das Einzelobjekt und ermöglichen so einen Überblick über den Gesamtbestand einer Einrichtung. Grundlage für ein Digitalisierungsprojekt können auch virtuelle Bestände und Sammlungen (z.B. sachthematisch, materialspezifisch) sein.

Es wird erwartet, dass Digitalisierungsprojekte mindestens Gegenstand und Umfang der jeweiligen Material- bzw. Objektauswahl auf einer Seite im Netz, möglichst auch in Englisch, darstellen. Sinnvoll dafür ist eine normierte Beschreibung, um diese Informationen in Zukunft besser in nationalen oder internationalen Portalen zusammenführen und gezielt recherchieren zu können. Diese Beschreibung kann in demselben Metadatenstandard erfolgen, in dem auch die Objektbeschreibungen verfügbar gemacht werden: MODS oder TEI-Header, die gängigen EAD-Profilen sowie LIDO bieten entsprechende Möglichkeiten. Zunehmend relevant ist außerdem das RDF-basierte Data Catalog Vocabulary (DCAT).⁷² Darüber hinaus sollte für eine Sammlung die eindeutige Identifizierung und Beschreibung gemäß ISO 27730, Information and Documentation – International Standard Collection Identifier (ISCI) erwogen werden, die auf der ISIL⁷³ einer Institution aufbaut.

3.3.4 Austausch und Weitergabe der Metadaten

Für den Aufbau übergreifender Nachweissysteme ist die Schaffung eines Standards zum Austausch von Metadaten für Digitalisate von zentraler Bedeutung (→ 6). Allerdings können sich Standards nur innerhalb der jeweiligen Community ausbilden und etablieren. Dabei können ein und dieselben Ressourcen durchaus im Horizont ganz verschie-

⁷¹ <https://wiki.deutsche-digitale-bibliothek.de/pages/viewpage.action?pageId=19006651>.

⁷² <https://www.w3.org/TR/vocab-dcat-3/>. Siehe auch Data Catalog Vocabulary (DCAT) - Version 2. W3C Recommendation 04 February 2020: <https://www.w3.org/TR/vocab-dcat-2/>.

⁷³ <https://sigel.staatsbibliothek-berlin.de/vergabe/isil>.

dener Fragestellungen erscheinen und entsprechend divergierende Sets von Metadaten erfordern. Ein generalisiertes Verfahren zum Austausch von Metadaten muss daher flexibel unterschiedliche Metadatenformate bzw. community-abhängige Spezifikationen verwalten können. Dieses lässt sich mit dem Protokoll der OAI gut erreichen. Die Verwendung von OAI ist mit Blick auf alte Drucke und Handschriften vor allem als technisches Austauschprotokoll sinnvoll. OAI schreibt vor, dass mindestens Dublin Core-Daten geliefert werden müssen; das ist zwar für die deskriptive Beschreibung sowohl alter Drucke und Handschriften⁷⁴ als auch nichttextueller Ressourcen⁷⁵ ungenügend, als zusätzliche Information aber von Nutzen. Der OAI-Standard sieht die parallele Unterstützung weiterer Metadatenformate explizit vor; eine Verbindung von OAI mit allen XML-basierten Metadatenformaten ist daher grundsätzlich möglich (METS/MODS mit spartenspezifischen Mappings, METS/TEI-HSS, MARCXML, MABxml, gängige EAD-Profilen, TEI P5, LIDO etc.).

Die DFG setzt zwar die Bereitstellung von Metadaten über OAI als Schnittstelle des DFG-Viewers voraus. Dabei müssen neben den von OAI vorgeschriebenen Dublin-Core-Metadaten materialspezifische Metadaten nach METS/MODS für alte Drucke, METS/MODS für Archivgut mit einem für den DFG-Viewer angepassten Mapping, METS/TEI für mittelalterliche Handschriften und LIDO für bildhafte und dreidimensionale Materialien ausgeliefert werden (→ [7](#) und [Anhang](#)). Jedoch ist es an dieser Stelle sinnvoll, auf die IIF Presentation API⁷⁶ als gleichwertige Möglichkeit der Datenbereitstellung zu verweisen. Die Bereitstellung kann auch über ein übergreifendes Portal (z. B. DDB, Archivportal-D) erfolgen. Eine Abweichung von dieser Verpflichtung muss im Antrag ausdrücklich begründet werden.

3.4 Volltextgenerierung

Eine vielseitige wissenschaftliche Nachnutzbarkeit, die unter anderem automatisierte Recherchen, quantitative Auswertungen im Rahmen von Text- oder Datamining, semantische Analysen, Mustererkennungen in nicht textuellen Materialien, Anreicherungen, Kontextualisierungen und Weiterverarbeitungen – auch im Rahmen von virtuellen Forschungsumgebungen – ermöglicht, basiert auf der leichten und ungehinderten Nachnutzbarkeit der entsprechenden Daten, entsprechenden Rechtseinräumungen und dem Angebot des digitalen Volltextes. Wo immer möglich und sinnvoll, wird die DFG-Förderung daher im textuellen Bereich auf die Bereitstellung maschinenlesbarer Volltexte abzielen. Von allen Anträgen zur Digitalisierung textueller Materialien wird daher eine Auseinandersetzung mit Möglichkeiten der Volltextbereitstellung erwartet. In diesem Zusammenhang wird auf das Koordinierungsprojekt OCR-D⁷⁷ verwiesen, das Verfahren zur OCR-gestützten Volltexterfassung, Spezifikationen und Guidelines bereitstellt sowie exemplarische Workflows entwickelt. Ergebnisse dieses Vorhabens sollen in die nächste Version der Praxisregeln einfließen. Unabhängig von diesen Entwicklungen gilt für Druckwerke ab Erscheinungsjahr 1850 verpflichtend, dass Volltext hergestellt werden muss und eine bloße Bilddigitalisierung nicht ausreicht.

Die Herstellung von digitalen Volltexten kann grundsätzlich durch manuelle oder automatische Verfahren erfolgen: entweder durch OCR/HTR oder durch Abschreiben bzw. Transkription. Die Verfahren unterscheiden sich sowohl in der Vorgehensweise als auch in weiteren Bereichen wie bspw. Format und Qualität des Ergebnisses. Die Frage, welches Verfahren gewählt werden soll, ist u.a. mit Blick auf die jeweiligen Anforderungen sowie abhängig vom Alter und Zustand der Vorlage zu entscheiden.

⁷⁴ Vgl. Hillmann, Diane I.: Choices: MARC or Dublin Core? In: Anne R. Kenny/Oya Rieger (Hrsg.): Moving Theory into Practice. Digital Imaging for Libraries and Archives. Mountain View: Research Library Group 2000, S. 89f.

⁷⁵ Die unzulässig vereinfachende Beschreibung musealer Objekte in Dublin Core und deren Darstellung in Online-Umgebungen war der zentrale Anlass für die Museums-Community zur Entwicklung des Harvestingformats LIDO.

⁷⁶ <https://iif.io/api/presentation/3.0/>.

⁷⁷ OCR-D: Koordinierungsprojekt zur Weiterentwicklung von Verfahren der Optical Character Recognition (<https://www.ocr-d.de/>).

3.4.1 Texterfassung

Textgenauigkeiten

Unabhängig davon, ob man einen Volltext durch OCR/HTR oder manuelle Transkription herstellt, ist die Frage zu beantworten, welche Qualität für welche Zwecke benötigt wird und welche Kosten für das jeweilige Ziel angemessen sind. Je nach Projekt werden die Anforderungen an die Textgüte variieren. Denn was dabei noch als ausreichend gilt und welche Kosten für welche Qualität akzeptabel sind, muss in Projekten fallweise, materialabhängig und nach Maßgabe der wissenschaftlichen Anforderungen sorgfältig begründet werden. Gewöhnlich wird die Qualität von mit OCR bearbeiteten Texten in Prozent angegeben. Dabei herrscht wenig Einigkeit darüber, welche Messkriterien und -verfahren angelegt werden. Genauigkeiten können sich auf die Richtigkeit der Buchstaben, aber auch auf die Richtigkeit der Wörter beziehen. Im ersten Fall ist bei 99 % jeder hundertste Buchstabe falsch, im letzteren jedes hundertste Wort. Ob falsche Layoutinformationen (Marginalie zwar richtig erkannt, aber an der falschen Stelle eingeordnet) oder fehlende Worttrennungen als Fehler gelten, wird von Fall zu Fall anders bewertet.

Um eine gewisse Einheitlichkeit bei der Beurteilung der Genauigkeit zugrunde legen zu können, werden Antragstellende gebeten, diesbezügliche Angaben hinsichtlich der Buchstabengenauigkeit zu machen, d.h. mangelhafte Worttrennungen und Layoutfehler unberücksichtigt zu lassen. Es ist also zu überprüfen, wie viele Zeichen der Quelle, einschließlich der Interpunktionszeichen, korrekt erkannt wurden. Ideal sind Messungen auf der Basis zuverlässiger Referenztexte, jedoch stehen diese nicht immer zur Verfügung, so dass auf Stichproben zurückgegriffen werden muss.

Um transkribierte oder mit OCR erstellte Texte auf ihre Genauigkeit hin zu überprüfen, wird hier unbeschadet späterer neuer Empfehlungen, die z.B. aus den Ergebnissen des OCR-D Projektes kommen, ein statistisches Verfahren empfohlen. Ziel ist, mittels einer Stichprobe zu überprüfen, ob eine vom Dienstleister behauptete Erkennungsquote als korrekt eingestuft werden kann, wobei einerseits die Wahrscheinlichkeit für einen Irrtum möglichst gering gehalten werden soll, andererseits aber die Stichprobengröße noch praktikabel ist. Das dazu erforderliche statistische Verfahren ist ein sogenanntes Bernoulli-Experiment. Da die Berechnung relativ kompliziert ist, wird hier mit fest vorgegebenen Werten gearbeitet, die typische Angaben enthalten. Vorgeschlagen wird eine Stichprobengröße von 500 beliebigen Zeichen; zu empfehlen ist die Benutzung eines Zufallsgenerators, der die Position der Zeichen bestimmt

(1. Zeichen: 15. Seite, 24. Zeile, 7. Zeichen. 2. Zeichen: 73. Seite, 3. Zeile, 32. Zeichen usw.). Unter dieser Voraussetzung gilt folgende Tabelle:

| Behauptete Erkennungsquote | Mindestzahl der korrekt erkannten Zeichen (Stichprobengröße = 500). |
|----------------------------|---|
| 95 % | 485. |
| 96 % | 489. |
| 97 % | 493. |
| 98 % | 496. |
| 99 % | 499. |
| > 99 % | 500. |

In der linken Spalte ist die behauptete Erkennungsquote angegeben, in der rechten die Zahl der in der Stichprobe mindestens korrekt erkannten Zeichen, die vorliegen muss, um überprüfen zu können, ob eine von einem Dienstleister behauptete Erkennungsquote als korrekt eingestuft werden kann. Wenn also ein Dienstleister behauptet, dass ein Text eine Genauigkeit von 96 % hat, müssen in der Stichprobe von 500 Zeichen mindestens 489 Zeichen korrekt

erkannt werden, damit bei einer Irrtumswahrscheinlichkeit von 2,5 % die Behauptung des Dienstleisters akzeptiert werden kann. Eine Genauigkeit unter 95 % sollte möglichst nicht vereinbart werden.

Bei einer behaupteten Erkennungsquote von über 99 % müsste die Stichprobengröße erhöht werden. Nachstehend zwei Tabellen, die beispielhaft zeigen, wie hoch die ermittelte Erkennungsquote in Abhängigkeit von einer bestimmten Stichprobengröße sein muss, wenn Texte eine behauptete Erkennungsquote von 99,5 % bzw. 99,7 % haben:

Behauptete Genauigkeit: 99,5 %

| Stichprobengröße . | Mindestzahl der korrekt erkannten Zeichen. |
|--------------------|--|
| 500 . | 500 . |
| 1000 . | 999 . |
| 2000 . | 1996 . |
| 5000 . | 4985 . |
| 10000 . | 9960 . |

Behauptete Genauigkeit: 99,7 %

| Stichprobengröße . | Mindestzahl der korrekt erkannten Zeichen |
|--------------------|---|
| 500 . | 500 . |
| 1000 . | 1000 . |
| 2000 . | 1998 . |
| 5000 . | 4995 . |
| 10000 . | 9990 . |

Natürlich ergeben sich hier pragmatische Grenzen im Aufwand der Stichprobenerhebung. Die Frage, welche Genauigkeit gut, ausreichend oder mangelhaft ist, hängt von der konkreten wissenschaftlichen Fragestellung ab und lässt sich nur aus den konkreten Projektanforderungen her beantworten.

OCR und HTR

Die OCR/HTR-Technik hat in jüngerer Zeit erhebliche Fortschritte gemacht. Wesentliche Verbesserungen konnten bei Frakturschriften sowohl der Handpressen- als auch Maschinenpresszeit sowie bei Handschriften erzielt werden. Gleichwohl können die Praxisregeln angesichts der dynamischen Entwicklung und mit Blick auf die zu erwartenden Ergebnisse insbesondere des OCR-D Projektes zur Frage der OCR/HTR-Software und ihrer Anwendbarkeit derzeit noch keine abschließenden Empfehlungen geben.⁷⁸

Der OCR/HTR-Prozess selbst kennt verschiedene Stufen, die aufeinander aufbauen. Grob unterscheiden kann man das Preprocessing, in dem ein Image vorbearbeitet wird (cropping, despeckle, deskewing, binarisation), die Optical

⁷⁸ Für die maschinelle Erkennung historischer Drucke sowie unterschiedlicher Schrifttypen und Handschriften wurde beispielsweise die freie Software OCR4all entwickelt, vgl. <http://www.ocr4all.org/>. Mittlerweile gibt es eine Verknüpfung mit den in OCR-D erarbeiteten Ansätzen.

Layout Recognition⁷⁹ (OLR = Segmentierung bzw. Layoutanalyse = Identifikation von Bild- und Textregionen, Strukturanalyse), Optical Character Recognition (OCR = eigentliche Texterkennung) bzw. Handwritten Text Recognition (HTR) und Postprocessing (Textkorrektur). Insofern hat die Qualität des Ausgangsbilds ganz entscheidenden Einfluss auf den Prozess der Binarisierung und dieser wiederum auf das Erkennungsergebnis. Daher sollten nur solche digitalen Vorlagen einer OCR-Bearbeitung unterzogen werden, bei denen sichergestellt ist, dass die Bildqualität ausreicht. Problematisch sind in diesem Zusammenhang auch intrinsische Phänomene wie Verschmutzungen, Widerdrucksschatten, manuelle Unterstreichungen oder Annotationen u.ä., die sich nachteilig auf den OCR/HTR-Prozess auswirken. Der Binarisierung folgt die Segmentierung, bei der die Software versucht, die jeweiligen Textbereiche auf einem Imagedigitalisat zu identifizieren, um die eigentlichen Textbereiche von Illustrationen oder anderen Bildelementen zu trennen. Als besonders problematisch erweisen sich bei diesem Prozess komplexe Layouts, etwa Marginalien in Handschriften und älteren Drucken oder auch Zeitungen und andere mehrspaltige Dokumente. Danach folgt der eigentliche OCR/HTR-Prozess der Texterkennung, der durch einen nachgeordneten Prozess der Textverbesserung (automatisch mit Hilfe von Wörterbüchern oder manueller Korrektur) unterstützt werden kann.

