

DFG-Praxisregeln

„Digitalisierung“



Deutsche Forschungsgemeinschaft

Kennedyallee 40 · 53175 Bonn · Postanschrift: 53170 Bonn

Telefon: + 49 228 885-1 · Telefax: + 49 228 885-2777 · postmaster@dfg.de · www.dfg.de

DFG

Inhaltsverzeichnis

Vorbemerkung zu den Praxisregeln „Digitalisierung“	4
1. Ziele und Auswahl	5
1.1 Ziele	5
1.2 Auswahl	5
1.3 Dublettenprüfung und Datenabgleich bei der Retrodigitalisierung publizierter Texte	6
2. Digitalisierung	7
2.1 Bereitstellung der Materialien, konservatorische Prüfung	7
2.2 Technische Parameter der digitalen Reproduktion	8
2.2.1 Allgemeine Erläuterungen und Parameter	8
2.2.1.1 Auflösung	8
2.2.1.2 Farbtiefe	10
2.2.1.3 Digitaler Aufnahmearlauf	11
2.2.1.4 Dateiformate	15
2.2.2 Materialspezifische Parameter	17
2.2.2.1 Textwerke	17
2.2.2.2 Grafische Darstellungen	17
2.2.2.3 Fotografien	18
2.2.2.4 Mikroformen	21
2.2.2.5 Dreidimensionale Objekte	21
2.2.3 Metadaten	25
2.3.1 Erschließung, deskriptive Metadaten	27
2.3.2 Strukturelle Metadaten für digitale Faksimiles	27
2.3.3 Sammlungs- und Bestandsbeschreibung	28
2.3.4 Austausch und Weitergabe der Metadaten	29
2.4 Volltextgenerierung	30
2.4.1 Texterfassung	30
2.4.2 Zeichenkodierung	34
2.4.3 Markup von Volltexten	34
2.4.4 Layout	35
2.5 Langzeitverfügbarkeit	35
3. Organisatorische Fragen	37
4. Zitieren digitaler Ressourcen, persistente Adressierung	39
5. Bereitstellung der Digitalisate für die Öffentlichkeit	40
5.1 Rechteklärung und Open Access	40
5.2 Mindestanforderungen für die Bereitstellungssysteme von Digitalisaten	41
5.2.1 Funktionalitätsanforderungen	41
5.2.2 Technische Mindestanforderungen	42
6. Präsentationsstandards für Textwerke, DFG-Viewer	43

6.1 Basisanforderungen und Architektur	43
6.2 Funktionalitätsanforderungen	44
6.3 DFG-Viewer	44
7. Prüfliste für Antragsteller und Gutachter	45
7.1 Allgemeine technische Abläufe/Voraussetzungen	46
7.2 Technische Parameter der digitalen Reproduktion	46
7.3 Metadaten	47
7.4 Volltextgenerierung	48
7.5 Langzeitverfügbarkeit	49
7.6 Organisatorische Fragen	49
7.7 Zitieren, persistente Adressierung	50
7.8 Bereitstellung der Digitalisate für die Öffentlichkeit	50
8. Regeln für die Durchführung von Digitalisierungsprojekten	51

[Anhang A: METS/MODS-Profil für die Darstellung im DFG-Viewer und Übermittlung per OAI](#)

Anhang B: METS/TEI-Spezifikation für die Darstellung von digitalisierten Handschriften (Der Anhang B wird derzeit noch erarbeitet und wird zu einem späteren Zeitpunkt hier veröffentlicht werden.)

[Anhang C: LIDO-Kernelemente für die Publikation](#)

[Anhang D: Sammlungs- und Bestandsbeschreibung](#)

Vorbemerkung zu den Praxisregeln „Digitalisierung“

Die Deutsche Forschungsgemeinschaft (DFG)¹ fördert im Bereich der Wissenschaftlichen Literaturversorgungs- und Informationssysteme (LIS) Projekte an wissenschaftlichen Einrichtungen, insbesondere Service- und Informationseinrichtungen in Deutschland. Förderziel ist der Aufbau leistungsfähiger Informationssysteme für die Forschung unter überregionalen Gesichtspunkten. Die Ergebnisse der Projekte sollen für die Wissenschaft frei und dauerhaft zugänglich sein.

Ziel der Praxisregeln „Digitalisierung“ im Förderbereich LIS ist es, Antragstellern die Planung von Digitalisierungsprojekten zu erleichtern und die Begutachtung von Anträgen vergleichbar zu gestalten. Die Praxisregeln wollen keine Hürden aufbauen, sondern durch die Formulierung von Standards einen Beitrag zur Nachhaltigkeit und Zukunftsfähigkeit der unterstützten Projekte leisten.

Die Standards ergänzend finden sich in den Praxisregeln auch weiterführende Ausführungen, beispielsweise zur konservatorischen Prüfung der zur Digitalisierung vorgesehenen Materialien, zum Erheben von Metadaten, zur Herstellung der Digitalisate, zur Indexierung von Bildinhalten, zur Herstellung von Volltexten oder auch zur Perspektive der Langzeitsicherung digitaler Inhalte.

Im Folgenden wird in den Kapiteln 1 bis 5 eine grundsätzliche und weiter ausholende Einführung in die Probleme und Vorgehensweisen gegeben, die sich für Projekte stellen, die Materialien aus Bibliotheken, Archiven oder musealen Sammlungen in digitale Form überführen wollen. Diese Kapitel wenden sich bewusst auch an Personen, die derartige Projekte neu planen, also unter Umständen über keine detaillierten Vorkenntnisse verfügen. Kapitel 6 verweist auf für textuelle Materialien erforderliche DFG-Präsentationsstandards und Formate. In Kapitel 7 sind knapp noch einmal die wichtigsten Anforderungen zusammengefasst. Abweichungen von den dort beschriebenen Regeln sind zwar möglich, müssen aber im Einzelfall begründet werden. In Kapitel 8 werden abschließend wichtige Verfahrens-Regeln für die Durchführung eines DFG-Vorhabens benannt.

Auch wenn dies im Folgenden öfter wiederholt werden wird, ist ein Prinzip so wichtig, dass es explizit vorangestellt werden soll: Die wissenschaftlich motivierte Digitalisierung wird nicht als neue Technik, sondern als Normalfall verstanden. In der Praxis bedeutet dies für die Konzeption von Projekten, dass es zwar weiterhin wichtig ist, Digitalisate zu erzeugen, deren Qualität den Anforderungen der die Digitalisate nutzenden Forschung genügt, dass aber gleichzeitig auch darauf zu achten ist, dass die gewählten Vorgehensweisen effektiv und

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 Forschern (<http://www.dfg.de>). Die DFG unterstützt auch Vorhaben zur Verbesserung der wissenschaftlichen Informations-Infrastrukturen in Deutschland. Die Ergebnisse der geförderten Projekte sollen für die Wissenschaft frei und dauerhaft zugänglich sein (<http://www.dfg.de/lis>). Zu beachten ist, dass die definierten Trägeraufgaben und -finanzierungen der antragstellenden Einrichtungen durch die Förderung nicht substituiert werden dürfen. Projekte müssen daher in ihrer Profilierung über die regulären Grundaufgaben einer Einrichtung hinausgehen, zeitlich und inhaltlich begrenzt sein sowie herausragende und überregional bedeutende Materialien zum Gegenstand haben. Nicht förderfähig sind damit Vorhaben, die vorrangig der Kulturförderung, Kulturgutvermittlung oder vergleichbaren Zielsetzungen dienen, sowie kommerziell orientierte Projekte.

kostenbewusst genug sind, um systematisch auf große Mengengerüste angewendet werden zu können.