Um Daten aus der OCR/HTR leicht nachnutzen zu können, wird die Verwendung des XML-Standards ALTO (Analyzed Layout and Text Object) empfohlen, der von der Library of Congress gepflegt wird.⁸⁰

Manuelle Erfassung/double keying

Man unterscheidet bei der Direkterfassung – dem Abschreiben – von Texten zwei Verfahren, das einfache und das double key-Verfahren. Bei Letzterem wird ein Text zweimal erfasst und die Abweichungen beider Versionen durch automatischen Textabgleich herausgefiltert. Auf diese Weise sind Erfassungsgenauigkeiten von 99,97 % erreichbar, also ein praktisch fehlerfreier Text.

Bei dieser Art Erfassung sollte man sich durch hohe Prozentsätze von Anbietern nicht irreführen lassen. Unterhalb einer Genauigkeit von 99,5 % ist bei manueller Erfassung ein Ergebnis ungenügend.

Sollte die Erfassung durch einen Dienstleister vorgenommen werden, so muss eine brauchbare Textgenauigkeit als Zielvorgabe auch vertraglich fixiert werden. Diese Vorgabe ist an Stichproben des digitalisierten Textes zu überprüfen (s.o.).

Manuelles Erfassen bietet zwar eine hohe Genauigkeit, ist aber auch kostenintensiv, so dass Vor- und Nachteile gegenüber einer OCR/HTR-Erfassung abgewogen werden müssen. Die eigentliche Texterfassung wird hierbei zumeist im Ausland vorgenommen; der Kontakt mit einer Digitalisierungsfirma sollte jedoch über einen Ansprechpartner in Deutschland erfolgen, da in der Regel eine enge Kooperation und Absprache zu den Erfassungsdetails erforderlich ist.

In einem ersten Schritt muss durch das jeweilige Digitalisierungsprojekt festgelegt werden, welche Eigenschaften der Vorlage mittels eines strukturellen Markups erfasst werden sollen. Hierbei können nur solche Eigenschaften ausgezeichnet werden, die grafisch unterscheidbar sind. Einfache Strukturen können vom Dienstleister automatisch erkannt werden, bei weitergehenden Angaben müssen diese im Bild vor der Übergabe an den Dienstleister entsprechend markiert werden. Das verursacht einen entsprechenden Personalaufwand und muss bei der Kalkulation des Projekts berücksichtigt werden.

⁷⁹ Der Begriff "Optical Layout Recognition" stammt ursprünglich von Fraunhofer IAIS und wurde für die Vermarktung ihrer OCR-Software myDec geprägt und hat sich seither etwas verselbstständigt. Im wissenschaftlichen Diskurs ist hingegen eher "(Document) Layout Analysis" bzw. "Segmentation" gebräuchlich.

⁸⁰ <http://www.loc.gov/standards/alto/>.

Da die meisten Dienstleister den Text nach Zeichenmenge inklusive des Markups berechnen, ist es ratsam, eine zeichenarme Auszeichnungsvariante⁸¹ für diese Zwecke zu verwenden.

3.4.2 Zeichenkodierung

Alle verbreiteten Betriebssysteme unterstützen Unicode. Unicode ist auch das Zeichenkodierungsformat von XML, das die Grundlage für die wichtigsten Strukturdaten- Auszeichnungssysteme darstellt.

Daher wird empfohlen, die Texte in Unicode abzuspeichern. Zu favorisieren ist das bei europäischen Sprachen sparsamere UTF-8 (ohne Byte Order Mark BOM). Zeichen, die nicht im Unicode-Standard enthalten sind, können durch Nutzung des private plane-Bereichs von Unicode⁸² abgebildet und durch entsprechende Grafiken oder Fonts repräsentiert werden. Vor der Definition eigener Zeichen in der privat use plane von Unicode ist zu prüfen, ob die Zeichen im community-basierten MUFI-Standard⁸³ abgebildet sind. In allen Fällen sind Möglichkeiten der Standardisierung zu prüfen.

Die Kodierung von langem und kleinem s in Frakturschriften oder Ligaturen in Frakturschriften (ch, tz etc.) bzw. die Darstellung von Diphthongen (æ etc.) ist abhängig von fachspezifischen Anforderungen oder editorischen Entscheidungen, die nicht Gegenstand dieser Empfehlungen sein können, aber bei der Erstellung von Volltext im Blick zu behalten und idealerweise zu dokumentieren sind.

3.4.3 Markup von Volltexten

Wenn nicht triftige Gründe dagegensprechen, müssen Volltexte von Drucken und Handschriften nach dem Modell der TEI kodiert bzw. mit Markup versehen werden. Als transparentes XML-Format ist TEI, sofern sorgfältig dokumentiert, auch für die Langzeitarchivierung die prospektiv beste Wahl. Von PDF/A sollte trotz bestehender ISO-Normen (19005-1:2005 bzw. 19005-2:2011) abgesehen werden, da deren Nachnutzung gerade durch die digital arbeitenden Geistes- und Kulturwissenschaften mangels struktureller Auszeichnung beschränkt ist.

Dessen ungeachtet ist PDF – neben zunehmend auch ePub für mobile Geräte – als derivatives Format für z.B. dynamisch generierte Lesefassungen oder für den Druck aufbereitete Texte gut geeignet und sollte wegen seiner weiten Verbreitung auch zusätzlich von digitalen Bibliotheken angeboten werden (vgl. z.B. die Angebote bei archive.org).

Bei der Kodierung von XML-Strukturen in TEI-Dokumenten muss zunächst entschieden werden, wie und in welchem Umfang man textsortenspezifische Gliederungen wie z.B. Jahresbände, Aufsätze, Kapitel, Unterkapitel etc. berücksichtigt. Hinzu kommen weitere denkbare Strukturmerkmale, z.B. Inhaltsverzeichnisse, Register, Zeilenumbruch, Spaltenumbruch, Seitenumbruch, Kopfzeile/Fußzeile/Kolumnentitel, Seitenzahl, Bilder oder bildähnliche Elemente, Bildunterschriften, Marginalien, aber auch Schriftwechsel, z.B. der Wechsel von Fraktur zu Antiqua (etwa für fremdsprachige Zitate), der Wechsel der Schriftgröße, der Wechsel der Schriftart (recte, kursiv, fett usw.) u.a.m., Formeln, z.B. mathematische (MathML) oder chemische (CML) Formeln, eingebetteter Code, Fortsetzungsmarkierungen (Kustoden) am Fußende von Seiten (für Anschlussbogen) u.a.m.

⁸¹ z.B. TEI Tite: http://www.tei-c.org/release/doc/tei-p5-exemplars/html/tei_tite.doc.html.

⁸² <http://de.wikipedia.org/wiki/Unicode>.

⁸³ <https://mufl.info/m.php?p=mufl>. Vor der Definition eigener Zeichen in der privat use plane von Unicode sollte auch geprüft werden, ob bereits im Rahmen der OCR-D Kodierung (https://ocr-d.de/de/gt-guidelines/trans/ocr_d_koordinationsgre-mium_codierung.html) eine standardisierte Kodierung vorliegt.

Die Wahl des Markup unterliegt in der Regel projektspezifischen Besonderheiten. Um die Austauschbarkeit und Nachnutzung von auf diese Weise mit Markup versehenen Volltexten sicherzustellen, sollten daher die verwendeten XML-Elemente und Attribute im TEI-Header dokumentiert werden. Besondere Beachtung verdienen in diesem Zusammenhang die Bemühung der TEI Community um ein strikt formuliertes Austauschformat: TEI-Simple.⁸⁴

3.4.4 Layout

In manchen Fällen ist es bei der Präsentation eines Volltextes wichtig, das Layout eines Dokumentes langfristig zu sichern. Die Praxisregeln empfehlen für diese Fälle bevorzugt den Einsatz einer geeigneten Formatierungssprache (z.B. XSLT, CSS), die die Unabhängigkeit von spezieller Software weitgehend sicherstellt. Falls eine Archivierung des Formats mit XML-Techniken aus nachvollziehbaren Gründen nicht möglich ist, können Layoutinformationen zu Textdokumenten auch in PDF nach der ISO-Norm 19005-1 archiviert werden. PDF-Dateien sind aber, wie in [3.4.2](#) dargelegt, kein Ersatz für eine Bereitstellung des mit Markup versehenen Volltextes in XML, da sie nicht strukturiert ausgewertet werden können. Neben TEI haben sich für Layoutinformationen inzwischen reichhaltigere XML-Formate wie PAGE XML, und den genannten Standard ALTO XML (Analyzed Layout and Text Object, ALTO v4) etc. etabliert, die ebenfalls empfohlen werden.⁸⁵ Für Präsentationszwecke können diese in HTML/CSS gewandelt werden, die Ausgangsdaten sollten aber in erstgenannten Formaten vorgehalten werden.

Das Wesen einer Publikation als XML + Formatierungssprache bringt es mit sich, dass dynamische Anzeigen nach Maßgabe des jeweils vom Nutzer gewünschten Zwecks generiert werden können. Dies sollte bei der Präsentation berücksichtigt und ein möglichst breites Spektrum angeboten werden. Typische Ausgabeformate von XML-Daten sind z.B. HTML/XHTML, PDF, ePub oder plain text.

3.5 Langzeitverfügbarkeit

Die Langzeitsicherung und Archivierung digitaler Inhalte ist technisch und organisatorisch gelöst, gut durch Modelle beschrieben und anhand von Standards prüfbar.⁸⁶ Man unterscheidet dabei zwischen der Archivierung/Langzeitsicherung genuin digitaler Informationen und der von (Retro-)Digitalisaten, die von vorhandenen analogen Objekten erstellt werden. Hier können unterschiedliche Maßstäbe an die Archivierung gelegt werden.

Im Rahmen der Langzeitsicherung werden Dateien in stabilen, migrationsfähigen Formaten auf einem technisch wie organisatorisch sicheren Speichersystem gesichert (Bitstream Preservation).

Die Archivierung digitaler Daten setzt auf einem solchen Speicher auf, beinhaltet jedoch noch weitreichendere technische und organisatorische Festlegungen, die nicht nur eine physische Erhaltung der Daten, sondern auch Strategien zur Bereitstellung für die Nutzung (Access) der Daten, auch im Kontext vorhandener und zukünftiger Informationssysteme umfassen. Die Speichersysteme müssen auf Redundanz ausgelegt und revisionssicher sein.

Für die Archivierung sind besonders technische Informationen und solche über die Veränderungshistorie eines Objekts wichtig. Besonders, wenn es um die Veränderungshistorie eines Objekts geht, hat sich PREMIS (Preservation Metadata: Implementation Strategies)⁸⁷ als Datenmodell etabliert.

⁸⁴ <https://github.com/TEIC/TEI-Simple>.

⁸⁵ S.o. Kapitel (→1.4), Anm. 21 und 22.

⁸⁶ Für einen guten Überblick: https://www.langzeitarchivierung.de/Webs/nestor/DE/Publikationen/publikationen_node.html.

⁸⁷ http://www.loc.gov/standards/premis/understanding_premis_german.pdf.

Ob die erzeugten digitalen Inhalte langzeitgesichert oder archiviert werden, hängt von der Strategie der jeweiligen Institution ab, die mit einer Antragstellung erläutert werden muss.⁸⁸ Wesentliche Erfolgsfaktoren für eine Langzeitsicherung von digitalen Dokumenten sind

- (1) die Schaffung der organisatorischen und wirtschaftlichen Rahmenbedingungen,
- (2) die Schaffung der technischen Rahmenbedingungen bzw. die Auswahl einer geeigneten technischen Methode/Strategie.

Für die Archivierung digitaler Daten ist das Open Archival Information System (OAIS) als Referenzmodell anzuwenden.⁸⁹ Die „Kriterien für vertrauenswürdige digitale Langzeitarchive“ sind essentiell. Mit ihnen werden u.a. der organisatorische Rahmen, die gesetzlichen Rahmenbedingungen, das Qualitätsmanagement sowie die Authentizität für ein vertrauenswürdigen Archiv geschaffen.⁹⁰ Die Umsetzung der Langzeitsicherung erfolgt in „Digitalen Magazinen“ oder Reproduktionenverwaltungen, die sich am OAIS-Modell orientieren.

Die Langzeitverfügbarkeit der Ergebnisse von Digitalisierungsprojekten ergibt sich einerseits aus der Wahl der Daten- und Metadatenformate. Überlegungen dazu flossen in die vorigen Kapitel ein. Andererseits ist sicherzustellen, dass die digitalen Daten auch physikalisch verfügbar bleiben. Dabei gilt: Kosten für die projektbezogene Sicherung der Daten werden in DFG-geförderten Digitalisierungsprojekten als Eigenleistung für die Laufzeit des Projekts anerkannt. Eine Förderung dieser Kosten aus DFG-Mitteln kann nicht erfolgen.

Die Langzeitsicherungs- bzw. Archivierungsfrage ist ein integraler Bestandteil jedes Digitalisierungsvorhabens. Aufwand und Kosten sollten nicht unterschätzt werden. Nicht nur die Kosten für den Speicherplatz, der je nach Projekt mehrere Terabyte betragen kann, sondern auch der Aufwand für die physische Erhaltung müssen langfristig berücksichtigt werden.

Es sei darauf hingewiesen, dass Digitalisierungsprojekte aus Sicht der DFG stets Projekte der gesamten Einrichtung sind: Die Unterstützung der das Projekt durchführenden Fachabteilung durch die IT-Infrastruktur des Hauses wird vorausgesetzt. Dabei wird begrüßt, wenn sich kleinere Einrichtungen der Kompetenz und der Dienstleistung größerer Einrichtungen bedienen.

Anträge müssen nachvollziehbare Aussagen zur institutionellen Langzeitsicherung und Archivierung enthalten.

4. Organisatorische Fragen – Eigendigitalisierung versus Dienstleistung

Die Digitalisierung kann inhouse oder in Dienstleistung erfolgen. Im ersten Fall können neben Personalkosten ggf. Gerätekosten beantragt werden, sofern sie projektbedingt und nicht der Grundausstattung der Einrichtung zuzurechnen sind.

Die Entscheidung darüber, ob ein Vorhaben als Inhouse-Projekt konzipiert oder ob auf das Angebot eines Dienstleisters zurückgegriffen wird, ist immer eine projektspezifische Entscheidung unter Abwägung aller Sach- und Finanzierungsgründe, die allein in die Verantwortung der Antragstellerinnen und Antragsteller fällt. Schließlich ist das

⁸⁸ Eine Institution kann in ihrer Strategie z.B. schlüssig festlegen, dass born digital-Objekte archiviert werden und Abbilder von analogen Objekten auf langzeitsicheren Speichern abgelegt werden. Ebenso ist es aus guten Gründen möglich, auch die vom Original erstellten Digitalisate zu archivieren, wenn z.B. die Originale fragil sind.

⁸⁹ Das OAIS-Referenzmodell ist als ISO 14721-Standard verabschiedet: <http://public.ccsds.org/publications/archive/650x0m2.pdf>. Deutsche Übersetzung unter: <http://nbn-resolving.de/urn:nbn:de:0008-2013082706>.

⁹⁰ Vgl. Schoger, Astrid/Susanne Dobratz/Reinhard Altenhöner: Kriterienkatalog vertrauenswürdige digitale Langzeitarchive, Frankfurt am Main, 2008. Vgl.: <http://nbn-resolving.de/urn:nbn:de:0008-2008021802>. Vgl. dazu auch: DIN 31644:2012-04.

Heranziehen eines Dienstleisters vor allem eine Sache des Vertrauens. Anders als bei der Filmdigitalisierung, die für das Objekt unkritisch ist, sollte der Dienstleister im Umgang mit unikalem Material oder bibliothekarischem Altbestand Erfahrungen vorweisen können.

Die Arbeit mit einem Dienstleister ist auch dann möglich, wenn die zu digitalisierenden Materialien nicht außer Haus gegeben werden können, da die Dienstleister bei größeren Projekten mit eigenen Geräten und Personal vor Ort arbeiten.

Bei der Formulierung der einschlägigen Verträge ist zu beachten:

- Die geforderten Leistungsparameter müssen exakt festgelegt werden, insbesondere die Anforderungen und Formatstandards für die zu liefernden Rohdaten. Sie sind gemäß den Vorgaben der „Praxisregeln Digitalisierung“ projektspezifisch in einem Lastenheft zu formulieren. Die beauftragten Unternehmen sollten zertifizierte Verfahren zur Qualitätssicherung nachweisen können. Der Auftraggeber ist zu einer sorgfältigen Qualitätskontrolle der Lieferungen verpflichtet, bevor Rechnungen vollständig bezahlt werden.
- Die DFG erwartet, dass ein angemessener Prozentsatz des Rechnungsbetrags als Sicherheit zurückbehalten und erst nach einer durchgeführten Qualitätskontrolle zur Auszahlung an das Dienstleistungsunternehmen kommt. Gleichzeitig soll vom Dienstleistungsunternehmen die schriftliche verbindliche Zusicherung eingefordert werden, dass kurzfristig und kostenlos Ersatzleistung oder Nachbesserung geleistet wird, wenn dies wegen Nichteinhaltung des Lastenhefts oder wegen sonstiger berechtigter Qualitätsmängel nötig werden sollte.

5. Zitieren digitaler Ressourcen, persistente Adressierung

In den Anfangszeiten der Digitalisierung wurde das Problem der Zitierfähigkeit digitaler Ressourcen meist unterschätzt. Dabei ist es gerade die Zitierbarkeit, die die wissenschaftliche Nutzung einer digitalisierten Quelle auch im Internet erlaubt. Im Unterschied zu vormaligen Sekundärformen wie Mikrofilm oder Papierausdruck ist die Internetressource nicht nur eine Kopie des Originals, die wie das Original behandelt und so zitiert wird, sondern ein selbständiges Objekt in einem sich dynamisch entwickelnden integralen Forschungsraum. Wird eine digitale Ressource ins Internet gestellt, muss sie, um von anderen Dokumenten oder Datenbanken aus erreichbar zu sein, eindeutig angesprochen werden können. Dazu ist über die im Analogen übliche Zitierform hinaus, die durch ein Angebot in der Navigationssoftware als klassische Form weiter genutzt werden kann und sollte, die Festlegung und die online zugängliche Dokumentation von Adressierungstechniken nach Maßgabe der in einem Forschungsvorhaben erforderlichen Granularitäten (Buch, Seite, Sammlung, Akte, Objekt, Objektteil, etc.) erforderlich.

Ein positiver Effekt der netzbasierten Zitierform ist, dass das Referenzieren eindeutig wird, was man wegen der vielen Fehler in alten Drucken oder auch aufgrund des Mangels an Paginierungen oder Follierungen etwa in Inkunabeln meist nicht erreichen kann. Daher sollte neben die „inhaltliche“ Zitierform (z.B. p.8, a4, 213r etc.) auch eine „formale“ Zitierform der Bildsequenz treten.

So lassen sich auch Digitalisate eindeutig zitieren, die nicht zum eigentlichen Werk gehören (Deckel, Spiegel, zusätzliche digitalisierte Wasserzeichen, Teilaufnahmen von Illustrationen etc.). Bei Volltexten gelten andere Mechanismen, zu denen derzeit noch keine konkreten Empfehlungen ausgesprochen werden können (eindeutige Referenzierungsmöglichkeiten bestünden z.B. mit XPointer, ID-Vergaben oder vergleichbaren Techniken).

Sichergestellt werden muss die Erreichbarkeit und Zitierbarkeit mittels URI bzw. IRI bei Druckwerken und Handschriften sowohl des Werkes als Ganzem als auch die Erreichbarkeit und Zitierbarkeit von einzelnen physischen Seiten aus diesem Werk. Einrichtungen sollten durch geeignete Mechanismen (PURL, URN, DOI, Handle, etc.) die Persistenz einer Ressource und der Verknüpfung zu ihr gewährleisten, um zuverlässiges Arbeiten mit den bereitgestellten Quellen in wissenschaftlichen Kontexten zu ermöglichen. Ausdrücklich sei auch auf den Nutzen der eindeutigen Zitierbarkeit mittels URI von Ressourcen im Zuge der Weiterentwicklung des Semantic Web und Linked Open Data (LOD) verwiesen.

Die Erzeugung von URNs über die Deutsche Nationalbibliothek mindestens auf Werkebene wird nachdrücklich empfohlen.⁹¹

6. Bereitstellung der Projektergebnisse für die Öffentlichkeit

6.1 Rechte, Lizenzierung und Open Access

Die Digitalisierung wissenschaftlicher Materialien wird durch die DFG gefördert, um diese Materialien der Forschung in Deutschland und weltweit zugänglich zu machen. Alle Projekte sind also so anzulegen, dass die Ergebnisse der Forschung frühzeitig und nachhaltig zugänglich werden. Dies wird in nahezu allen Fällen die Bereitstellung der Digitalisate im Internet einschließen. Sollten Metadaten und/oder Digitalisate aus rechtlichen Gründen nicht frühzeitig nach ihrer Erstellung frei im Open Access verfügbar gemacht werden können, ist sicherzustellen, dass dies zeitnah mit dem Wegfall entgegenstehender rechtlicher Beschränkungen erfolgt.⁹²

Die Deutsche Forschungsgemeinschaft ist Mitunterzeichnerin der „Berlin Declaration“⁹³ zu Open Access. Die Ergebnisse der durch sie geförderten Digitalisierungsprojekte stehen der weltweiten Forschung daher grundsätzlich unentgeltlich und im Geiste dieser Erklärung zur Verfügung. Sollte in Projekten davon abgewichen werden, ist dies im Antrag ausdrücklich zu begründen. Eine Einschränkung des Open Access zu den erzeugten Daten kann z.B. datenschutz-, urheber- oder archivrechtlich begründet sein, darf aber maximal 10 % der insgesamt zu digitalisierenden Materialien betreffen.⁹⁴

6.1.1 Vorgehen vor der Veröffentlichung

Vor der Digitalisierung sind die Rechte an den zu digitalisierenden Materialien zu klären und ggf. für die Digitalisierung und Zurverfügungstellung einzuholen (→ [1.8](#); [2.2](#)). Bei der Digitalisierung (auch bei der Beauftragung von Dienstleistern) entstehen bei einer rein maschinellen Reproduktion (Buchscanner, Scanroboter) keine neuen eigenen Rechte. Dies ist grundsätzlich bei Massendigitalisierungsverfahren von Flachware so. Infolge der DSM-Richtlinie wurde in Deutschland darüber hinaus im Sommer 2021 klargestellt, dass durch die Reproduktion gemeinfreier visueller Werke generell keine neuen Schutzrechte entstehen (vgl. § 68 UrhG).

Wenn das digitalisierte Objekt selbst noch urheberrechtlich geschützt ist, entstehen auch bei der Reproduktionsfotografie neue, eigene Rechte. Gleiches gilt, wenn es sich bei einem Foto um keine reine Reproduktion handelt,

⁹¹ <http://www.persistent-identifier.de/>.

⁹² Rechtlich gesehen ist der Ansatzpunkt des Urheberrechts immer das Werk. Digitalisierungen sind Vervielfältigungen eines Werkes. Lizenzrechtlich sind dabei Unterschiede im Format oder in der Auflösung unerheblich.

⁹³ <http://oa.mpg.de/openaccess-berlin/berlindeclaration.html>.

⁹⁴ Eine Förderung der Bearbeitung nicht zugänglichen Materials mit DFG-Mitteln ist nicht möglich und ein Konzept zur Zugänglichmachung sowohl hinsichtlich der technischen Lösung als auch der zeitlichen Dimension Voraussetzung.

sondern dieses selbst Werkcharakter hat. Bei dreidimensionalen Objekten gibt es Abgrenzungsschwierigkeiten, unter Umständen gelten Fotografien hier als Lichtbildwerke.

Wenn Rechte entstehen, ist sicherzustellen, dass sie der Einrichtung zufallen, die das Vorhaben durchführt und die Institution hier mithilfe der Freigabeerklärung CC0 die Reproduktionen so behandelt, als würden diese neuen Rechte nicht bestehen. Dies gilt insbesondere für die Reproduktionsfotografie, aber auch für dreidimensionale Reproduktionen.

Bei der Erstellung von Metadaten entstehen in der Regel keine Rechte im Hinblick auf die einzelnen Datensätze.⁹⁵ Werden existierende Erschließungsinformationen digitalisiert, die über rein formale Angaben hinausgehend auch Beschreibungstexte enthalten und damit rechtlich geschützt sein können, so ist ggf. eine Rechteeinholung für die Zugänglichmachung und die Nutzungsmöglichkeit der Daten einzuholen.

Im Sinne von Open Access und Open Source wird vorausgesetzt, dass alle Ergebnisse – Metadaten, Digitalisate und Volltexte bzw. das den Volltexten zu Grunde liegende XML, DTDs oder XML-Schema-Dateien sowie Skripte und Programmcode – so frei nachnutzbar wie rechtlich möglich angeboten werden, sei dies durch die Markierung als Public Domain oder die Verwendung freier Creative Commons-Lizenzen⁹⁶ bzw. FLOSS-Lizenzen für freie Software.

Bei der Digitalisierung gemeinfreier Materials wird die Markierung der Digitalisate als Public Domain erwartet. Bei allen Ergebnissen, bei denen dies nicht möglich ist, weil Rechte bestehen oder bei der Digitalisierung entstanden sind, ist ebenfalls sicherzustellen, dass die Daten im Open Access und so frei nachnutzbar wie möglich zur Verfügung gestellt und lizenziert werden. In Frage kommt ein vollständiger Verzicht auf entstandene Rechte durch die Markierung mit CC0 (z.B. bei Metadaten), aber auch die Verwendung der freien Lizenztypen CC-BY oder CC-BY-SA (→ [1.8](#)). Wird noch geschütztes Material digitalisiert und kann das Digitalisat wegen der noch bestehenden Rechte am Ausgangsmaterial nicht frei lizenziert werden, sollte der Rechtsstatus mit standardisierten Rechteinweisen versehen werden.⁹⁷ Bei Projekten, in denen nicht nur gemeinfreie Materialien digitalisiert werden und/oder mit kommerziellen Partnern oder Verlagen zusammengearbeitet wird, kann eine verzögerte Open Access-Publikation (moving wall) von bis zu einem Jahr nach Projektende vereinbart werden.

6.1.2 Umgang mit verwaisten/nicht verfügbaren Werken

In den Beständen von Bibliotheken, Archiven und Museen befinden sich oft große Mengen vergriffener Werke, die von hohem Wert für Forschung und Bildung sind. Die Rechtereklärung von urheberrechtlich geschützten Materialien erweist sich jedoch gerade bei älteren Materialien und solchen, die nicht in einem Verwertungszusammenhang erstellt wurden, als äußerst schwierig.

Der Gesetzgeber hat daher mit den Regelungen zu verwaisten Werken (§§ 61 UrhG ff) und zu nicht verfügbaren Werken (§§ 61d UrhG ff) gesetzliche Erlaubnisse geschaffen, die Kulturerbe-Einrichtungen die Online-Stellung von Digitalisaten ermöglicht – auch ohne dass die dafür notwendigen Nutzungsrechte geklärt wären. Die dadurch erreichte Zugänglichkeit genügt jedoch nicht den Anforderungen an eine Open Access-Publikation, da beispielsweise eine Creative Commons-Lizenzierung und damit eine Nachnutzung nicht zulässig ist. Auch eine dauerhafte Online-

⁹⁵ Siehe: Klimpel, Paul, Das Eigentum an Metadaten, in Euler, Ellen/Monika Hagedorn-Saupe/Gerald Maier/Werner Schweibenz/Jörn Sieglerschmidt; Handbuch Kulturportale. Online-Angebote aus Kultur und Wissenschaft, Berlin/Boston 2015, S. 57 – 64.

⁹⁶ Es ist immer die aktuellste Version der Creative Commons-Lizenzen zu verwenden: <https://creativecommons.org/>.

⁹⁷ www.rightsstatements.org. Siehe auch die Empfehlungen für Rechteinformationen in Metadaten auf der Wiki-Seite der DINI AG KIM: <https://wiki.dnb.de/pages/viewpage.action?pageId=217533652>.

Verfügbarkeit ist nicht sichergestellt, da sowohl bei verwaisten als auch bei nicht verfügbaren Werken der Rechteinhaber jederzeit Widerspruch gegen die Onlinestellung erheben kann. Hinzu kommt, dass eine Präsentation der Objekte selbst in der Deutschen Digitalen Bibliothek (DDB) oder anderen Portalen dann nicht möglich ist, wenn dafür die Übertragung von Nutzungsrechten an das Portal erforderlich ist.

Gleichwohl kann es dort, wo eine Online-Stellung rechtlich schlicht nicht zulässig ist, unter Umständen sinnvoll und geboten sein, nach den Regelungen für verwaiste oder nicht verfügbare Werke zu verfahren. Da dies eine Ausnahme vom Gebot der Open Access-Publikation ist, ist es auch gesondert zu begründen, etwa wenn die entsprechenden Materialien ohnehin in absehbarer Zeit gemeinfrei werden und die gesetzlichen Erlaubnisse nur für eine kurze Übergangszeit benötigt werden.

Um eine Publikation in die Wege zu leiten, ist zunächst die Recherche der Rechtesituation (Ermittlung möglicher Urheber) und die Eintragung der Werke auf den entsprechenden EU-Portalen erforderlich, damit diese dort von Rechteinhabern gefunden werden können.⁹⁸ Für Kulturerbeeinrichtungen gilt außerdem:

- Entfernung der Images auf Verlangen eines Rechteinhabers,
- keinen Download online ermöglichen,
- keine Weitergabe an Portale,
- keine Nachnutzung der Digitalisate.

Wichtig ist festzustellen, wie lange ein urheberrechtlicher Schutz noch besteht und dafür Sorge zu tragen, dass nach Ablauf der Schutzfrist die Materialien als gemeinfrei gekennzeichnet und ohne Einschränkungen zugänglich gemacht werden.