1. Ziele und Auswahl

1.1 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, sondern auch und vor allem das Vernetzen der unterschiedlichen Ressourcen zu einer virtuellen Forschungsinfrastruktur.

Es gibt heute einen breiten Stand abgesicherter 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 Bearbeitung von Massenverwaltungsschriftgut des späten 19. Jahrhunderts unnötig zeitraubend und überteuert sein. Diese Empfehlungen gehen von der Vorstellung aus, dass die Digitalisierung für eine wissenschaftliche Informationseinrichtung des 21. Jahrhunderts eine normale Dienstleistung und kein „Elitemerkmal“ darstellt. Der digitale Zugang wird in naher Zukunft die Regel und nicht die Ausnahme darstellen. Die Digitalisierung hervorragender Einzelstücke oder bedeutender Sammlungen steht hierbei in keinem Gegensatz zu großflächigen Digitalisierungsmaßnahmen.

Da sich die Mehrzahl der bisher durchgeführten und in Planung befindlichen DFG-geförderten Projekte mit textuellen Materialien befasst haben, werden im Folgenden die Techniken und Parameter für diesen Bereich besonders gründlich erörtert. Hinzu kommen Ausführungen zu bildhaften Materialien, seien es grafische Darstellungen, Fotografien oder Bilddigitalisate von dreidimensionalen Sammlungsobjekten. Die vorliegende Version der Praxisregeln integriert dabei neuere Entwicklungen in diesen Bereichen, ohne den Anspruch auf Vollständigkeit zu erheben. In laufenden und geplanten Pilotprojekten werden über die DFG-Förderung spezifische Techniken und Parameter für die Digitalisierung historischer Zeitungen und unikalen Materials (Archivgut, mittelalterliche Handschriften, dreidimensionale Objekte) ermittelt, die nach Abschluss der Pilotphasen Eingang in die Praxisregeln finden und damit eine Fortschreibung der vorliegenden Version darstellen werden.

1.2 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, ausdrücklich 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, so dass die wechselseitige Vernetzung möglich wird.² Alternativ kann man auf einschlägige Fachbibliografien zurückgreifen, die einen Kanon formulieren, oder aber auch eine Mischung aus beidem anstreben. Interessant sind sicher auch Konzepte des *digitization on demand* oder *digitization on use* (Digitalisierung von in Bibliotheken und Archiven genutzttem oder stark genutzttem Material), wodurch in jedem Fall ein konkretes Nutzungsinteresse sichergestellt ist.³

Die Digitalisierung von wissenschaftsrelevanten Materialien wird durch die DFG gefördert, um diese Materialien der Forschung in Deutschland und weltweit zugänglich zu machen. Bei der Auswahl der für ein Digitalisierungsprojekt in Frage kommenden Materialien ist unbedingt zu beachten, dass die Ergebnisse der durch die DFG geförderten Digitalisierungsprojekte der Forschung im *open access* zur Verfügung gestellt werden müssen. Bereits mit Beginn der Projektplanung, spätestens aber zum Zeitpunkt der Antragstellung muss die Rechteklärung bezüglich der zu digitalisierenden Materialien erfolgt sein. Insbesondere eventuell vorliegende Urheber-, Persönlichkeits- und Leistungsschutzrechte sind dabei zu berücksichtigen.

Größere Digitalisierungsprojekte, die auf die Schaffung einer umfassenden Infrastruktur zielen, sollten sich in ein Gesamtprogramm (z.B. Masterplan VD17) einfügen und in den jeweiligen Communities abgestimmt sein.

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

1.3 Dublettenprüfung und Datenabgleich bei der Retrodigitalisierung publizierter Texte

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

² In der Regel bauen Forschungsprojekte auf geförderten Infrastrukturen auf, es ist jedoch ebenfalls möglich, Projekte, die Forschung und Infrastrukturelemente integrieren, zu beantragen.

³ Zu weiteren Gesichtspunkten der Auswahl vgl. u.a. das Minerva Good Practice Handbook.
Vgl.: <http://www.minervaeurope.org/publications/goodhand.htm>

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 Digitalisierungsprojekten (insbesondere Google) verhält. Ziel sollte es sein, die Zahl der gegebenenfalls auftretenden Doppeldigitalisierungen mit pragmatischem Aufwand möglichst gering zu halten. Bei Materialien 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.

2. Digitalisierung

Die Digitalisierung umfasst immer die Herstellung digitaler Images und die Erzeugung von Metadaten, im Falle von Textwerken gegebenenfalls zusätzlich auch die Volltexterfassung und die Erzeugung von Strukturdaten und Markup. Wenn im Folgenden von Digitalisierung gesprochen wird, so ist der gesamte Arbeitsgang gemeint (Vorbereitung, Digitalisierung im engeren Sinne, Erzeugung von Metadaten sowie Langzeitsicherung/digitale Bestandserhaltung). Auf eine Differenzierung nach Materialarten (u.a. Druckwerke, unikale Dokumente, Objekte) wird in den allgemeinen Kapiteln im Folgenden verzichtet.

2.1 Bereitstellung der Materialien, konservatorische Prüfung

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

Die konservatorische Prüfung kann sehr viel Zeit in Anspruch nehmen, sollte aber zum Schutz der Materialien keinesfalls unterbleiben. Es wird empfohlen „Checklisten“ für die Prüfung der Digitalisierungsfähigkeit zu verwenden und daraus die Vorgaben für die Digitalisierung zu entwickeln.⁴ Stellt die Reproduktion mit den vorhandenen Reproduktionstechniken ein Risiko bzw. eine besondere Belastung für die Originale dar,⁵ so sollte sie – sofern möglich – vom

⁴ 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 weilig; Blätter/Lagen sehr steif; Textverlust unvermeidbar; Planlage der Vorlage unmöglich; extremes Format.

⁵ Vgl. Petersen, Dag-Ernst: Die Mikroform: Chance und Gefahr für das Buch. In: IADA Preprints 1999. 9. IADA Kongress, 16. - 21. August 1999 in Kopenhagen, S. 181-183.

bestehenden Mikrofilm⁶ 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 Scanteknik und der Einsatz von Hilfsmitteln, wie Buchkeile oder Fixierhilfen, sind entsprechend den Vorgaben der konservatorischen Prüfung zu wählen.

2.2 Technische Parameter der digitalen Reproduktion

Ziel der Digitalisierung ist die möglichst originalgetreue Wiedergabe des Materials nach Maßgabe der wissenschaftlichen Erfordernisse. Die anzuwendenden Parameter für die Digitalisierung sind mit Blick auf die Qualität des Bildes, seine Langzeitverfügbarkeit und Interoperabilität zu wählen. Neben den folgenden Regeln können zur vergleichenden Orientierung auch die Richtlinien der Federal Agencies Digitization Guidelines Initiative (FADGI)⁷ oder anderer einschlägiger Institutionen herangezogen werden.

2.2.1 Allgemeine Erläuterungen und Parameter

Bei der Herstellung sind grundsätzlich zwei Formen von Digitalisaten zu berücksichtigen. Zum einen der sogenannte digitale Master, also das Roh- oder Archivformat, und für die Nutzung hergestellte Derivate, meist verkleinerte Kopien in anderen Dateiformaten. Derivate wie z.B. JPEG-Dateien sind in Abhängigkeit von der gewählten Präsentation aus den Mastern zu erzeugen und können, sofern die Master ausreichende Qualität bieten, nach Belieben modifiziert werden. Eine Modifikation wird z.B. nötig, wenn sich die bei den Nutzerinnen und Nutzern zu erwartenden Bildschirmauflösungen ändern oder Bildformate zum Einsatz kommen sollen, die optimierte Eigenschaften für die gewünschte Anzeige haben (stufenfreies Zoomen; fließende Übergänge zwischen Abschnitten größer, gleichzeitig aber detailreicher Objekte, wie zum Beispiel Landkarten, mittelalterliche Urkunden oder Kupferstiche).