In jedem Fall ist sicherzustellen, dass die Metadaten zu solchen Digitalisaten dauerhaft online sind, damit die jeweiligen Objekte in den Institutionen, die sie beherbergen, am Terminal genutzt werden können.

6.1.3 Zugang und Nachnutzung

Maßgebliches Ziel der freiest möglichen Verfügbarmachung der Ergebnisse und der vereinheitlichten Rechteangaben ist es, für nicht unter Abschnitt 6.1.2 fallende Daten Möglichkeiten zu schaffen, Digitalisate, Metadaten und Volltexte auch in anderen Nutzungskontexten als der unmittelbaren Projektumgebung (z.B. zum Zwecke des data mining oder der Datenaggregation) auswerten zu können; dies erfordert das Herunterladen, die Neuindexierung und das Anbieten in eigenen Forschungs- und Präsentationskontexten. In diesem Kontext gilt für Bilddigitalisate aktuell, dass sie in einer Form bereitgestellt werden, die grundsätzlich eine vollumfängliche wissenschaftliche Nachnutzung in anderen Forschungskontexten erlaubt (z.B. durch Nutzung von einzelnen Images „in Frames“ nach dem Modell des DFG-Viewers bzw. der iFrame-Einbettung von Google LLC oder durch generelle Erteilung einer Erlaubnis, Vorschaubilder auf fremden Servern anzeigen zu dürfen). Die Daten sollten ohne technische Zugangsbeschränkungen (Login, IP-Schranken, etc.) verfügbar sein und so ausgeliefert werden, dass CORS (Cross-Origin Resource Sharing) gestattet ist. Nur so kann technisch gewährleistet werden, dass die Daten tatsächlich unbeschränkt auch im Kontext anderer Applikationen/Webangebote nachgenutzt werden können. Dazu sollen bei den Bilddigitalisaten entweder hochaufgelöste Derivate im TIFF-Format oder vollaufgelöste JPEG-Bilder mit einer Kompression von 90, maximal 80 in Kombination mit der vollen Scanauflösung zur Verfügung gestellt werden. Die Erhebung

⁹⁸ <https://euipo.europa.eu/ohimportal/de/web/observatory/orphan-works-db> (verwaiste Werke); <https://euipo.europa.eu/ohimportal/de/web/observatory/outofcommerceworks> (vergriffene Werke). Nähere Informationen zu verwaisten Werken beim Deutschen Patent- und Markenamt, vgl. <https://www.dpma.de/dpma>.

von Entgelten für Kopien in darüberhinausgehenden Qualitäten, Derivaten oder für die Herstellung anderer Medienformen (CD, Druck, etc.) bleibt davon unberührt.

Die DFG erwartet, dass DFG-geförderte Projekte auf ihren im Internet angebotenen Daten eindeutige Herkunftsnachweise sowie soweit möglich den Hinweis auf die Förderung durch die DFG anbringen. Dies erfolgt im Fall von Bilddigitalisaten noch immer durch das Hinzurechnen einer Nachweisleiste in die veröffentlichte Nutzerkopie auf der Website (Derivat) oder eine Footerzeile beim Download von Digitalisaten (zum Beispiel im Grafikformat JPEG).⁹⁹ Die Bilder können dann später besser zugeordnet werden (Provenienzfrage). TIFF-Dateien zum Download sind ohne Footer bereitzustellen. Eingriffe in den Darstellungsbereich der Digitalisate (etwa durch Ergänzung von Wasserzeichen) sind nicht zulässig. Im Fall von Volltexten geschieht der Nachweis durch entsprechende Hinweise im Header der Textdatei, ebenso können die Nachweise in die Bildheader eingebettet werden. In jedem Fall aber müssen die Herkunftsangabe und der Hinweis auf die Förderung in den begleitenden Metadaten enthalten sein und im Bereitstellungssystem auch ausgegeben werden.

6.2 Mindestanforderungen für die Bereitstellungssysteme von Digitalisaten

6.2.1 Funktionalitätsanforderungen

Sammlungen/Bestände können in der Regel über eine Vielzahl von Wegen zugänglich gemacht werden:

- über die Website der das Projekt durchführenden Institution
 - alle Objekte sind dort mit Navigationsmöglichkeiten zu versehen
 - sie sind dort so anzubieten, dass langfristig zitierfähige URLs möglichst niedriger Granularität entstehen. Die Form, in der zitiert werden soll, ist deutlich anzuzeigen
- über den lokal implementierten oder an einem anderen Ort betriebenen DFG-Viewer, sofern für das angebotene Material zutreffend (→ [7.3](#))
- über Nachweise im lokalen und regionalen Bibliothekskatalog / das lokale und regionale Archivportal
- über überregionale Nachweis- und Präsentationssysteme, insbesondere die DDB bzw. das Archivportal-D
- sofern vorhanden über eines der von der DFG geförderten materialspezifischen Portale, die einen integrierten Zugriff auf alle im DFG-Programm geförderten digitalen Sammlungen ermöglichen, bzw. ein gemeinsames fachliches Portal z.B. der Fachinformationsdienste
- über eine OAI-Schnittstelle. Über die OAI-Funktionalität ist auch der DFG-Viewer zu bedienen (sofern das jeweilige Format unterstützt wird), wobei die Schnittstelle in jedem Fall vorrangig ist.

Neu eingerichtete OAI-Schnittstellen sollten den in Frage kommenden Portalen und Fachinformationsdiensten gemeldet werden. Ferner sollten geeignete Maßnahmen ergriffen werden, die dazu führen, dass die Metadaten von Suchmaschinen gefunden werden (z.B. mit dem sitemap protocol¹⁰⁰).

Den Nutzerinnen und Nutzern sollte neben dem gezielten Zugriff auf spezifizierte Dokumente über eine Recherche in Metadaten (searching) auch die Möglichkeit geboten werden, in vorab definierten Sammlungen, Sammlungsteilen oder Beständen strukturiert zu navigieren (browsing). Bei der Suche ist darauf zu achten, dass im Zweifelsfall

⁹⁹ Derzeit wird über eine Verpflichtung zu einem sichtbaren Herkunftsnachweis ohne sichtbare Veränderung der Derivate beraten. Herkunftsnachweise könnten alternativ etwa in den deskriptiven Metadaten, in den Image-Metadaten und optisch innerhalb der Webpräsentation untergebracht werden. Dies ist jedoch aktuell (noch) keine Empfehlung der DFG.

¹⁰⁰ <http://www.sitemaps.org/>.

einfache, googleorientierte Suchinstrumente für Nutzerinnen und Nutzer intuitiver bedienbar sind als stark gefelderte Suchmasken, die ein eingehendes Verständnis der Datenstruktur der jeweiligen Sammlung oder des jeweiligen Bestands erfordern. Idealerweise werden beide Aspekte durch die Facettierung der Suche kombiniert, bei der die Nutzerinnen und Nutzer Möglichkeiten an die Hand bekommen, selbst größere Treffermengen nach definierten Kriterien (Facetten) zu reduzieren. Darüber hinaus bleibt auch die differenzierte, stark gefelderte Suche als zusätzliches Hilfsmittel für eine hochspezialisierte Fachcommunity ein wünschenswertes Angebot.

Außerdem sind folgende Funktionen zu implementieren (sofern es sich nicht um verwaiste oder nicht verfügbare Werke handelt):

- Downloadfunktion¹⁰¹
- Druckfunktion für die ausgegebene Dokumentansicht¹⁰²
- Aus zentralen DFG-geförderten Informationssystemen (VD16, VD17, VD18, Fachportalen, materialspezifischen Portalen etc.) sollte zuerst auf eine Ansicht im Style „DFG-Viewer“ verlinkt werden.

6.2.2 Technische Mindestanforderungen

Die Präsentationsanwendung/-plattform ist so zu konzipieren, dass sie, soweit anwendbar¹⁰³

- alle Materialien in einer hinreichend guten Qualität anbietet, so dass die Benutzung zu wissenschaftlichen Zwecken auf an den Hochschulen üblicherweise vorhandenem Equipment mühelos möglich ist. Dies wird beispielsweise dadurch erreicht, dass Schrift in einer Größe angeboten wird, die problemlos gelesen werden kann.
- alle Materialien in einer Qualität anbietet, die ohne unzumutbare Verzögerungen bearbeitbar sind.
- zum wissenschaftlichen Gebrauch den kostenlosen Download kompletter Einheiten in einer einzigen Datei (beispielsweise einzelne gedruckte Werke) ermöglicht.
- alle jeweils gängigen Browser unterstützt, soweit dies der Natur der Sache nach nicht unmöglich ist.¹⁰⁴

7. Präsentationsstandards für Textwerke, DFG-Viewer

Die oben angesprochenen Grundsätze gelten für alle Arten von Projekten, die digitale Inhalte bereitstellen. Bei der Bereitstellung von Digitalisaten, die den Charakter von digitalen Büchern oder mehrseitigen Dokumenten haben, gelten zusätzlich folgende Mindestanforderungen. Diese betreffen einige Basisanforderungen und die mindestens zu realisierende Funktionalität.

¹⁰¹ Ein Download nach Abschnitten oder einzelnen Seiten ist dann vorzusehen, wenn die Größe der Gesamtdatei nicht mehr handhabbar wäre. Grundsätzlich ist auch eine Bereitstellung in Originalgröße des Bildmasters für die Forschung unentgeltlich (mindestens JPEG 80% Kompression) zu ermöglichen.

¹⁰² Eine Druckfunktion nach Abschnitten oder einzelnen Seiten ist dann vorzusehen, wenn die Größe der Gesamtdatei nicht mehr handhabbar wäre.

¹⁰³ Hier wird nachdrücklich das Kriterium der Alltagstauglichkeit gefordert, es geht nicht um die Realisierung abstrakter Wünschbarkeiten. Können Objekte eines Projekts ihrer Natur nach auf Bildschirmen mit einer Auflösung unter 1600 x 1200 nicht sinnvoll dargestellt werden, wird kein Aufwand für Scheinlösungen gefordert; ist ein Objekt unter 3 MB nicht sinnvoll bearbeitbar, verletzt es das Kriterium der DSL-Tauglichkeit des Servers nicht, wenn kein Angebot kleinerer Versionen erfolgt.

¹⁰⁴ Wird ein für eine fortgeschrittene 3D-Anwendung notwendiges Format von einem Browser nicht unterstützt, ist es nicht erforderlich den Aufwand zu betreiben, ein geeignetes Plugin zu entwickeln.

7.1 Basisanforderungen und Architektur

Das Bereitstellungssystem verknüpft die digitalisierten Bild- oder Volltextdateien zu einer Dokumentstruktur, um für die Nutzerinnen und Nutzer eine Navigation innerhalb des Dokumentes zu ermöglichen. Weiterhin stellt es Verbindungen zwischen digitalen Dokumenten bzw. Teilen davon (z.B. Kapitel, Seiten) und Metadaten her, um den Nutzerinnen und Nutzern, ausgehend von einer Recherche in Metadatenbeständen, den Zugriff auf das einzelne Dokument oder auf bestimmte Dokumentteile zu ermöglichen. Schließlich organisiert es digitale Dokumente zu sachlich oder durch Provenienz zusammengehörigen digitalen Sammlungen bzw. Beständen, um den Nutzerinnen und Nutzern die Navigation zwischen Dokumenten und Sammlungen fachgerecht anzubieten. Es stellt entsprechende Benutzeroberflächen für Recherche, Navigation, Zugriff und Abruf von Metadaten, Dokumenten, Sammlungen und Beständen zur Verfügung und unterstützt den weitgehend automatisierten Ex- und Import von standardkonformen Rohdaten. Die Bereitstellungssysteme der einzelnen Informationsinfrastruktureinrichtungen sollten sowohl bei der Navigation in digitalen Sammlungen/Beständen als auch bei der Recherche über Indizes einen institutionenübergreifenden Zugriff ermöglichen. Darüber hinaus ist eine transparente Verknüpfung der Bereitstellungssysteme sowohl mit lokalen Katalog-/Informationssystemen als auch übergreifenden Informationssystemen wünschenswert.

Zur Lösung dieser Aufgaben können unterschiedliche Systemarchitekturen eingesetzt werden. Folgende grundlegende Alternativen sind denkbar:

- Die Metadaten werden zentral in einem Online-Informationssystem (z.B. dem lokalen OPAC bzw. Online-Findmittelsystem oder einem übergreifenden Informationssystem wie einem Bibliotheksverbundkatalog oder dem Archivportal-D) gehalten, die digitalen Dokumentdateien inklusive XML-codierter Strukturdaten werden auf einem gesonderten Dokumentenserver für den Online-Zugriff bereitgestellt. Die Struktur der digitalisierten Sammlung bzw. die interne Struktur der digitalisierten Dokumente kann dabei durch die Hierarchie des Dateisystems abgebildet werden.
- Ein Dokumenten-Management-System (DMS) oder Content-Management-System (CMS) kommt zum Einsatz, bei dem sowohl die Metadaten als auch die Digitalisate im Datenbank-System gespeichert sind.

Normalerweise kommt die erste Variante zum Einsatz, die eine verteilte und transparente Informationsinfrastruktur ermöglicht.

7.2 Funktionalitätsanforderungen

Ein wesentliches funktionales Qualitätskriterium ist der Komfort bei der Navigation innerhalb eines aufgefundenen Dokuments. Die folgenden Navigationsmöglichkeiten gelten als Basisanforderungen:

- Ansteuern eines beliebigen Bilddigitalisats
- Anfang: Springen an den Anfang eines Dokuments
- Ende: Springen an das Ende eines Dokuments
- Vor: Eine Seite vorgehen
- Zurück: Eine Seite zurückgehen
- Volltextrecherche innerhalb der Digitalisate, sofern vorhanden (für Bücher ab 1850 verpflichtend)¹⁰⁵

¹⁰⁵ Nach dem Stand der gegenwärtigen Technik ist eine OCR-Erkennung bei Drucken der Maschinenpressenzeit ab 1850 verpflichtend.

- Metadaten-Info: Hier können Nutzerinnen und Nutzer die Informationen aus den im Informationssystem gespeicherten Beschreibungsfeldern zu „ihrem“ digitalen Dokument einsehen.
- Hilfe: Über das Hilfemenü sollte eine detaillierte Beschreibung mit Fallbeispielen zur Navigation und für die Suche in der Digital Library zugänglich sein.

Wenn sachlich möglich und geboten, sind Inhaltsverzeichnisse/Strukturbäume oder funktionale Äquivalente vorzusehen; diese sind grundsätzlich durchsuchbar zu gestalten. Wünschenswert sind Navigationshilfen, z.B. grafische Repräsentationen in einer Kopfzeile, die den Nutzerinnen und Nutzern signalisieren, an welcher Stelle des digitalen Dokuments sie sich gerade befinden. Enthält ein Informationssystem Materialien, die konzeptuell in der Benutzung üblicherweise zu übergeordneten Einheiten zusammengefasst werden (mehrbändige Werke), so sollen diese übergeordneten Einheiten als solche sichtbar werden.

7.3 DFG-Viewer und IIF

Um für die wissenschaftliche Benutzung neben den unterschiedlich gestalteten und dezentral verantworteten Webangeboten der jeweiligen Einrichtungen einen einheitlichen Zugriff auf die Daten (Inhalte) aller DFG-geförderten Digitalisate bieten zu können, werden derzeit zwei sich ergänzende Strategien verfolgt:

- Festlegung eines einheitlichen Designprofils für die Visualisierung von in DFG-Förderung erstellten Digitalisaten („DFG-Viewer“). Eine einheitliche Präsentation aller DFG-geförderten Projekte ist auf diese Weise möglich. Der Viewer dient zugleich als Referenzimplementierung für die geforderten Standards, anhand dessen digitalisierende Einrichtungen prüfen können, ob sie diesen Standards entsprechen (Überprüfung der gewünschten Implementierung). Diese Funktion erfüllt der DFG-Viewer auch weiterhin.
- Wichtiger als die Anbindung des Viewers ist die Schaffung einer OAI-PMH Schnittstelle mit Unterstützung des METS-Standards: Diese dient zunächst der einheitlichen Anzeige von Derivaten und ihren Metadaten in DFG-geförderten Projekten und transportiert materialabhängig METS/MODS für Drucke und Archivgut und METS/TEI für Handschriften.

Zum Zweck der überregionalen Erstpräsentation wird für DFG-geförderte Digitalisierungsvorhaben die Bedienung der genannten Schnittstellen empfohlen. Es ist jedoch nicht verpflichtend, den DFG-Viewer an der eigenen Einrichtung zu implementieren.¹⁰⁶ Stattdessen können andere Viewer mit dazu vergleichbarem Funktionsumfang eingesetzt werden. [Anhang A](#) definiert das METS/MODS- und [Anhang B](#) das METS/TEI-Format der Schnittstelle sowie das Viewer-Design.

Da das International Image Interoperability Framework (IIIF), das gleich mehrere APIs vereinigt, zunehmend Verbreitung findet, soll an dieser Stelle zusätzlich auf die IIIF Presentation API als gleichwertige Möglichkeit der Datenbereitstellung verwiesen werden. IIIF garantiert eine hohe Interoperabilität im Austausch von digitalen Objekten. Es kann jedoch den oben genannten Referenzstatus des DFG-Viewer-Standards nicht ersetzen, weshalb dieser in den vorliegenden Praxisregeln vorrangig empfohlen wird.

Für die nun folgenden Anhänge A-C soll an dieser Stelle zusätzlich auf die Dokumentation zu den Lieferformaten für die Datenlieferungen an die DDB verwiesen werden:

<https://wiki.deutsche-digitale-bibliothek.de/display/DFD/Lieferformate>

¹⁰⁶ Eine lokale Installation durch das aggregierende Portal oder die datenliefernde Einrichtung ist in diesem Fall nicht erforderlich.

Anhänge

Anhang A: METS/MODS-Profil für die Darstellung im DFG-Viewer und Übermittlung per OAI

1. DFG-Viewer

Zum Zweck der einheitlichen Präsentation beim Zugriff aus überregionalen Nachweissystemen (z.B. Deutsche Digitale Bibliothek, Archivportal-D) auf lokale digitale Angebote sollten DFG-geförderte Projekte die als DFG-Viewer bekannte Browserdarstellung nutzen sowie die dem Viewer hinterlegten Schnittstellen bedienen. Wissenschaftlerinnen und Wissenschaftlern soll auf diese Weise die Nutzung digitalisierter Inhalte erleichtert werden. Aus dem DFG-Viewer heraus ist die Weiterleitung auf die speziellen lokalen Angebote der jeweiligen Einrichtungen möglich.

Der Viewer¹⁰⁷ wurde im Rahmen der Aktionslinie „Digitalisierung des VD16/VD17“ durch die in der ersten Ausschreibungsrunde geförderten Bibliotheken aufgebaut. Er wird durch die SLUB Dresden in Zusammenarbeit mit weiteren Partnern beständig weiterentwickelt. Gehostet wird die Referenzanwendung des DFG-Viewers derzeit an der SLUB Dresden.

Um den DFG-geförderten Projekten eine größtmögliche Sicherheit bei der Planung ihrer Anträge zu ermöglichen und um die Konformität der Metadaten mit den Anforderungen des DFG-Viewers sicherzustellen, sollten Metadaten aus Projekten gegen eines der von der Website des Viewers abzurufenden Anwendungsprofile valide sein.¹⁰⁸

Zur Darstellung von Metadaten im Viewer wird METS¹⁰⁹ verwendet. METS dient dabei als ein Rahmenformat (wrapper), innerhalb dessen sowohl deskriptive, administrative und strukturelle Metadaten als auch Ressourcen (z.B. Bilddigitalisate, Volltexte) verzeichnet werden. Der DFG-Viewer benötigt zum Zwecke der Anzeige von deskriptiven Metadaten nach MODS¹¹⁰ kodierte Metadaten (für Drucke, Zeitschriften, Zeitungen, Archivgut) bzw. TEI kodierte Metadaten (für mittelalterliche und frühneuzeitliche Handschriften). Für administrative Metadaten (u.a. Rechtsinformationen und Verlinkung auf die lokale Anwendung) wird ein spezielles, für den Viewer entwickeltes Format (Namespace dv) verwendet.

Eine ausführliche Dokumentation zur Implementierung des METS-Formats befindet sich auf der Homepage der Referenzanwendung des DFG-Viewers.

Diese Vorgaben gelten grundsätzlich für alle digitalisierten Textwerke. Für mittelalterliche und frühneuzeitliche Handschriften kann als Formatalternative bzw. zusätzlich METS/TEI verwendet werden (→ [Anhang B](#)). Für bildhaltende und dreidimensionale Objekte ist anstelle von METS/MODS bzw. METS/TEI die Verwendung von LIDO verpflichtend (→ [Anhang C](#)). Zur Kodierung von archivischen Findmitteln muss EAD(DDB) verwendet werden.¹¹¹

¹⁰⁷ <https://dfg-viewer.de/>.

¹⁰⁸ <https://dfg-viewer.de/metadaten>.

¹⁰⁹ <http://www.loc.gov/standards/mets/>.

¹¹⁰ <https://www.loc.gov/standards/mods/>.

¹¹¹ s. <https://wiki.deutsche-digitale-bibliothek.de/pages/viewpage.action?pageId=19010180>.

2. MODS-DFG-Standard-Set (Bibliotheksbestände)

Für die Anzeige im DFG-Viewer sind nur wenige Pflichtfelder erforderlich (s. nachstehende Tabellen).

1 Titelinformation

| Element/Unterelement | Wiederholbar | Bemerkungen/Erläuterungen | Status |
|--------------------------|--------------|---|---|
| <titleInfo> | ja | Titelinformationen; sofern das Werk keinen Titel enthält, muss ein aussagekräftiger Titel fingiert werden. In Fällen von Bänden bzw. Heften von mehrbändigen Werken ohne Stücktitel, Zeitschriften oder Zeitungen reicht es aus, den Titel des Gesamtwerks in dem Element <mods:relatedItem type="host"> anzugeben. Verpflichtend ist dann aber die Angabe der Zählung in dem Element <mods:part> mit den Unterelementen <mods:detail> und <mods:number>. | Pflicht bei einteiligen Textwerken und bei den Bänden mehrbändiger Werke mit Stücktitel |
| <titleInfo> / <title> | nein | Enthält den Haupttitel des Werks. | Pflicht |
| <titleInfo> / <subTitle> | ja | Enthält den Zusatz zum Haupttitel des Werks. | Pflicht, wenn vh. |

Beispiele mods:titleInfo s. <https://wiki.deutsche-digitale-bibliothek.de/display/DFD/titleInfo>

2 Person

| Element/Unterelement | Wiederholbar | Bemerkungen/Erläuterungen | Status |
|--|--------------|--|-----------------------------|
| <name type="personal" valueURI="..."> | ja | Person, die mit dem Werk in Beziehung steht (z.B. der Autor). Das Attribut @valueURI enthält den URI der Person bzw. des Personennamens. Handelt es sich bei dem @valueURI nicht um einen http-URI, muss @authority den Code für das Regelwerk enthalten | Pflicht, wenn vh. |
| <mods:part> | ja | Enthält Namensbestandteile, deren Art in @type festgelegt ist. Mögliche Werte sind „date“, „family“, „given“, „termsOfAddress“ | |
| <name> / <displayForm> | nein | Name in der gewünschten Anzeigeform | Pflicht, wenn mods:name vh. |
| <name> / <role> | ja | Rahmenelement zur Funktion der Person | Pflicht, wenn mods:name vh. |
| <name> / <role> / <roleTerm type="code" authority="marcrelator"> | nein | Funktion der Person; der Wert in mods:roleTerm muss ein Code aus dem MARC RelatorVokabular sein (https://id.loc.gov/vocabulary/relators.html) | Pflicht, wenn mods:name vh. |

Beispiele, s. <https://wiki.deutsche-digitale-bibliothek.de/display/DFD/name>

3 Körperschaft

| Element/Unterelement | Wiederholbar | Bemerkungen/Erläuterungen | Status |
|---|--------------|--|-----------------------------|
| <code><name type="corporate" authority="..."></code> | ja | Körperschaft, die mit dem Werk in Beziehung steht. | Pflicht, wenn vh. |
| <code><name> / <namePart></code> | ja | | |
| <code><name> / <displayForm></code> | nein | s.o. | Pflicht, wenn mods:name vh. |
| <code><name> / <role></code> | nein | s.o. | Pflicht, wenn mods:name vh. |
| <code><name> / <role> /</code> | nein | s.o. | |
| <code><roleTerm type="code" authority="marcrelator"></code> | | | |

Beispiele mods:name s. <https://wiki.deutsche-digitale-bibliothek.de/display/DFD/name>

4 Erscheinungsvermerk

| Element/Unterelement | Wiederholbar | Bemerkungen/Erläuterungen | Status |
|---|---|---|-------------------|
| <code><originInfo eventType="publication"></code> | ja | Der erste <code><originInfo></code> -Block ist für Angaben zur Quelle vorgesehen, Angaben zur digitalen Edition können in einem zweiten <code><originInfo></code> Block mit dem <code>eventType="digitisation"</code> gemacht werden. | Pflicht, wenn vh. |
| <code><originInfo>/<place>/<placeTerm type="text"></code> | | Enthält den Namen des Erscheinungsorts. | |
| <code><originInfo> / <publisher></code> | ja | Enthält den Namen des Verlags bzw. des Druckers. | Pflicht, wenn vh. |
| <code><originInfo> / <dateIssued encoding="8601"></code> | nur, wenn Attribut <code>@point</code> verwendet wird | Enthält das Erscheinungsjahr der Quelle. | Pflicht, wenn vh. |

Beispiele mods:originInfo s. <https://wiki.deutsche-digitale-bibliothek.de/display/DFD/originInfo>

5 Ausgabevermerk

| Element/Unterelement | Wiederholbar | Bemerkungen/Erläuterungen | Status |
|---|--------------|--|-------------------|
| <code><originInfo> / <edition></code> | ja | Enthält die Ausgabebezeichnung (s.a. oben bzgl. Angaben zur elektronischen Ausgabe). | Pflicht, wenn vh. |

6 Kollationsvermerk/physische Beschreibung

| Element/Unterelement | Wiederholbar | Bemerkungen/Erläuterungen | Status |
|-------------------------------------|--------------|---------------------------|---------|
| <physicalDescription> | nein | Kollationsvermerk | Pflicht |
| <physicalDescription> / <extent> | | | |

Beispiele mods:physicalDescription

s. <https://wiki.deutsche-digitale-bibliothek.de/display/DFD/physicalDescription>

7 Übergeordnetes Werk

| Element/Unterelement | Wiederholbar | Bemerkungen/Erläuterungen | Status |
|---|--------------|---|--------------------------|
| <relatedItem type="host"> / <recordInfo> / <recordIdentifier @source> | nein | <recordIdentifier> innerhalb von mods:relatedItem type="host" ist ein Identifier, der zur Verknüpfung mit hierarchisch übergeordneten Datensätzen dient. Das Attribut @source muss Informationen zu Herkunft des Datensatzes enthalten und ist verpflichtend. | Pflicht, wenn Hierarchie |

Beispiele mods:relatedItem s. <https://wiki.deutsche-digitale-bibliothek.de/display/DFD/relatedItem>

8 Bandangabe

| Element/Unterelement | Wiederholbar | Bemerkungen/Erläuterungen | Status |
|--|--------------|---|-------------------|
| <part order="..."> / <detail> / <number> | nein | Angaben zu Teilen; Wert des Attributs @order ist ein beliebiger numerischer Wert (integer), der die Reihenfolge der Teile sicherstellt. <number> enthält die Zählung von Bänden, Heften u.a. | Pflicht, wenn vh. |

Beispiele mods:part s. <https://wiki.deutsche-digitale-bibliothek.de/display/DFD/part>

9 Sprache

| Element/Unterelement | Wiederholbar | Bemerkungen/Erläuterungen | Status |
|--|--------------|--|---------|
| <language> | ja | Sprachangaben | Pflicht |
| <language> / <languageTerm type="code" authority="iso639-2b"> | nein | Enthält die Sprache des Werks in gemäß ISO 639-2/B kodierter Form. | |

Beispiele mods:language s. <https://wiki.deutsche-digitale-bibliothek.de/display/DFD/language>

10 Zitierfähige Identifizierung

| Element/Unterelement | Wiederholbar | Bemerkungen/Erläuterungen | Status |
|---------------------------|--------------|---|---------|
| <identifizier type="..."> | ja | Weltweit eindeutige Identifizierung der Ressource (Attribut @type z.B. urn, purl, doi, handle, uri etc.). Angegeben werden muss auch, wenn vorhanden, die GW und VD-Nummer des Drucks (Attribute @type VD16, VD17, GW). | Pflicht |

Beispiele mods:identifizier s. <https://wiki.deutsche-digitale-bibliothek.de/display/DFD/identifizier>

11 Datenbank-ID

| Element/Unterelement | Wiederholbar | Bemerkungen/Erläuterungen | Status |
|---|--------------|--|---------|
| <recordInfo> / <recordIdentifizier @source=> | nein | Datensatz-Identifizier zur eindeutigen Identifizierung innerhalb eines Datenbanksystems, z.B. Pica Produktionsnummer. Das Attribut @source muss Informationen zu Herkunft des Datensatzes enthalten und ist verpflichtend. | Pflicht |

Beispiele mods:recordInfo s. <https://wiki.deutsche-digitale-bibliothek.de/display/DFD/recordInfo>

12 Signatur

| Element/Unterelement | Wiederholbar | Bemerkungen/Erläuterungen | Status |
|----------------------|--------------|-------------------------------------|-----------|
| <location> | ja | Standort und Signatur des Originals | Empfohlen |
| <physicalLocation> | nein | | |
| <shelfLocator> | ja | | |

Beispiele mods:location s. <https://wiki.deutsche-digitale-bibliothek.de/display/DFD/location>

Je nach Materialart und Projektzuschnitt ist eine Erweiterung dieses Basissets zu erwägen. Für diesen Zweck stehen auf der Seite des DFG-Viewers differenziertere Formatdefinitionen zur Verfügung. Diese enthalten auch weitere Erläuterungen zur Besetzung der jeweiligen Felder.