Der Master ist die Grundlage für alle weiteren Prozesse. Daher sollte man seiner Herstellung besondere Aufmerksamkeit schenken und bei der Langzeitsicherung entsprechende Empfehlungen (s. z.B. Nestor⁸) berücksichtigen.

Die folgenden Richtlinien für die Mindestanforderungen an die Digitalisierung beziehen sich ausschließlich auf den **digitalen Master**.

2.2.1.1 Auflösung

Das gebräuchliche Maß für das Auflösungsvermögen einer Datei ist dpi (*dots per inch*), gleich Bildpunkte pro inch. Dieses Maß gibt Auskunft über die Detailwiedergabe der Datei.

Als Untergrenze sollte daher ein Scanauflösung für die Digitalisate gewählt werden, bei der die Details einer Vorlage vollständig in einer gleichgroßen Reproduktion wiedergegeben werden können. 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 ermüdfrei über einen längeren Zeitraum

⁶ In Archiven ist die Digitalisierung vom Mikrofilm unabhängig von konservatorischen Erwägungen durchaus Praxis. Zu den materialspezifischen Parametern für eine Digitalisierung vom Mikrofilm vgl. Kapitel 2.2.2.

⁷ Federal Agencies Digitization Guidelines Initiative (FADGI): Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files, August 2010.

Vgl.: http://www.digitizationguidelines.gov/guidelines/FADGI_Still_Image-Tech_Guidelines_2010-08-24.pdf

⁸ <http://www.langzeitarchivierung.de>

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 dazwischen liegt. Eine Rasterfrequenz von 60 Linien/cm bei 20 cm Betrachtungsabstand kann nicht mehr als getrennte Linien wahrgenommen werden.⁹ Daraus resultiert folgende Auflösungsanforderung, die auch in der Druckindustrie gebräuchlich ist:

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

$120 \text{ px/cm} \times 2,54 \text{ cm/inch} = 304,8 \text{ dpi}$, rund **300 dpi**

Daher gilt die grundsätzliche Empfehlung von **300 dpi bezogen auf das Vorlagenformat als Zielauflösung** (Vorlagenformat = Ausgabeformat des Digitalisats bei 300 dpi).

Die obige Herleitung bezieht sich allerdings nur auf Vorlagen, die für die Betrachtung mit bloßem Auge gedacht sind, wie etwa: Textwerke, Grafiken und fotografische Aufsichtvorlagen. Anders verhält es sich bei Medien, deren vollständige Bildinformation nur in Vergrößerung erfassbar wird. Das sind Miniaturen jeglicher Art und ganz besonders die fotografischen Durchsichtmedien.

Hierzu ein 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 26 cm nach obiger Maßgabe im Digitalisat des KB-Negativs zu erreichen, müsste es mit mindestens 3.000 dpi digitalisiert werden.

Weiterhin gibt es Objekte, insbesondere Objekte der Architektur, aber auch überformatige Malereien wie beispielsweise wandgroße Fresken oder Skulpturen im öffentlichen Raum, die sich auf Grund ihrer Dimensionen den oben beschriebenen Verfahren zur Bestimmung der Auflösung entziehen.

In der bildmäßigen Fotografie für Sammlungsobjekte und Architekturfotografie wird beispielsweise mit einer modernen 21-Megapixel-Digitalkamera eine Auflösung von 300 dpi bei DIN A3 Ausgabegröße erreicht. Diese Auflösung ist ausreichend für vielseitige Verwendungen wie Druckerzeugnisse unterschiedlicher Größe und die Präsentation im Internet. Jedoch können höhere Auflösungen notwendig sein, wenn bildwichtige Details mit diesen Vorgaben nicht dargestellt werden können (z.B. bei Reproduktionen von Gemälden größer als doppelt DIN A0). Bei einem Flächensor mit 3.800 x 5.600 Pixeln ergibt sich also nach der Beziehung Kamerapixel/Zielauflösung in dpi = Zielgröße in inch folgende Ausgabegröße bei 300 dpi:

$$5600/300 = 18,7 \text{ inch} \times 2,54 \text{ cm/inch} = 47 \text{ cm}$$

$$3800/300 = 12,7 \text{ inch} \times 2,54 \text{ cm/inch} = 32 \text{ cm}$$

⁹ Vgl. Weber, Helen: Digitale Farben in der Medienproduktion und Druckvorstufe. Heidelberg: MITP Verlag 2006, S.24ff.

Hinweise für eine gegebenenfalls sinnvolle Abweichung von der grundsätzlichen Auflösungsempfehlung von 300 dpi bezogen auf das Vorlagenformat werden nachstehend bei den materialspezifischen Parametern genannt. (Vgl. Kapitel [2.2.2](#))

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.

Ob ein digitales Aufnahmesystem für ein Digitalisierungsvorhaben geeignet ist, lässt sich vorab im Rahmen einer Testdigitalisierung mit genormten Testcharts ermitteln. Testcharts sind in der Regel definierte Linienmuster, an denen sich feststellen lässt, ab welcher Linienwiederholfrequenz in Abhängigkeit vom Abbildungsmaßstab ein digitales Aufnahmesystem diese Linien nicht mehr unterscheiden kann.

Das ISO-12233-Testchart wurde speziell für die digitale Fotografie entwickelt. Das USAF-1951-Testchart wurde von der US Airforce zur Beurteilung der Qualität optischer Instrumente entwickelt. Obwohl schon 1951 entwickelt, wird dieses Testchart bis heute verwendet.

Die effektive Auflösung der Testbilder lässt sich dann mit einer Analysesoftware oder mit entsprechenden, in der Regel beigefügten Auflösungstabellen ermitteln.

2.2.1.2 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.

- Bei einer Farbtiefe von 1 Bit sind zwei unterschiedliche Zustände möglich: Weiß und Schwarz.
- Bei einer Farbtiefe von 8 Bit/Farbkanal sind es $2^8 = 256$ Helligkeitsstufen/Farbkanal von Weiß nach Schwarz.
- Bei einer Farbtiefe von 16/Farbkanal Bit sind es $2^{16} = 65.536$ Helligkeitsstufen/Farbkanal von Weiß nach Schwarz.
- Graustufenbilder mit nur einem Kanal für die Helligkeit haben somit eine Farbtiefe von 8 oder 16 Bit.
- Farbbilder im RGB-Modus (je ein Farbkanal für Rot, Grün und Blau) besitzen somit eine Farbtiefe von 3×8 Bit = 24 Bit, bzw. 3×16 Bit = 48 Bit.

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

Digitale Aufnahmesysteme zeichnen Farben im RGB-Farbmodus auf. Die Analog-/Digitalwandler in hochleistungsfähigen chipbasierten Scanbacks und Zeilenscanbacks können Helligkeitsunterschiede mit 16 Bit differenzieren. Bei den meisten digitalen KB-Kameras und Flachbettscannern sind es 14 Bit.

Digitale Aufnahmesysteme zeichnen grundsätzlich mit einer hohen Farbtiefe von 14 Bit bzw. 16 Bit pro Kanal auf. Um die Informationsdichte der erzeugten Bilddateien zu erhalten, wird eine notwendige Nachbearbeitung der Bilddaten in 16 Bit pro Kanal, d.h. 48 Bit durchgeführt.

Für die Sicherung des finalen digitalen Master ist eine **Farbtiefe von 8 Bit pro Kanal**, d.h. **24 Bit** ausreichend, da die heute gängigen Ausgabe- und Anzeigegeräte nur eine Tonwertwiedergabe mit 8-Bit-Differenzierung unterstützen.