Die Anforderungen an METS/MODS-Daten für die Lieferung der Daten an die DDB sind im DDBInfo-Wiki definiert.¹¹² Hier findet sich auch ein Link auf die Schematron-Validierung der DDB.¹¹³

¹¹² <https://wiki.deutsche-digitale-bibliothek.de/pages/viewpage.action?pagelid=19006651>.

¹¹³ <https://wiki.deutsche-digitale-bibliothek.de/display/DFD/Schematron-Validierungen+der+Fachstelle+Bibliothek#SchematronValidierungenderFachstelleBibliothek-METS/MODS-Anwendungsprofilf%C3%BCdigitalisierteMedien>.

3. Hinweise für Archive zur Konversion von EAD(DDB) nach MODS

Das Mapping von EAD(DDB) auf entsprechende MODS-Elemente ist grundsätzlich leicht durchführbar. Für die vereinfachte Bearbeitung von archivischen Datensätzen stehen in der archivischen Community auch Tools zur Verfügung, die von interessierten Einrichtungen genutzt werden können.¹¹⁴

| Feldbezeichnung | EAD(DDB) | MODS | Status |
|-----------------|---|--|--------------------|
| Titel | c/did/unittitle | mods:titleInfo / mods:title und gegebenenfalls weitere Unterelemente | Pflicht |
| | | MODS-Definition: A word, phrase, character, or group of characters, normally appearing in a resource, that names it or the work contained in it. | |
| | | Kommentar: Für das Wurzelstrukturelement in einem METS/MODS-Satz ist die Angabe mindestens eines Titels unter Verwendung des Elements mods:titleInfo mit dem Unterelement mods:title verpflichtend. | |
| Institution | ancestor::arch-desc/did/repository/corpname | mods:location / mods:physicalLocation MODS-Definition: The institution or repository that holds the resource or where it is available. | Pflicht |
| Signatur | c/did/unitid[not(@type)] | mods:location / mods:shelfLocator MODS-Definition: Shelfmark or other shelving designation that indicates the location identifier for a copy. | Pflicht |
| Laufzeit | c/did/unit-date | mods:originInfo / mods:dateCreated MODS-Definition: The date of creation of the resource Kommentar: Datum der Entstehung bzw. der Laufzeit der analogen Vorlage. | Pflicht, wenn v.h. |

¹¹⁴ Die DDB-Fachstelle Archiv stellt hierfür das Data Preparation Tool zur Verfügung: <https://wiki.deutsche-digitale-bibliothek.de/display/DFD/DDB+Data+Preparation+Tool>, Aufruf 25.10.2022. Die Anwendung wird von der Fachstelle weiterentwickelt und gepflegt.

| Feldbezeichnung | EAD(DDB) | MODS | Status |
|------------------------|-------------------------------------|---|-----------|
| Kontext ¹¹⁵ | ances- tor::c/did/unit -title | mods:relatedItem type="host" / mods:title gegebenenfalls mit weiteren Unterele- menten MODS-Definition: Information that identifies other re- sources related to the one being de- scribed. Über mods:relatedItem lassen sich nicht nur hierarchische Beziehungen abbilden. Vielmehr lassen sich mittels des @type-Attributs eine Vielzahl von Beziehungen darstellen. Um die Beziehung zwischen Dokumen- ten oder Dokumentteilen mit MODS herzustellen, müssen eindeutige Identi- fier verwendet werden. | Empfohlen |
| ID | c/@id | mods:recordInfo / mods:recordIdenti- fier MODS-Definition: Contains the system control number assigned by the organi- zation creating, using, or distributing the record. | Pflicht |

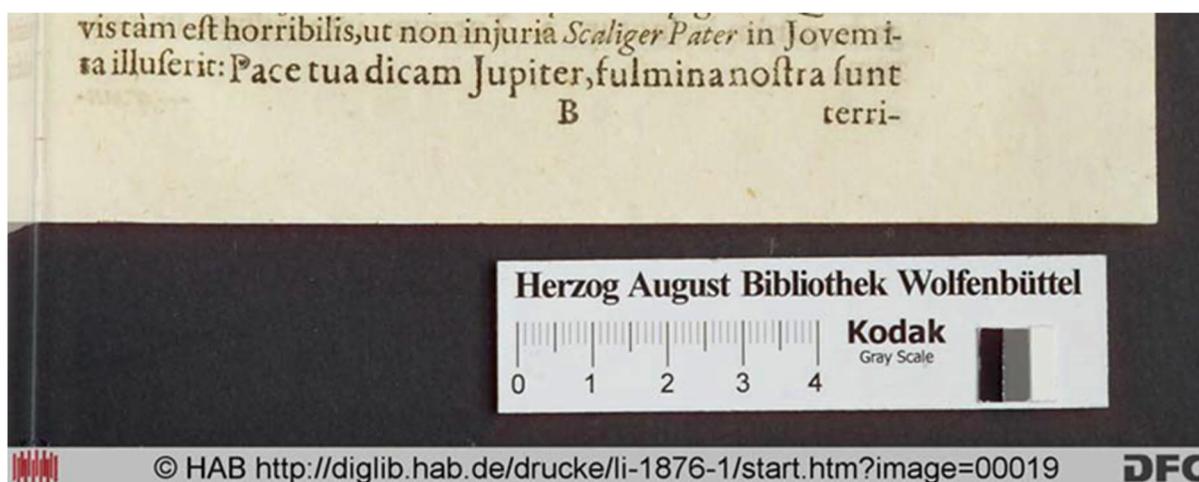
4. Beispiele für DFG-Viewer-kompatible METS/MODS Datensätze

- (1) <https://wiki.deutsche-digitale-bibliothek.de/pages/viewpage.action?pagelId=19006655>
- (2) <https://wiki.deutsche-digitale-bibliothek.de/display/DFD/5.+Beispiele+und+Schemata+zur+Validierung#id-5.BeispieleundSchematazurValidierung-Beispieldateienf%C3%BCrdasarchivischeMETS/MODS-Profil>

5. Hinzurechnen einer Fußzeile

Die Fußzeile sollte bei den an die Nutzer ausgelieferten bzw. im Internet angebotenen Formaten am unteren Rand des Bilddigitalisats hinzugerechnet werden. Links sollte sich das Institutionslogo, bei Förderung durch die DFG nach Möglichkeit rechts zusätzlich das Logo der DFG befinden. Die Angabe einer zitierfähigen URL im Mittelbereich (ergänzend zur Übergabe in der Viewer-XML-Datei) wird empfohlen. Schrift und Logos müssen der Auflösung entsprechend skaliert werden. In PDF-Dateien zum Download können – zusätzlich zu einem Vorblatt – ebenfalls Fußzeilen an jedes Bild angefügt werden. Vergleiche die untenstehenden Beispiele: Das Mapping von EAD(DDB) auf entsprechende MODS-Elemente ist grundsätzlich leicht durchführbar.

¹¹⁵ Dies bezieht sich auf den Bestandskontext der Ressource, also die darüber liegenden Verzeichnungsstufen aus Tektonik und Klassifikation.



Anhang B: METS/TEI-Spezifikation für die Darstellung von digitalisierten Handschriften

1. DFG-Viewer

Ziel der vorliegenden METS/TEI-Spezifikation für digitalisierte mittelalterliche und frühneuzeitliche Handschriften ist die Darstellung der Digitalisate unter der vereinheitlichten Oberfläche des DFG-Viewers. Der ursprünglich für die Präsentation von digitalisierten Druckwerken geschaffene DFG-Viewer wird somit medientypologisch erweitert, ohne dabei die gewohnte Bedienoberfläche aufzugeben. Diese wird lediglich um wenige Funktionen ergänzt, um den spezifischen Anforderungen des neuen Medientyps gerecht zu werden.

Da es sich nicht um eine grundsätzlich neue Implementierung des DFG-Viewers handelt, sondern lediglich um die Unterstützung eines zusätzlichen Anwendungsprofils für METS/TEI, gelten die in Anhang A gemachten Angaben zu Druckwerken analog auch für digitalisierte Handschriften – insbesondere die allgemeinen Informationen zum DFG-Viewer in Abschnitt 1 sowie die Hinweise zum Hinzurechnen einer Fußzeile in Abschnitt 5.

Weitere Informationen zu den vom DFG-Viewer unterstützten Datenformaten und Funktionen sowie zur Einbindung des Webdienstes in das eigene Angebot finden Sie auf der Website des DFG-Viewers.¹¹⁶

2. TEI-DFG-Standard-Set (Handschriftenbestände)

Die folgenden Angaben zur Kodierung einer digitalisierten Handschrift in TEI beschränken sich auf die nötigen Pflichtfelder und stellen nicht das vollständige Spektrum aller Möglichkeiten dar. Eine ausführliche Dokumentation aller Optionen findet sich auf der Webseite des DFG-Viewers. Grundlagen des Anwendungsprofils sind die TEI-Spezifikation der Text Encoding Initiative¹¹⁷ sowie die Vorarbeiten des Europeana Regia-Projekts¹¹⁸ zu denen das DFG-Standard-Set kompatibel ist.

1 Angaben zur Identifikation

Alle in diesem Abschnitt beschriebenen Elemente sind Unterelemente von `teiHeader/fileDesc/sourceDesc/msDesc`

| Element/Unterelement | Wiederholbar | Bemerkungen/Erläuterungen | Status |
|--|--------------|---|-------------------|
| <code><head> / <title></code> | nein | Enthält den Titel der Handschrift nach DFG-Richtlinien zur Handschriftenkatalogisierung. | Pflicht, wenn vh. |
| <code><msIdentifier> / <settlement></code> | nein | Enthält den geographischen Aufbewahrungsort der Handschrift. | Pflicht |
| <code><msIdentifier> / <repository></code> | nein | Enthält den Namen der die Handschrift aufbewahrenden Einrichtung. | Pflicht |
| <code><msIdentifier> / <idno></code> | nein | Enthält einen innerhalb der aufbewahrenden Einrichtung eindeutigen Identifikator, in der Regel die Signatur ggf. ergänzt um eine Angabe der Blätter, über die sich die logische Struktur erstreckt. | Pflicht |

¹¹⁶ <https://dfg-viewer.de/>.

¹¹⁷ <https://tei-c.org/>.

¹¹⁸ <http://www.europeanaregia.eu/>.

| Element/Unterelement | Wiederholbar | Bemerkungen/Erläuterungen | Status |
|---------------------------|--------------|--|-----------|
| <msIdentifier> / <msName> | ja | Enthält einen nicht-kanonischen Titel der Handschrift. | Empfohlen |

2 Angaben zur Entstehung

Alle in diesem Abschnitt beschriebenen Elemente sind Unterelemente von `teiHeader/fileDesc/sourceDesc/msDesc/history`

| Element/Unterelement | Wiederholbar | Bemerkungen/Erläuterungen | Status |
|------------------------|--------------|--|-----------|
| <origin> / <origPlace> | nein | Enthält den Entstehungsort der Handschrift wie er im DFG-Viewer erscheinen soll. | Empfohlen |
| <origin> / <origDate> | nein | Enthält Angaben zur Entstehungszeit der Handschrift wie sie im DFG-Viewer angezeigt werden sollen. Exakte numerische Angaben werden in den Attributen <code>@when</code> , <code>@notBefore</code> , <code>@notAfter</code> , <code>@from</code> und <code>@to</code> gemacht, wenn möglich. | Empfohlen |

3 Angaben zur physischen Beschaffenheit

Alle in diesem Abschnitt beschriebenen Elemente sind Unterelemente von `teiHeader/fileDesc/sourceDesc/msDesc/physDesc/objectDesc`

| Element/Unterelement | Wiederholbar | Bemerkungen/Erläuterungen | Status |
|---|--------------|---|-------------------|
| <supportDesc material="..."> | | | Pflicht, wenn vh. |
| <supportDesc> / <support> | nein | Enthält Angaben zum Beschreibstoff wie sie im DFG-Viewer angezeigt werden sollen. | Pflicht, wenn vh. |
| <supportDesc> / <extent> / <measure> | nein | Enthält Mengenangaben zum Umfang wie z.B. die Blattzahl, die auch im DFG-Viewer angezeigt wird. | Pflicht, wenn vh. |
| <supportDesc> / <extent> / <dimensions> | nein | Enthält Angaben zur physischen Abmessung wie etwa das Blattformat, das auch im DFG-Viewer angezeigt wird. Die konkreten Größenangaben erfolgen in den Unterelementen <code>depth</code> , <code>height</code> sowie <code>width</code> . | Pflicht, wenn vh. |

4 Angaben zum Inhalt

Alle in diesem Abschnitt beschriebenen Elemente sind Unterelemente von `teiHeader/fileDesc/sourceDesc/msDesc/msContent`

| Element/Unterelement | Wiederholbar | Bemerkungen/Erläuterungen | Status |
|---|--------------|--|-----------|
| <code><summary></code> | nein | Enthält eine summarische Inhaltsbeschreibung. | Empfohlen |
| <code><textLang mainLang="..." otherLangs="..."></code> | nein | Enthält Angaben zu den Sprachen, in denen die Texte der Handschrift verfasst sind. Das Attribut <code>@mainLang</code> sollte die überwiegend verwendete Sprache benennen, während alle weiteren vorkommenden Sprachen jeweils mit Leerzeichen getrennt im Attribut <code>@otherLangs</code> kodiert werden sollten. Die Attributwerte müssen Sprachcodes nach den Standards ISO 639-2 oder ISO 639-3 sein. | Empfohlen |

5 Administrative Angaben zum Digitalisat

Alle in diesem Abschnitt beschriebenen Elemente sind Unterelemente von `teiHeader/fileDesc`

| Element/Unterelement | Wiederholbar | Bemerkungen/Erläuterungen | Status |
|--|--------------|---|-------------------|
| <code><titleStmt> / <title></code> | nein | Enthält einen Titel für das Digitalisat. In der Regel ist dies die Schlagzeile nach DFG-Richtlinien bestehend aus Signatur der digitalisierten Handschrift und ihrem Titel. Dieses Feld ist nicht mit dem Titel zu verwechseln und kommt im DFG-Viewer nicht zur Anzeige. Die Verpflichtung ergibt sich aus der allgemeinen TEI-Spezifikation und ist nicht Viewer-spezifisch. | |
| <code><titleStmt> / <funder></code> | ja | Enthält einen für die Digitalisierung relevanten Geldgeber, also etwa „Deutsche Forschungsgemeinschaft“. Das kann aber z.B. auch die digitalisierende Institution sein. | Empfohlen |
| <code><publicationStmt> / <publisher></code> | ja | Enthält den Herausgeber des Digitalisats, in aller Regel also die digitalisierende Einrichtung. Mindestens ein Unterelement von <code>publicationStmt</code> ist verpflichtend. | Pflicht, wenn vh. |
| <code><publicationStmt> / <pubPlace> / <ptr target="..." cRef="..." type="..." /></code> | ja | Enthält einen persistenten Identifier des Digitalisats. Handelt es sich beim Identifier um eine URL, so ist diese im Attribut <code>@target</code> anzugeben, andernfalls im Attribut <code>@cRef</code> . Darüberhinaus muss die Art des Identifiers im Attribut <code>@type</code> mit den Werten <code>urn</code> oder <code>purl</code> angegeben werden. Mindestens ein Unterelement von <code>publicationStmt</code> ist verpflichtend. | Pflicht, wenn vh. |
| <code><publicationStmt> / <idno></code> | ja | Enthält einen Identifier des Datensatzes. Optional kann dessen Typ im Attribut <code>@type</code> angegeben werden. | Pflicht, wenn vh. |