2.2.1.3 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. Über dieser Pixelmatrix liegen abwechselnd für jeweils 2 x 2 Pixel Farbfilter in einmal rot, zweimal grün und einmal blau. Mit dieser sogenannten Bayermatrix werden Farben erzeugt. Man unterscheidet zwischen dem Oneshot- und dem Multishotverfahren:

- Das Oneshotverfahren ist die am weitesten verbreitete digitale Aufnahmetechnik. Hier werden, um die volle Auflösung des Flächensensors zu nutzen, die fehlenden Farbinformationen für einen Bildpunkt aus den umliegenden Pixeln interpoliert. Die Algorithmen zur Farbinterpolation sind heute so ausgereift, dass bezüglich Farb- und Detailwiedergabe das Oneshot-Verfahren für die meisten Digitalisierungsaufgaben geeignet ist.
- Beim Multishotverfahren, das nur bei speziellen hochwertigen Digitalrückteilen zum Einsatz kommt, wird der Bildsensor im Pixelabstand verschoben und für jeweils eine Farbe eine Aufnahme gemacht. Die einzelnen Bilder werden anschließend in der Aufnahmesoftware zu einem Bild verrechnet, eine Farbinterpolation ist nicht nötig. Hierdurch können Farbverschiebungen und Moirée-Effekte vermieden werden.

Zeilenscanner und multishotfähige Flächensensorkameras finden hauptsächlich Anwendung bei filigranen Motiven wie bildhaften Tiefdruckwerken, Kartenwerken oder Textilien. Zeilenscanner ermöglichen bei Vorlagen bis DIN A0 hohe Auflösungen und extreme Detailgenauigkeit. Für große zweidimensionale Vorlagen ab doppelt DIN A0, die zur Gewährleistung der Lesbarkeit mit höchster Detailwiedergabe aufgezeichnet werden sollen, sind Zeilensensoren in Verbindung mit einem Kamerasystem vorzuziehen.

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 restauratorisch unbedenkliche Vorlagen.

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.

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. 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 Kamerasytem können für Reproduktionen in diesem Maßstabsbereich auch hochwertige Vergrößerungsobjektive eingesetzt werden.

Die Abbildungsleistung eines Objektivs lässt sich auch an den dazugehörigen Modulationtransferfunktions-Diagrammen (MTF-Diagrammen) ablesen. Diese Diagramme werden von diversen Objektivherstellern veröffentlicht. Die Modulationtransferfunktion beschreibt, wie gut ein Objektiv den Kantenkontrast in einer Vorlage auf das fotografische Abbild übertragen kann. Sie bietet somit ein Hinweis auf das Auflösungsvermögen einer Optik. Die technischen Datenblätter der Objektive geben auch Auskunft über den Helligkeits- und Schärfeabfall 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.

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.¹⁰

Ein Moiré-Effekt tritt in der digitalen Bilderzeugung auf, wenn in Abhängigkeit von der Pixelgröße der Scanzeile oder des Flächensensors, des Abbildungsmaßstabes und der

¹⁰ 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.

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 Scanzeiten länger sind und eine absolut konstante Beleuchtung während des Aufnahmevergangs unabdingbar ist.

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 konservatorischen 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 Multishotsystemen ist darauf zu achten, dass die Blitzleistung und -farbe während der einzelnen Belichtungen konstant bleiben. Gleches gilt für Dauerlicht.

Belastend für Vorlagen sind Wärme und Lichtintensität, vor allem der UV-Anteil. Blitzlampen sind in der Regel mit einer UV-Sperfilter-Schutzglocke ausgestattet. Grundsätzlich ist bei allen Beleuchtungsarten darauf zu achten, dass das UV-Licht herausgefiltert wird.

Auf jeden Fall ist eine kurzfristige hohe Wärmefbelastung für die Vorlagen zu vermeiden. Im ungünstigsten Fall können die Vorlagen durch zu hohe Hitze und den plötzlichen Temperaturanstieg beschädigt werden. Das Wärmeproblem betrifft vorrangig Dauerlicht. Halogenlampen haben eine sehr hohe Wärmeentwicklung. Hier sind Lösungen mit Tageslichtleuchtstofflampen oder LED vorzuziehen. Sie haben eine geringere Wärmeentwicklung.

Blitzlicht gibt innerhalb einer kurzen Zeitspanne (1/200 sec und kürzer) eine hohe Lichtmenge ab, die Wärmeentwicklung ist zu vernachlässigen. Vorteile sind eine kurze Digitalisierungszeit, das Vermeiden von Verwacklungen und die Möglichkeit des Arbeitens bei Raumbeleuchtung.

Unter konservatorischen Gesichtspunkten lässt sich die Meinung vertreten, dass eine kurzfristige hohe Lichtbelastung weniger schädlich ist als eine hohe Wärmefbelastung, zumal

bezüglich der Schadwirkung 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.

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. Solche RAW-Bilder repräsentieren die originalen Kameradaten. Bildkorrekturen sollten möglichst in der RAW-Software des jeweiligen Kameraherstellers vorgenommen werden. Die vorgenommenen Änderungen werden dabei nicht in der Datei gespeichert, sondern von der Software als Bearbeitungsvorschrift in einer Datenbank bzw. in einer XML-Datei abgelegt. Die RAW-Datei wird also nicht verändert und kann bei Bedarf verlustfrei neu korrigiert werden. Bearbeitungsvorschriften können überdies als *Preset* auf beliebig viele Bilder angewendet werden, wodurch eine hohe Arbeitseffizienz erreicht wird.

Als Profil für Farbaufnahmen sind entweder Adobe RGB, der von der European Color Initiative (ECI)¹¹ empfohlene Arbeitsfarbraum ECI-RGB v2 oder der dazu identische Farbraum L-Star RGB¹² zu wählen. CMYK-Farträume kommen als ausschließliche Druckausgabeprofile nicht in Frage.

Für eine verlässliche Farbabstimmung in der späteren Bildverarbeitung ist es notwendig, Graukeile oder Farbkeile mit zu fotografieren. Sofern man keine Einstellungsänderung vornimmt, ist es ausreichend, den Graukeil oder Farbkeil pro Bildserie nur einmal repräsentativ für die ganze Serie aufzunehmen.

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. Nach heutigem Stand sind Tonwerte von 95 % für Schwarz und 5 % für Weiß differenziert druckbar. Als maximale Helligkeitswerte empfehlen sich:

Schwarz: RGB 16/16/16, Graustufen 90 %
Weiß: RGB 232/232/232, Graustufen 10 %

Bei Reproduktionen von zweidimensionalen Vorlagen sind Freisteller so zu gestalten, dass die gesamte Vorlage mit leichtem, umlaufendem Rand abgebildet wird. Nur so ist erkennbar, dass nichts von der Vorlage abgeschnitten wurde. Das Nachschärfen („Unscharfmaskieren“) sollte für den digitalen Master nur moderat erfolgen. Anwendungsgebunden kann für unterschiedliche Derivate eine zusätzliche Schärfung erfolgen.

Notwendige Bildbearbeitungen in der Bilddatei sind bei 16 Bit Farbtiefe pro Kanal vorzunehmen. Sind alle Korrekturen vorgenommen, wird die RAW-Datei anschließend in eine TIFF-Datei mit 24 Bit/RGB oder 8 Bit/Graustufen als Master konvertiert. Nur falls noch weitere Bearbeitungen in der TIFF-Datei vorgesehen sind, erfolgt die Konvertierung aus der RAW-Datei mit 48 Bit/RGB oder 16 Bit/Graustufen. Nach Beendigung aller notwendigen Bearbeitungsschritte wird der finale digitale Master als TIFF mit 24 Bit/RGB oder

¹¹ European Color Initiative: <http://www.eci.org>

¹² <http://www.colormanagement.org/de/workingspaces.html>

8 Bit/Graustufen gesichert und gilt als archivtauglich. Das RAW-Format ist als proprietäres Dateiformat allerdings nicht archivtauglich.

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. Monitormodelle, die in der grafischen Industrie Verwendung finden, liefern in diesem Zusammenhang die verbindlichsten Darstellungen. Sie bieten über die gesamte Grauachse eine konsistente Farbwiedergabe. Die Monitore sollten mit einem Colorimeter auf die entsprechenden Zielwerte für eine D50-Normlichtumgebung kalibriert werden. Dabei ist zu beachten, dass eine Farbtemperatur von 5.800 K statt 5.000 K am Monitor eingestellt wird, da 5.000 K am Monitor gelber wahrgenommen werden als 5.000 K der D50-Normlichtumgebung. Das Schweizer Kompetenzzentrum für Medien- und Druckereitechnologie empfiehlt für eine Kalibration die folgenden Eckwerte:¹³

- Luminanz: mindestens 120 cd/m²
- Weißpunkt: 5800 K mit chromatischer Adaption
- Gradation: 1.8 Gamma oder alternativ L*
- Farbräume: ISOcoated v2 und ECI-RGB 1.0/2.0

Bei hardwarekalibrierbaren Monitoren wird die interne Farbtabelle (LUT) des Monitors so angepasst, dass Farbfehler zugunsten einer neutralen und farbrichtigen Darstellung korrigiert werden. Der Vorteil in der Hardware-Kalibrierung liegt darin, dass der Monitor auf die gewünschten Zielwerte bezüglich Gradation, Weißpunkt, Luminanz und Farbraum abgeglichen wird und nicht, wie bei der Software-Kalibrierung, Korrekturwerte in der Grafikkarte gespeichert werden, die für die korrekte Monitoransicht auf die Bilddateien angewendet werden. Vorsicht ist bei älteren Colorimetern geboten, da diese bei Monitoren mit neuer Technologie, wie LED-Hintergrundbeleuchtung oder erweiterten LUTs, oftmals nicht die gewünschten Ergebnisse liefern.

Digitale Bildnachbearbeitung

Zur Optimierung der Bildqualität digitaler Aufnahmen ist in der Regel eine zusätzliche Nachbearbeitung notwendig. Dabei sollte sich diese 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 Verfremdungsfiltern. Zur Steigerung der Bildintegrität können Positionierungshilfen nachträglich aus dem Bild entfernt werden oder Hintergründe bereinigt werden.

2.2.1.4 Dateiformate

Bildmaster von Graustufen oder Farbbildern sollten nach dem derzeitigen Kenntnisstand in „**TIFF uncompressed**“¹⁴ langzeitgesichert werden. Das Format TIFF gibt es schon seit den 1980er Jahren. Es hat sich als einer der wichtigsten Standards etabliert und es ist damit zu rechnen, dass es auch in Zukunft von allen Standardprogrammen unterstützt wird. Allerdings gilt dies nur für so genannte „Baseline-TIFFs“. Die weitgehenden Optionen von Extended TIFFs sollten für den digitalen Master nicht genutzt werden.

¹³ <http://www.ugra.ch>

¹⁴ TIFF uncompressed, Revision 6: <http://www.digitalpreservation.gov/formats/fdd/fdd000022.shtml>

Neben TIFF kann auch **JPEG2000**¹⁵ 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 lizenzenfreien Bereiche von JPEG2000 Verwendung finden.

In den letzten Jahren drängt sich gerade angesichts der zunehmenden Speicherproblematik JPEG2000 als effizientes Kompressionsformat auf. Allerdings ist JPEG2000 mit Blick auf die Langzeitarchivierung mit Risiken behaftet und Betreiber von Repositoryn sollten Vor- und Nachteile des Einsatzes sorgfältig abwägen. Denn einerseits sind Formate mit Kompression grundsätzlich problematisch, auch wenn JPEG2000 weniger störungsanfällig und robuster ist als andere Kompressionsformate. Andererseits ist seine Verbreitung und Marktdurchdringung, unabhängig davon, dass internationale große und einflussreiche Bibliotheken wie die Library of Congress und die British Library auf JPEG2000 setzen, immer noch unzureichend. Namhafte Softwarehersteller wie Adobe haben nach anfänglicher Unterstützung des Formats diese wieder eingestellt und aus Standardsoftware wie Photoshop entfernt, was die weitere Verbreitung beeinträchtigt hat. Letztlich ist auch die Lizenzierungssituation noch nicht abschließend geklärt, selbst wenn Bereiche von JPEG2000 als frei deklariert wurden.

Andere Kompressionsformate wie JPEG oder PNG sollten nach Möglichkeit nicht verwendet werden, weil schon durch geringe Beschädigung der Datei oder Kippen einzelner Bits das gesamte Bild beeinträchtigt wird. Solche Beschädigungen können durch Defekte der Speicherträger oder auch beim Kopieren der Dateien auftreten. Jüngere Untersuchungen zeigen, dass neben JPEG auch PNG für Defekte dieser Art anfällig ist, weshalb es abweichend von der letzten Version der Praxisregeln als Archivformat nicht mehr empfohlen werden kann.

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 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 JPEG und PNG.

Für AV-Medien muss ebenso zwischen Archivformaten und Nutzungsformaten unterschieden werden. Im Bereich Audio hat sich das Waveform Audio File-Format (WAV) als Quasi-Standard etabliert. Als Nutzungsformat hat sich mp3 (MPEG-2 Audiolayer III) durchgesetzt. Für Videoformate kann zurzeit noch keine Empfehlung ausgesprochen werden. Gerade bei Videodateien liegt das Problem in der enormen Größe. Die Anforderung einer unkomprimierten Speicherung ist bei größeren Projekten kaum möglich.¹⁶

Digitale Repliken (vgl. Kapitel [2.2.5](#)) werden als 3D-Modelle von 3D-Objekten abgelegt. Diese Modelle ermöglichen es, die Form, die Textur und die optischen Materialeigenschaften des Ursprungsobjekts originalgetreu abzubilden. Infrage kommende Datenformate sollten

¹⁵ <http://www.jpeg.org/jpeg2000/>

¹⁶ Folgende Formate werden für die Langzeitarchivierung empfohlen, wobei zurzeit noch keine befriedigende Lösung vorliegt: MJPEG2000, MPEG-4, MXF/AAF

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. Formate wie Collada¹⁷, ein auf XML-basierendes offenes Austauschformat von 3D-Daten, oder X3D¹⁸, welches speziell zur Visualisierung von 3D-Modellen über Web-Browser erschaffen wurde, bieten sich an.

2.2.2 Materialspezifische Parameter

2.2.2.1 Textwerke

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

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. So empfiehlt die FADGI für Textdokumente mit dem kleinsten signifikanten Zeichen ab 1,0 mm eine Auflösung von 400 dpi und erst ab 1,5 mm Mindestzeichengröße eine Auflösung von 300 dpi.¹⁹

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

2.2.2.2 Grafische Darstellungen

Heutige digitale Aufnahmesysteme bieten eine ausreichende Auflösung, um auch feinste Details in der Ausführung einer grafischen Arbeit aufzuzeichnen. So können Informationen zu einer Originalvorlage visuell vermittelt werden, ohne das Original dazu heranzuziehen.

Wie unter [2.2.1.1](#) beschrieben liegt die empfohlene Mindestauflösung für die Digitalisierung bei 300 dpi unter Berücksichtigung der Ergebnisse einer Testdigitalisierung mit genormter Testtafel.

Häufig wird jedoch bei kleinen Vorlagen eine Auflösung von 300 dpi 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 Rechenbeispieleispiel ohne Berücksichtigung der Objektivqualität: Eine

¹⁷ Vgl. <http://www.khronos.org/>

¹⁸ Vgl. <http://www.web3d.org>

¹⁹ 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.