3. Beispiel für einen METS/TEI Datensatz nach dem DFG-Standard

Das folgende Beispiel zeigt den METS/TEI-Datensatz aufgrund seines Umfangs nur ausschnittsweise. Der Auszug stellt einen vollständigen deskriptiven Metadatensatz dar, der alle oben beschriebenen Pflicht- und Optionalfelder exemplarisch enthält. Dieser Auszug befindet sich in einer METS-Datei innerhalb eines dmdSec-Abschnitts.

```

<tei:teiHeader>
  <tei:fileDesc>
    <tei:titleStmt>
      <!--Fingierter Titel für das Digitalisat (obligatorisch)
      Es handelt sich hierbei nicht um den bibliografischen Titel der Handschrift, sondern um die nach DFG-
      Richtlinien gebildete Schlagzeile bestehend aus Signatur und wissenschaftlichem Titel
      Der Titel kommt im DFG-Viewer nicht zur Anzeige, ist aber aufgrund der TEI Spezifikation dennoch ver-
      pflichtend. -->
      <tei:title>
        Zeugbuch Kaiser Maximilians I. BSB Cod.icon. 222
      </tei:title>
      <!--Förderer der Digitalisierung (optional) -->
      <tei:funder>Deutsche Forschungsgemeinschaft</tei:funder>
    </tei:titleStmt>
    <tei:publicationStmt>
      <!--Herausgeber des Digitalisats (obligatorisch, sofern bekannt) -->
      <tei:publisher>Bayerische Staatsbibliothek</tei:publisher>
      <!--Persistenter Identifikator für das Digitalisat (obligatorisch)-->
      <tei:pubPlace>
        <tei:ptr
          target="http://daten.digitale-sammlungen.de/~db/0002/bsb00020956/"
          type="purl"/>
        </tei:pubPlace>
      <!--Identifikator für den Datensatz (obligatorisch) -->
      <tei:idno type="zend">bsb00020956</tei:idno>
      <!--Lizenzangabe-->
      <tei:availability>
        <tei:licence
          target="http://creativecommons.org/publicdomain/zero/1.0/">Public domain (CC0)
        </tei:licence>119
      </tei:availability>
    </tei:publicationStmt>
    <tei:sourceDesc>
      <tei:msDesc>
        <tei:msIdentifier>
          <!--Geographischer Aufbewahrungsorts (obligatorisch) -->
          <tei:settlement>München</tei:settlement>
          <!--Name der bewahrenden Institution (obligatorisch) -->
          <tei:repository>Bayerische Staatsbibliothek</tei:repository>
          <!--Signatur o. eindeutiger Identifikator (obligatorisch) -->
          <tei:idno>Cod.icon. 222</tei:idno>
          <!--Nicht-kanonischer Titel (optional, wiederholbar) -->
          <tei:msName>Zeugbuch Kaiser Maximilians I.</tei:msName>
        </tei:msIdentifier>
        <tei:head>
          <!--Handschriftentitel (obligatorisch, sofern vorhanden) -->
          <tei:title/>
        </tei:head>
        <tei:history>
          <tei:origin>
            <!--Entstehungsort (empfohlen, sofern bekannt) -->

```

¹¹⁹ Wenn für den DFG Viewer nötig, muss dieselbe Angabe noch einmal in dv:licence ergänzt werden.

```

<tei:origPlace>Innsbruck</tei:origPlace>
<!--Entstehungszeit (empfohlen, sofern bekannt) -->
<tei:origDate notBefore="1500" notAfter="1505">
  um 1502
</tei:origDate>
</tei:origin>
</tei:history>
<tei:physDesc>
<tei:objectDesc>
  <!--Angabe des Materials (obligatorisch, sofern ermittelbar) -->
  <tei:supportDesc material="mixed">
    <!--Angaben zum Beschreibstoff (obligatorisch, sofern
      ermittelbar) -->
    <tei:support>
      Pergament (2-8, 11-30, 34, 48, 295, 296) und Papier
    </tei:support>
    <tei:extent>
      <!--Angaben zum Umfang wie z.B. die Blattzahl
        (obligatorisch, sofern ermittelbar) -->
      <tei:measure type="leavesCount">2 + 276 Blätter</tei:measure>
      <!--Blattformat (obligatorisch, sofern ermittelbar) -->
      <tei:dimensions unit="mm" type="leaves">
        <tei:height>420</tei:height>
        <tei:width>285</tei:width>
      </tei:dimensions>
    </tei:extent>
  </tei:supportDesc>
</tei:objectDesc>
</tei:physDesc>
<tei:msContents>
  <!--Summarische Beschreibung des Inhalts (empfohlen) -->
  <tei:summary/>
  <!--Textsprache (empfohlen, sofern bekannt) -->
  <tei:textLang mainLang="ger">Deutsch</tei:textLang>
</tei:msContents>
</tei:msDesc>
</tei:sourceDesc>
</tei:fileDesc>
</tei:teiHeader>

```

Der übrige Aufbau der METS-Datei folgt analog der Dokumentation für Drucke. Innerhalb des structMap-Elements mit dem Attribut TYPE=LOGICAL wird in der METS-Datei die logische Struktur der Handschrift kodiert. Jede logische Struktureinheit (z.B. ein Textfragment) kann wiederum auf einen eigenen Abschnitt mit deskriptiven Metadaten verweisen, der ebenfalls dem oben beschriebenen Aufbau folgen muss.

Anhang C: LIDO-Kernelemente für die Publikation

1. LIDO-Kernelemente für bildhafte und dreidimensionale Objekte

Der LIDO-Standard ermöglicht die Weitergabe und Bereitstellung von deskriptiven und administrativen Metadaten für die Recherche und Präsentation von in der Regel unikal, bildhaften und dreidimensionalen Objekten in Online-Umgebungen und unterstützt die gemeinsame Nutzung und Verknüpfung von Daten im Internet. Nähere Informationen zu LIDO finden Sie auf der LIDO-Website unter <http://www.lido-schema.org>.

LIDO definiert nur wenige Pflichtelemente. Die unten folgende Tabelle enthält neben diesen rot markierten LIDO-Pflichtelementen weitere für die Objektdokumentation in Digitalisierungsprojekten sinnvolle LIDO-Elemente. Für Elemente mit der Kennzeichnung „Pflicht, wenn vh.“ kann es ggf. Gründe geben, warum sie nicht belegt werden können, z. B. weil die Angaben für die jeweilige Objektgattung nicht sinnvoll sind. Hier können für den konkreten Anwendungsfall Anpassungen der Vorgaben vorgenommen werden. Zur Anwendung von LIDO kann je nach Projektkontext auf material- bzw. fachspezifische LIDO-Profile zur Orientierung zurückgegriffen werden: Zur Erfassung und Publikation von Metadaten zu Zeichnungen und druckgraphischen Werken ist das LIDO-Handbuch Graphik¹²⁰ heranzuziehen, für Gemälde und Skulpturen das LIDO-Handbuch Malerei und Skulptur¹²¹. Weitere Informationen zu LIDO-Profilen werden über das Themenportal „LIDO für kulturelle Objekte“¹²² und die LIDO Website¹²³ publiziert. LIDO-Profile können neben den allgemeinen LIDO-Pflichtelementen weitere profilspezifische Pflichtelemente enthalten und geben Informationen zu weiteren empfohlenen Elementen.

Für weitere Informationen zu den unten aufgeführten und weiteren LIDO-Elementen sei auf die Dokumentation der jeweils aktuellen LIDO-Version¹²⁴ verwiesen. Dort ist zu allen LIDO-Elementen eine Beschreibung nach folgendem Muster abrufbar: <https://lido-schema.org/schema/latest/lido.html#objectWorkType>

Zahlreiche LIDO-Elemente (z.B. Objekttyp, Material, Technik) werden mit Datenwerten belegt, die als Begriffe („units of thought“¹²⁵) in kontrollierten Vokabularen (Thesauri, Klassifikationssystem, Wortlisten)¹²⁶ definiert sein sollen. Solche Elemente werden im Folgenden in der jeweiligen Erläuterung als Begriff markiert. Diese LIDO-Elemente haben zur Referenzierung des kontrollierten Vokabulars die in <http://lido-schema.org/schema/latest/lido.html#conceptComplexType> definierten Unterelemente. Manche LIDO-Elemente (z.B. Art der Zuschreibung) können vorzugsweise als Begriff, alternativ aber auch mit Freitext belegt werden, diese folgen der Definition in <http://lido-schema.org/schema/latest/lido.html#conceptMixedComplexType>. Das Dokument „LIDO Terminology Recommendation“¹²⁷ gibt vielfältige Empfehlungen zum Einsatz der LIDO Terminologie¹²⁸ und zum Einsatz weiterer Vokabulare.

Empfehlungen aus dem LIDO Terminology Recommendation Document zum jeweiligen Element sind (so für dieses Element vorhanden) nach folgendem Muster abrufbar:

<https://lido-schema.org/documents/terminology-recommendation.html#objectWorkType>

¹²⁰ <https://doi.org/10.11588/arthistoricum.382.544>.

¹²¹ <https://doi.org/10.11588/arthistoricum.1026>.

¹²² <https://www.arthistoricum.net/themen/portale/lido>.

¹²³ <http://www.lido-schema.org>.

¹²⁴ <https://lido-schema.org/schema/latest/lido.html>.

¹²⁵ <https://www.w3.org/TR/2009/REC-skos-reference-20090818/#concepts>.

¹²⁶ Dies können (vorzugsweise) publizierte, aber auch unpublizierte, lokal gepflegte Vokabulare sein.

¹²⁷ <http://lido-schema.org/documents/terminology-recommendation.html>.

¹²⁸ <http://terminology-view.lido-schema.org/>.

1 Identifikatoren für LIDO Datensatz und Objekt

| Element/Unterelement | Wiederholbar | Bemerkungen/Erläuterungen | Status |
|----------------------|--------------|---|-------------------|
| <lidoReclID> | ja | https://lido-schema.org/schema/latest/lido.html#lidoReclID Eindeutige Identifikationsnummer für den LIDO Datensatz, typischerweise zusammengesetzt aus der ISIL-Nummer des Datenlieferanten sowie der lokalen Datensatznummer. (siehe https://sigel.staatsbibliothek-berlin.de/ , zu ISIL-Nummern für Museen siehe https://isil.museum/) | Pflicht |
| <objectPublishedID> | ja | https://lido-schema.org/schema/latest/lido.html#objectPublishedID Ein publizierter Identifikator des Objekts oder Werks, möglichst in Form eines dereferenzierbaren URIs. | Pflicht, wenn vh. |

Deskriptive Metadaten

2 Objekttypen

| Element/Unterelement | Wiederholbar | Bemerkungen/Erläuterungen | Status |
|----------------------|--------------|--|---------|
| <objectWorkType> | ja | https://lido-schema.org/schema/latest/lido.html#objectWorkType Objekt- oder Werktyp Begriff | Pflicht |

3 Klassifikation

| Element/Unterelement | Wiederholbar | Bemerkungen/Erläuterungen | Status |
|-----------------------------|--------------|---|-----------|
| <classification type="..."> | ja | https://lido-schema.org/schema/latest/lido.html#classification Ein Begriff, der ein Objekt in einem größeren Kontext kategorisiert. Das (optionale) @type-Attribut qualifiziert die Art der Klassifikation, mögliche Werte s. http://terminology.lido-schema.org/classification_type | Empfohlen |

4 Titel/Objektname

| Element/Unterelement | Wiederholbar | Bemerkungen/Erläuterungen | Status |
|------------------------------------|-------------------------|--|---------|
| <titleSet> | ja | https://lido-schema.org/schema/latest/lido.html#titleSet Umschlag für Titelinformationen | Pflicht |
| <titleSet /> <appellationValue> | nur für Sprachvarianten | https://lido-schema.org/schema/latest/lido.html#appellationValue Titel oder Name, der einem Objekt gegeben wurde. Sofern das Objekt keinen Titel hat, muss er generiert werden, z.B. aus dem Objekttyp und einem Identifikator, etwa der Inventarnummer. | Pflicht |

5 Bewahrungsort/Standort

| Element/Unterelement | Wiederholbar | Bemerkungen/Erläuterungen | Status |
|--|-------------------------|---|-------------------|
| <repositorySet type="http://terminology.lido-schema.org/lido00475"> | ja | https://lido-schema.org/schema/latest/lido.html#repositorySet Umschlag für Informationen zum aktuellen Bewahrungsort/Standort; kann mit dem type-Attribut="http://terminology.lido-schema.org/lido00476" für Angaben zu ehemaligen Bewahrungsorten/Standorten versehen werden. | Pflicht, wenn vh. |
| <repositorySet> / <repositoryName> / <legalBodyName> / <appellationValue> | nur für Sprachvarianten | https://lido-schema.org/schema/latest/lido.html#repositoryName Bezeichnung für die aufbewahrende Institution | Pflicht, wenn vh. |
| <repositorySet> / <workID> | ja | https://lido-schema.org/schema/latest/lido.html#workID Eine eindeutige numerische oder alphanumerische Identifikationsnummer, die die aufbewahrende Institution dem Objekt gegeben hat; typischerweise die Inventarnummer | Pflicht, wenn vh. |
| <repositorySet> / <repositoryLocation> / <placeID> | ja | https://lido-schema.org/schema/latest/lido.html#placeID Identifikator für den Ort. Referenziert idealerweise eine Normdatei. | Empfohlen |
| <repositorySet> / <repositoryLocation> / <namePlaceSet> / <appellationValue> | nur für Sprachvarianten | https://lido-schema.org/schema/latest/lido.html#repositoryLocation Bezeichnung des Standorts (insbes. für Architektur Pflichtangabe) | Pflicht, wenn vh. |

6 Objektbeschreibung

| Element/Unterelement | Wiederholbar | Bemerkungen/Erläuterungen | Status |
|--|-------------------------|---|-----------|
| <objectDescriptionSet type="..."> | ja | https://lido-schema.org/schema/latest/lido.html#objectDescriptionSet Umschlag für Beschreibungstexte und ihre Quellen. Das (optionale) type-Attribut qualifiziert die Art der Beschreibung, mögliche Werte s. http://terminology.lido-schema.org/objectDescriptionSet_type | Empfohlen |
| <objectDescriptionSet> / <descriptiveNoteValue> | nur für Sprachvarianten | https://lido-schema.org/schema/latest/lido.html#descriptiveNoteValue | Empfohlen |
| <objectDescriptionSet> / <objectDescriptionRights> | ja | https://lido-schema.org/schema/latest/lido.html#objectDescriptionRights Rechte am Beschreibungstext | Empfohlen |