Briefmarke ließe sich mit einer Digitalkamera mit einem Sensor im Kleinbildformat (24 x 36 mm) im Maßstab 1:1 reproduzieren. Bei einem Pixelmaß der Kamera von 3800 Pixeln/cm ergibt sich hochgerechnet eine Auflösung von 4021 dpi.

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, sie sind oftmals auf Signalwirkung und Fernsicht ausgelegt. In diesem Fall kann die Auflösung bis auf 150 dpi reduziert werden, da mit wachsendem Betrachtungsabstand auch die Punktgröße wächst, die vom menschlichen Auge nicht mehr unterschieden werden kann. Sind Überformate jedoch detailreich wie beispielsweise Generalkarten oder Kupferstiche sollte mit mindestens 300 dpi digitalisiert werden.

Aufgrund des großen Mehrwerts für die Benutzung sollte in einem ökonomisch vertretbaren Rahmen die volle Auflösung der eingesetzten 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.

2.2.2.3 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.

Bei Diapositiven lässt sich die Qualität eines Scans visuell mit Hilfe eines Normlichtleuchtkastens mit dem Original vergleichen. Zur Farb- und Tonwertkorrektur sollte die Bilddatei in 48 Bit/RGB vorliegen. Sind keine weiteren Änderungen an der Datei vorgesehen, wird das Endergebnis mit 24 Bit/RGB gesichert. Soll vorrangig der Zustand des Vorlagenmediums zum Zeitpunkt des Scans als digitaler Master dokumentiert werden, empfiehlt sich alternativ auch eine Sicherung mit 48 Bit/RGB, besonders wenn die originale Vorlage Alterungsscheinungen (Farbausbleichungen) aufweist, die für die diversen Derivate bearbeitungsintensiv korrigiert werden sollen.

Anders verhält es sich mit Negativen: 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 Colornegativen zusätzlich durch deren Farbmaske nicht farbrichtig. Um das Motiv im digitalen Fotoabzug klar erkennbar wiederzugeben, ist daher eine beträchtliche Bildkorrektur nötig. Soll vorrangig der Zustand des Mediums zum Zeitpunkt des Scans als digitaler Master gespeichert werden, empfiehlt sich alternativ auch eine Sicherung mit 48Bit/RGB bzw. 16 Bit/Graustufen.

Bei SW-Durchsichtmedien ist die Speicherung in Graustufen ausreichend. Das Speichern von Farbinformationen bei SW-Vorlagen ist nur dann zu vertreten, wenn wichtige Informationen zu dem Medium in Farbe vorliegen, wie Retuschen im Negativ mit rotem Abdecklack.

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 [2.2.1.1](#) beschriebenen Verfahren an:

Außer dem Aufnahmeformat sind die Aufnahmebedingungen in der Regel nicht bekannt oder dokumentiert. Das verwendete Filmmaterial lässt sich möglicherweise an Hand einer Kennung, wie Kerben im Planfilm, identifizieren, bei Glasnegativen entfällt dieses Merkmal jedoch.

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.

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 dpi, gerundet 4.000 dpi. Auf das Filmmaterial bezogen wäre die theoretisch erzielbare Auflösung für alle Aufnahmeformate konstant.

Für Kleinbildaufnahmen kann durchaus eine Scanauflösung von 4.000 dpi gewählt werden, da in der Regel die Abbildungsleistung von Kleinbildoptiken besser ist als die von Mittelformat- oder Großbildobjektiven. 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 Zielgrößen:²⁰

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

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

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

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

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

Aufsichtmedien

Das Fotopositiv ist ein Endprodukt. Es dient anders als das fotografische Negativ nicht als Reproduktionsvorlage oder Ausgangspunkt für eine Vergrößerung. Die Bildqualität ist von der Negativqualität abhängig. Anders als bei Negativfilmen gibt es zum Auflösungsvermögen von Fotopapieren (gemessen in Linienpaaren pro mm, L/mm) keine verlässlichen Angaben. Das Auflösungsvermögen der fotografischen Emulsion im Positiv ist jedoch ausreichend für eine scharfe Bildwiedergabe bei Betrachtung mit bloßem Auge. Als Untergrenze sollte daher eine Scanauflösung für die Digitalisate gewählt werden, die bei Ausgabe für den Druck oder auf Fotopapier die Bildqualität der Vorlage 1:1 wiedergeben kann.

Wie unter [2.2.1.1](#) beschrieben liegt die empfohlene Mindestauflösung für die Digitalisierung bei 300 dpi unter Berücksichtigung der Ergebnisse einer Testdigitalisierung mit genormter Testtafel.

Häufig wird jedoch bei kleinen Vorlagen diese Auflösung nicht ausreichen, um die Eigenschaften der Vorlage umfassend erkennbar wiederzugeben. In der ersten Hälfte des 20. Jahrhunderts war es gebräuchliche Praxis auch von 6 x 9 cm Negativen nur Kontaktkopien zu erstellen. So empfiehlt die FADG für Fotoaufsichtmedien im Format 4 x 5 inch eine Auflösung von 800 dpi.²¹

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. Großvergrößerungen sind hier als Beispiel zu nennen, sie sind auf Fernsicht ausgelegt. Bei zu nahem Betrachtungsabstand nimmt man das Filmkorn deutlich wahr. Hier kann die Auflösung bis auf 150 dpi reduziert werden, da mit wachsendem Betrachtungstand auch die Punktgröße wächst, die vom menschlichen Auge nicht mehr unterschieden werden kann.

Soll allerdings bei Mappenwerken mit historischen Fotopositiven die Konfektionierung inklusive Präsentationskarton dokumentiert werden, ist von einer verringerten Auflösung abzusehen, damit Bildunterschriften und Aufnahmedetails identifizierbar bleiben, da bei

²⁰ Diese Werte gibt auch die FADGI vor. 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. 60.

²¹ 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. 62.

großformatigen Mappenwerken der Präsentationskarton durchaus bis zu 50 % oder mehr der Gesamtfläche des Mediums einnehmen kann.

Das Speichern von Farbinformationen bei SW-Vorlagen ist nur dann zu vertreten, wenn wichtige Informationen zu dem Medium in Farbe vorliegen, wie z.B. getonte Abzüge.

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.

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

2.2.2.4 Mikroformen

Für Mikroformen (Mikrofilme 16 und 35 mm, Mikrofiche 105x148 mm, positiv und negativ, S/W und Farbe) gelten bezüglich der Parameter die Hinweise, die im vorletzten Abschnitt für Durchsichtmedien gegeben wurden. Hierbei ist aber für die Auflösung abweichend zu berücksichtigen, dass es sich um Sicherungs- 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 dpi 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 zurzeit 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 dpi 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.

2.2.2.5 Dreidimensionale Objekte

In den Bereich der 3D-Objekte fallen sowohl museales Sammlungsgut als auch Objekte der Architektur und ihrer Ausstattung.

Die digitale Erfassung eines unikalen, dreidimensionalen Objekts (wozu z.B. auch Gemälde gehören, die durch die verwendete Maltechnik häufig Reliefcharakter haben) kann zu einer großen technischen Herausforderung werden und insbesondere Aspekte wie Materialeigenschaften des Objekts, Größe, geometrische Komplexität, Ausleuchtung und Aufnahmestandpunkte spielen für die Qualität des Resultats naturgemäß eine wesentliche Rolle.

In vielen Fällen werden Fotografien zur digitalen Repräsentation dieser Objekte erstellt, teilweise von verschiedenen Aufnahmestandpunkten aus, um diese Objekte zu dokumentieren.

Eine Fotografie ist aber immer nur eine Momentaufnahme eines Objekts aus einer Perspektive und für eine bestimmte Lichtsituation.

In den letzten Jahren wenden sich daher kulturelle Institutionen vermehrt der digitalen Replikation von Objekten in 3D zu.

Im Gegensatz zur Fotografie geht man bei der digitalen Replikation eines Objekts nun einen Schritt weiter und versucht die komplette Geometrie eines Objekts, seine Oberflächentextur und nach Möglichkeit seine optischen Materialeigenschaften zu erfassen und zu einem integrierten, digitalen 3D-Modell zusammenzuführen, welches das Objekt möglichst originalgetreu abbildet.

Ein entscheidender Vorteil gegenüber einer Fotografie ist, dass hierbei die Form des Objekts und die Oberflächen-Licht-Interaktion des Objekts originalgetreu erfasst und wiedergegeben werden können, also eine digitale 3D-Replik dieses Objekts entsteht.

Diese 3D-Replik kann nun aus beliebigen Perspektiven, in der originalen Lichtsituation sowie in beliebigen neuen Lichtsituationen visualisiert werden.

Weitere Vorteile sind:

- praktisch beliebige Verfügbarkeit und parallele Zugreifbarkeit (auch in unterschiedlicher Qualität) von digitalen 3D-Repliken von Kulturobjekten für Wissenschaftlerinnen und Wissenschaftler
- Einsatz von digitalen 3D-Modellen im Museumsbetrieb, z.B. zur Ausstellungsplanung, Dokumentation, Beschaffungsplanung, etc.
- virtuelle Präsentation (in Kombination mit neuen Präsentationstechniken, wie z.B. hybride Exponate) für die Öffentlichkeit als Mittel zur Attraktivitätssteigerung
- Referenz für die Restaurierung beschädigter Originale bzw. Generierung physischer Repliken auf Basis des digitalen 3D-Modells
- Ersatz für Ausleihe (Vermeidung von Beschädigung, Versicherungskosten, Rechtsunsicherheit zu Eigentum)

Man unterscheidet demnach folgende Digitalisierverfahren:

- Die digitale Repräsentation eines dreidimensionalen Objekts ist eine fotografische Erfassung aller relevanten visuellen Eigenschaften des Objekts, meist von mehr als einem Aufnahmestandpunkt.
- Die digitale Replik eines dreidimensionalen Objekts beinhaltet eine originalgetreue Rekonstruktion seiner Form und Oberflächen-Licht-Interaktion.

Während die fotografische Digitalisierung von dreidimensionalen Objekten ein etabliertes Verfahren auch für große Mengengerüste ist, gilt für die Erstellung von digitalen Repliken musealer Objekte allerdings, dass es insbesondere für große Stückzahlen noch wenig Erfahrung gibt. Die Qualität der Ergebnisse hängt maßgeblich von Material und Zustand der Objekte ab, so bereiten z.B. glänzende und transparente Materialien oder auch Risse und Ränder für ein effektives wie qualitätsvolles 3D-Scanning bislang Probleme. Von einer 3D-Digitalisierung in geringen Auflösungen, die vorrangig zu Publizitätszwecken geeignet sind, ist abzuraten.

Digitale Repräsentation

Häufig wird man für Sammlungsobjekte, auf jeden Fall aber für Architekturobjekte nicht mit einer einzigen Aufnahme auskommen, sondern eine Serie von Aufnahmen anfertigen. Auf Details einer dokumentarischen Bildsprache soll jedoch an dieser Stelle nicht eingegangen werden.

Für die Gemäldereproduktion sowie die Reproduktion von Textilien kann man sich an den Vorgaben für Mindestauflösungen für die grafischen Darstellungen (vgl. Kapitel [2.2.2.2](#)) orientieren, jedoch empfiehlt es sich, im Vorfeld der Digitalisierungsaufgabe zu prüfen, ob die Ausführung der Kunstwerke nicht eine höhere Auflösung verlangt, um alle Details umfänglich abzubilden.

Darüber hinaus gilt: Unterschiedliche Aufgaben in der Sammlungsfotografie benötigen objektspezifische fotografische Lösungen. Neben der notwendigen Kameratechnik spielt die geeignete objekt- und materialspezifische Beleuchtung eine maßgebliche Rolle. So muss für die Abbildung eines Reliefs eine andere Beleuchtung als für die Reproduktion eines Plakats eingesetzt werden. Dafür müssen entsprechende Lampen und Lichtformer gewählt werden.

Gewöhnlich werden Museumsobjekte unter Studiobedingungen fotografiert. Dies gewährleistet eine objektspezifische Lichtführung. Starke Schatten und Spiegelungen auf den Objekten sollten dabei vermieden werden.

Aufgrund des großen Mehrwerts für die Nutzerinnen und Nutzer sollte in einem ökonomisch vertretbaren Rahmen die volle Auflösung der eingesetzten Aufnahmetechnik bei den 3D-Objekten genutzt werden.

Im Sinne einer für die meisten wissenschaftlichen Zwecke hinreichend verlässlichen Dokumentation ist es bei dreidimensionalen Objekten oftmals nötig, mehr als nur eine Aufnahme eines Objekts anzufertigen, um Form, Funktion, Dekor und Ornament umfassend wiederzugeben. Dazu können bei Bedarf noch Detailaufnahmen kommen, wenn diese in der Gesamtansicht nicht ausreichend erkennbar sind. Für Ensembles gilt: Neben der Gesamtaufnahme einer Gruppe zusammengehöriger Gegenstände sollten diese auch einzeln fotografiert werden. Ein Anhaltspunkt für eine solche Vereinzelung ist die Granulation von Inventarnummern. Hat nur das gesamte Ensemble eine Inventarnummer und sind alle Teilobjekte annähernd gleich, ist die eine Aufnahme eines Teilobjekts – stellvertretend für die anderen – ausreichend.

Die Wahl des Hintergrunds und seiner Farbe hat Einfluss auf die Bildwirkung und somit das Erscheinen eines fotografierten Objekts. Der Hintergrund sollte in Hinsicht auf Farbe und Helligkeit so gewählt werden, dass Objektbegrenzungen nicht mit dem Hintergrund verschwimmen. Es empfiehlt sich die Verwendung eines neutralgrauen Hintergrunds, der in den meisten Fällen am wenigsten mit dem fotografierten Objekt konkurriert.

Digitale Replik

Die digitale Replik eines dreidimensionalen Objekts beinhaltet eine originalgetreue Rekonstruktion seiner Form und Oberflächen-Licht Interaktion, es werden also Geometrie, Textur und die optischen Materialeigenschaften eines Objekts erfasst.

Die vielfältigen Eigenschaften des zu erfassenden Objekts, wie etwa Größe, Texturierung, Material (Reflektivität, Transparenz etc.), kombiniert mit der Vielzahl an Erfassungsansätzen bzw. -geräten mit all ihren Stärken und Schwächen bilden einen zweidimensionalen Lösungsraum für die Wahl des optimalen Erfassungsansatzes.

Schließlich ergibt sich durch die verschiedenen angestrebten Zwecke der Erfassung mit jeweils unterschiedlichen Anforderungen ein dreidimensionaler Lösungsraum.

Keiner der aktuell verfügbaren Erfassungsansätze deckt eine der genannten Lösungsraum-Dimensionen komplett ab, daher muss für jede Kombination aus Objekt und Erfassungsziel erneut eine optimale Lösung gefunden werden.

Eine Dimension ist repräsentiert durch das Objekt mit seinen charakteristischen Eigenschaften. Eine der wichtigsten Objekteigenschaften, die Größe, grenzt bereits die in Frage kommenden Verfahren ein, eng gekoppelt an die Anforderungen.