7 Maßangaben

| Element/Unterelement | Wiederholbar | Bemerkungen/Erläuterungen | Status |
|--|-------------------------|---|-------------------|
| <objectMeasurementsSet> | ja | https://lido-schema.org/schema/latest/lido.html#objectMeasurementsSet Umschlag für Maßangaben Die Angabe von Maßangaben ist auch ereignisbezogen in <eventSet> / <event> / <eventObjectMeasurements> möglich. s. https://lido-schema.org/schema/latest/lido.html#eventObjectMeasurements | Pflicht, wenn vh. |
| <objectMeasurementsSet> / <displayObjectMeasurements> | nur für Sprachvarianten | https://lido-schema.org/schema/latest/lido.html#displayObjectMeasurements Maßangabe in Textform | Pflicht, wenn vh. |
| <objectMeasurementsSet> / <objectMeasurements> / <measurementsSet> | ja | https://lido-schema.org/schema/latest/lido.html#measurementsSet Maßangabe in strukturierter Form mit den Unterelementen <measurementType>, <measurementUnit>, <measurementValue> | Empfohlen |

8 Ereignis

| Element/Unterelement | Wiederholbar | Bemerkungen/Erläuterungen | Status |
|--|-------------------------|---|-------------------|
| <eventSet> | ja | https://lido-schema.org/schema/latest/lido.html#eventSet Umschlag für Informationen über ein Ereignis im Lebenszyklus des Objekts | Pflicht, wenn vh. |
| <eventSet> / <event> / <eventType> | nein | https://lido-schema.org/schema/latest/lido.html#eventType Ereignistyp Je nach Objektart sind typische Ereignisse: Herstellung, Fund, Nutzung. Für empfohlene Ereignistypen siehe http://terminology.lido-schema.org/event-Type Begriff | Pflicht, wenn vh. |
| Am Ereignis beteiligte identifizierbare Akteure | | | |
| <eventSet> / <event> / <eventActor> | ja | https://lido-schema.org/schema/latest/lido.html#eventActor Umschlag für Information zur Beteiligung einer Person, Institution, Gruppe, Familie am Ereignis | Pflicht, wenn vh. |
| <eventSet> / <event> / <eventActor> / <actorInRole> / <actor> / <actorID> | ja | https://lido-schema.org/schema/latest/lido.html#actorID Identifikator für den beteiligten Akteur. Referenziert idealerweise eine Normdatei. Für DFG-Projekte soll die GND verwendet werden. | Pflicht, wenn vh. |
| <eventSet> / <event> / <eventActor> / <actorInRole> / <actor> / <nameActorSet> / <appellationValue> | nur für Namensvarianten | https://lido-schema.org/schema/latest/lido.html#actorInRole https://lido-schema.org/schema/latest/lido.html#appellationValue Name des beteiligten Akteurs | Pflicht, wenn vh. |
| <eventSet> / <event> / <eventActor> / <actorInRole> / <roleActor> | ja | https://lido-schema.org/schema/latest/lido.html#roleActor Rolle des beteiligten Akteurs Begriff | Empfohlen |

| Element/Unterelement | Wiederholbar | Bemerkungen/Erläuterungen | Status |
|--|-------------------------|---|-------------------|
| <eventSet> / <event> / <eventActor> / <actorInRole> / <attributionQualifierActor> | ja | https://lido-schema.org/schema/latest/lido.html#attributionQualifierActor Information zur Zuschreibung eines Objekts zu einem Beteiligten; besonders bei Künstlern relevant | Empfohlen |
| Kultureller Kontext des Ereignisses | | | |
| <culture> | ja | https://lido-schema.org/schema/latest/lido.html#culture Kultureller Kontext zum Ereignis; vor allem relevant für Herstellung und Nutzung, wenn kein konkretes Individuum als Akteur benannt werden kann Begriff | Empfohlen |
| Datierung des Ereignisses | | | |
| <eventDate> | nein | https://lido-schema.org/schema/latest/lido.html#eventDate Datierung des Ereignisses | Pflicht, wenn vh. |
| <eventDate > / <displayDate> | nur für Sprachvarianten | https://lido-schema.org/schema/latest/lido.html#displayDate Datierungsangabe in Textform, in der ggf. Ungenauigkeiten zum Ausdruck gebracht werden können. | Pflicht, wenn vh. |
| <eventDate > / <earliestDate> | nein | https://lido-schema.org/schema/latest/lido.html#earliestDate Jahresangabe oder exaktes Datum, das den frühestmöglichen Zeitpunkt angibt, zu dem das Ereignis stattgefunden resp. begonnen hat. | Pflicht, wenn vh. |
| <eventDate > / <latestDate> | nein | https://lido-schema.org/schema/latest/lido.html#latestDate Jahresangabe oder exaktes Datum, das den spätmöglichen Zeitpunkt angibt, zu dem das Ereignis stattgefunden resp. geendet hat. | Pflicht, wenn vh. |
| Zeitperiode zum Ereignis | | | |
| <periodName> | ja | https://lido-schema.org/schema/latest/lido.html#periodName Zeitperiode zum Ereignis; z. B. archäologische, historische oder künstlerische Periode. Begriff | Empfohlen |

| Element/Unterelement | Wiederholbar | Bemerkungen/Erläuterungen | Status |
|---|-------------------------|--|-------------------|
| Ortsangaben zum Ereignis | | | |
| <eventPlace> | ja | https://lido-schema.org/schema/lido-test/lido.html#eventPlace Umschlag für Ortsinformation zum Ereignis https://lido-schema.org/schema/lido-test/lido.html#namePlaceSet | Pflicht, wenn vh. |
| <eventPlace><placeID> | ja | https://lido-schema.org/schema/lido-test/lido.html#placeID Identifikator für den Ort. Referenziert idealerweise eine Normdatei. | Empfohlen |
| <eventPlace > / <place> / <namePlaceSet> <appellationValue> | nur für Namensvarianten | https://lido-schema.org/schema/lido-test/lido.html#namePlaceSet https://lido-schema.org/schema/lido-test/lido.html#appellationValue Name des Ortes | Pflicht, wenn vh. |
| <eventPlace > / <place> / <gml> | ja | https://lido-schema.org/schema/lido-test/lido.html#gml Georeferenzen des Ortes | Empfohlen |
| Material- und Technikangaben zum Ereignis | | | |
| <eventMaterialsTech> | ja | https://lido-schema.org/schema/lido-test/lido.html#eventMaterialsTech Umschlag für Material- und Technikangaben zum Ereignis; kann für verschiedene Objektteile differenziert werden. Alternativ bzw. zusätzlich ist die Angabe von Material und/oder Technik ereignisunabhängig in <objectIdentificationWrap> / <objectMaterialsTechWrap> / <objectMaterialsTechSet> möglich. s. https://lido-schema.org/schema/lido-test/lido.html#objectMaterialsTechSet | Pflicht, wenn vh. |
| <eventMaterialsTech > / <materialsTech> / <termMaterialsTech type="..."> | ja | https://lido-schema.org/schema/lido-test/lido.html#termMaterialsTech Ein Material oder eine Technik Mögliche Werte für das type-Attribut s. http://terminology.lido-schema.org/termMaterialsTech_type Begriff | Pflicht, wenn vh. |

9 Thema/Bildinhalt

| Element/Unterelement | Wiederholbar | Bemerkungen/Erläuterungen | Status |
|--------------------------|--------------|---|-------------------|
| <subjectSet> | ja | https://lido-schema.org/schema/latest/lido.html#subjectSet Umschlag für Informationen zum Thema oder Bildinhalt eines Objektes Beachte: Unterschiedliche Entitäten (Begriff, Akteur, Ort, Ereignis, Objekt) haben je eigene Unterelemente. | Pflicht, wenn vh. |
| <subjectSet> / <subject> | nein | https://lido-schema.org/schema/latest/lido.html#subject Indexierung der Themen oder Bildinhalte des Objekts. Dies kann Allgemeinbegriffe (für <subjectConcept>) oder Individualbegriffe (für <subjectActor>, <subjectPlace>, <subjectEvent>, <subjectObject>) als auch Zeitangaben (<subjectDate>) umfassen. https://lido-schema.org/schema/latest/lido.html#subjectConcept https://lido-schema.org/schema/latest/lido.html#subjectActor https://lido-schema.org/schema/latest/lido.html#subjectDate https://lido-schema.org/schema/latest/lido.html#subjectEvent https://lido-schema.org/schema/latest/lido.html#subjectPlace https://lido-schema.org/schema/latest/lido.html#subjectObject Begriff (für alle Unterelemente außer <subjectDate>) | Pflicht, wenn vh. |

Administrative Metadaten

10 Rechte am Objekt

| Element/Unterelement | Wiederholbar | Bemerkungen/Erläuterungen | Status |
|---|-------------------------|--|-------------------|
| <rightsWorkSet> | ja | https://lido-schema.org/schema/latest/lido.html#rightsWorkSet Umschlag für Rechte-Information zum Objekt / Werk. | Pflicht, wenn vh. |
| <rightsWorkSet> / <rightsType> | ja | https://lido-schema.org/schema/latest/lido.html#rightsType Angabe einer Rechte- oder Lizenzklärung Begriff. | Pflicht, wenn vh. |
| <rightsWorkSet > / <rightsHolder> / <legalBodyName>/ <appellationValue> | nur für Sprachvarianten | https://lido-schema.org/schema/latest/lido.html#rightsHolder https://lido-schema.org/schema/latest/lido.html#appellationValue Name des Rechteinhabers. | Pflicht, wenn vh. |

11 Datensatz

| Element/Unterelement | Wiederholbar | Bemerkungen/Erläuterungen | Status |
|--|-------------------------|---|-------------------|
| <recordID> | ja | https://lido-schema.org/schema/latest/lido.html#recordID Eindeutige Identifikationsnummer im (lokalen) System des Datenlieferanten. | Pflicht |
| <recordType> | nein | https://lido-schema.org/schema/latest/lido.html#recordType Datensatztyp: Angabe, ob der Datensatz ein Einzelobjekt, eine Sammlung, Serie, Objektgruppe o.a. beschreibt. Mögliche Werte s. http://terminology.lido-schema.org/recordType Begriff | Pflicht |
| <recordSource> | ja | https://lido-schema.org/schema/latest/lido.html#recordSource Umschlag für Informationen zur Datensatzquelle, i.d.R. die datenliefernde Institution | Pflicht |
| <recordSource> / <legalBodyID> | ja | https://lido-schema.org/schema/latest/lido.html#legalBodyID Ein Identifikator für die Datensatzquelle, z.B. ISIL | |
| <recordSource> / <legalBodyName> / <appellationValue> | nur für Sprachvarianten | https://lido-schema.org/schema/latest/lido.html#legalBodyName https://lido-schema.org/schema/latest/lido.html#appellationValue Name zur Datensatzquelle. | Pflicht |
| <recordRights> | ja | https://lido-schema.org/schema/latest/lido.html#recordRights Umschlag für Rechte-Information zum Datensatz. | Pflicht, wenn vh. |
| <recordRights> / <rightsType> | ja | https://lido-schema.org/schema/latest/lido.html#rightsType Angabe einer Rechte- oder Lizenzklärung Begriff | Pflicht, wenn vh. |
| <recordRights> / <rightsHolder> / <legalBodyName> / <appellationValue> | nur für Sprachvarianten | https://lido-schema.org/schema/latest/lido.html#rightsHolder https://lido-schema.org/schema/latest/lido.html#legalBodyName https://lido-schema.org/schema/latest/lido.html#appellationValue Name des Rechteinhabers. | Pflicht, wenn vh. |
| <recordInfoSet> / </recordInfoID> | ja | https://lido-schema.org/schema/latest/lido.html#recordInfoID Ein Identifikator für den publizierten Datensatz, möglichst in Form eines dereferenzierbaren URIs; | Pflicht, wenn vh. |
| <recordInfoSet> / <recordInfoLink> | ja | https://lido-schema.org/schema/latest/lido.html#recordInfoLink Link zum Metadatensatz oder der Objektpräsentation. | Pflicht, wenn vh. |

| Element/Unterelement | Wiederholbar | Bemerkungen/Erläuterungen | Status |
|----------------------|--------------|---|-----------|
| <collection> | ja | https://lido-schema.org/schema/latest/lido.html#collection Umschlag für Angaben zur (analogen oder digitalen) Sammlung/Objektauswahl, der das Objekt zugeordnet ist. Hierüber kann insbesondere die Zuordnung zu einem Digitalisierungsprojekt hinterlegt werden. Als Beispiel für die Belegung der Unterelemente <displayObject>, <objectWebResource>, <objectType>, <objectName> und <objectNote> s. https://doi.org/10.11588/data/CHEPS6/PPPEMA Sollte es für die Sammlung/Objektauswahl eine ID (z.B. DFG-Förderkennzeichen) geben, kann dafür das Unterelement <objectID> verwendet werden s. https://lido-schema.org/schema/latest/lido.html#objectID | Empfohlen |

12 Repräsentationen

| Element/Unterelement | Wiederholbar | Bemerkungen/Erläuterungen | Status |
|---|-------------------------|--|-------------------|
| <resourceSet> | ja | https://lido-schema.org/schema/latest/lido.html#resourceSet Umschlag für Angaben zu (digitalen) Repräsentationen des Objekts. | Pflicht, wenn vh. |
| <resourceSet> / <resourceRepresentation > | ja | https://lido-schema.org/schema/latest/lido.html#resourceRepresentation Link(s) zur digitalen Repräsentation des Objekts, ggf. in verschiedenen Auflösungen und mit technischen Angaben zur Wiedergabe (Audio, Video) | Pflicht, wenn vh. |
| <resourceSet> / <rightsResource> | ja | https://lido-schema.org/schema/latest/lido.html#rightsResource Umschlag für Rechte-Information zur (digitalen) Repräsentation. | Pflicht, wenn vh. |
| <resourceSet> / <rightsResource> / <rightsType> | ja | https://lido-schema.org/schema/latest/lido.html#rightsType Angabe einer Rechte- oder Lizenzklärung Begriff | Pflicht, wenn vh. |
| <resourceSet> / <rightsResource> / <rightsHolder> | ja | https://lido-schema.org/schema/latest/lido.html#rightsHolder Informationen zum Rechteinhaber | Pflicht, wenn vh. |
| <resourceSet> / <rightsResource> / <rightsHolder> / <legalBodyID> | ja | https://lido-schema.org/schema/latest/lido.html#legalBodyID Eine Identifikation für den Rechteinhaber, z.B. ISIL oder GND-ID | Pflicht, wenn vh. |
| <resourceSet> / <rightsResource> / <rightsHolder> / <legalBodyName>/ <appellationValue> | nur für Sprachvarianten | https://lido-schema.org/schema/latest/lido.html#appellationValue Name des Rechteinhabers. | Pflicht, wenn vh. |

2. Beispiele für LIDO-Datensätze

LIDO-Beispieldatensätze finden sich unter:

- (1) <https://lido-schema.org/examples>
(heranzuziehen sind die Beispiele zur jeweils aktuellsten LIDO-Version)
- (2) Knaus, Gudrun; Kailus, Angela; Stein, Regine, 2022, „LIDO-Handbuch für die Erfassung und Publikation von Metadaten zu kulturellen Objekten - Band 2: Malerei und Skulptur [Anwendungsbeispiele]“, <https://doi.org/10.11588/data/CHEPS6>, heiDATA, V1 (ggf. auf neuere Versionen achten)