Eine weitere wichtige Eigenschaft von Objekten ist ihr Material. So können Oberflächen komplexe Materialien aufweisen, die das Objekt transluzent oder gar transparent erscheinen lassen, und dadurch den Einsatz optischer Vermessungsansätze erschweren. Andere Eigenschaften wie etwa hohe Reflexionsgrade, die optischen Messmethoden oft Probleme bereiten, können durch den Einsatz mehrerer Sensoren gleichzeitig zuverlässig erfasst werden.

Die zweite Dimension ist definiert durch den Zweck der Digitalisierung, dieser kann von der bloßen Sicherung oder Archivierung physikalischer Objekte in Form von 3D-Modellen, über die Rekonstruktion von Kulturgütern oder ihrer Dokumentation, über die Präsentation von z.B. architektonischen Projekten in Verbindung mit bereits existierenden Teilen, bis hin zu wissenschaftlichen Zwecken reichen.

Die Menge existierender Ansätze der 3D-Digitalisierung repräsentiert die letzte Dimension.

Die gebräuchlichsten optischen Digitalisierungsverfahren sind Laser, aufgeteilt in die Laserbasierten Ansätze Lauflängenmessung, Triangulierung, und Liniencodierung, Strukturiertes Licht als Familie der 3D-Rekonstruktionsansätze basierend auf Oberflächenkodierung, und Fotogrammetrie, welche Information der Tiefe von Struktur von Objektoberflächen aus dem Vergleich und der Analyse von mehreren Bildern zieht.

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 bildbasierte Verfahren zeigen ihre Stärken.

Es 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.

Geometrie und Textur werden üblicherweise im selben Arbeitsschritt der Geometriekquise 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 Objektoberfläche kompensiert werden muss. Eine Farb-Kalibrierung ist hierbei wichtig, da Textur einen wesentlichen Einfluss auf die Realitätsnähe hat.

Ein weitgehend unabhängiger Bereich ist dagegen die Materialakquise. Grundsätzlich gibt es zwei strategische Ansätze, das Material eines Objektes zu erfassen. Einerseits die naive komplexe Erfassung der gesamten Objektoberfläche in bestimmter Auflösung und nach bestimmten Kriterien, und andererseits die Erfassung von Teilen der Oberfläche, die repräsentativ für eine komplexe Materialklasse des Objekts sind, welche wiederum die Ganzheit der vorhandenen Materialien des Objekts in ihrer Kombination erschließen. Diese Erfassung bedarf eines vorhergehenden Segmentierungsschritts der Objektoberfläche in Materialklassen, welche dann exemplarisch einzeln erfasst werden an beliebiger Stelle innerhalb des Oberflächensegments.

2.3 Metadaten

Die Erzeugung von Metadaten, welche erst die Auffindbarkeit der Objekte gewährleisten und eine kontextualisierende Präsentation ihrer digitalen Images 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 **von der 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 Metadatensatz in software-unabhängiger Form auch dann bereit steht, wenn das Projekt – aus welchen Gründen auch immer – zu einem vorzeitigen Ende kommt. Eine Projektplanung, in der die Bereitstellung herstellerunabhängiger Metadaten erst in einer späten Projektphase oder nach Projektende eingeplant wird, ist in hohem Maße problematisch. Es ist also von Anfang an darauf zu achten, dass ein Softwaresystem die Ausgabe der Daten in einem herstellerunabhängigen, sowohl semantischen als auch technischen Standardformat ermöglicht.

Wenn sich die im Rahmen eines geförderten Digitalisierungsprojekts ergebenden Materialien sachlich für die Einbindung in ein von der DFG-gefördertes Portal und/oder eine „Virtuelle Fachbibliothek“ eignen, wird erwartet, dass ein Projektantrag entweder erläutert, welche Vorehrungen 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, archivarische Erschließung, Beschreibung von unikalen, häufig zudem nicht-textuellen Objekten), strukturelle (Text-, Dokumentstruktur), administrative (z.B. Rechteverwaltung) und technische (z.B. Dateitypen) Metadaten unterschieden. Die folgenden Überlegungen beziehen sich ausschließlich auf deskriptive und strukturelle Metadaten.

Die Verknüpfung zwischen den Metadaten einerseits und den digitalen Images andererseits zu einem Objekt muss dabei immer auf der Ebene der Metadaten gewährleistet sein.

Zusätzlich können Metadaten auch in den Header der digitalen Images eingebettet werden, jedoch werden diese von den Software-Produkten unterschiedlich dargestellt und im schlimmsten Fall sogar korrumptiert, so dass die Einbettung in jedem Fall nur eine ergänzende Option ist.

Die Metadatenformate realisieren die Referenzierung der digitalen Images auf unterschiedliche Weise. Zum einen gibt es das Containerformat **METS**, innerhalb dessen Struktur die deskriptiven Metadaten in beliebigen Standardformaten (MODS, TEI) sowie die strukturellen Metadaten inklusive der Referenzen auf die digitalen Images transportiert werden können. Dies bietet sich insbesondere für Volldigitalisate von Textwerken an, die sowohl mit deskriptiven als auch strukturellen Metadaten erschlossen sind. Andere Formate wie **LIDO** enthalten zusätzlich zu den semantischen Elementen zur Objektbeschreibung eigene Elemente zur Referenzierung beliebiger digitaler Ressourcen zum Objekt (neben Images auch Audio- oder Videodaten). Auch hier können beliebig viele Ressourcen mit einem Objekt verknüpft und noch mit eigenen Beschreibungselementen versehen sein. Auch in **EAD** ist eine Verknüpfung zu beliebig vielen digitalen Ressourcen möglich. Mit **SAFT** ist eine Referenzierung nicht möglich, da es sich um ein reines Austauschformat für archivische Findmittel handelt. Wichtig ist, dass die digitalen Ressourcen innerhalb der Metadaten über global eindeutige, persistente Adressen, in der Regel URLs, referenziert werden (vgl. persistente Adressierung in Kapitel 4).

Der Nachweis der Digitalisate und Metadaten muss bei Bibliotheksgut entweder durch **Katalogisierung der elektronischen Ausgabe** oder durch Angabe der **PURL der Bilddateien** bzw. **Angabe einer persistenten Verknüpfung im Katalog** (OPAC, Verbundsystem) erreicht werden. Von Antragstellern aus universitären Einrichtungen wird erwartet, dass sie die Frage der Katalogisierung mit ihren örtlichen Bibliotheken abstimmen bzw. durch diese durchführen lassen. Der Nachweis digitalisierter Drucke im **Zentralen Verzeichnis digitalisierter Drucke (ZVDD)**²² wird erwartet. Die Digitalisate anderer bibliothekarischer Materialien sind in einschlägigen materialspezifischen Portalen nachzuweisen (z.B. Manuscripta Mediaevalia für mittelalterliche Handschriften). Materialien, die nicht in bibliothekarische Nachweisinstrumente eingebracht werden können, sollten in **geeigneten fachlichen oder fachübergreifenden Online-Anwendungen** präsentiert werden. Die Einbringung der Daten in die **Deutsche Digitale Bibliothek (DDB)**²³ sowie die **Europeana**²⁴ wird allen Projekten empfohlen.

Aus sämtlichen in übergeordnete Datenbanksysteme eingepflegten Daten oder Metadaten eines Projekts muss die Projektzugehörigkeit eindeutig hervorgehen. Es obliegt dem Datenbanksystem gegebenenfalls sicherzustellen, dass diese Informationen durch einen **Suchlink** oder andere **geeignete Filter** bzw. **Suchbedingungen** selektiert werden können, um das Arbeitsergebnis überprüfen und dem von der DFG geförderten Projekt eindeutig zuordnen zu können.

Die Ablieferung an diese Portale sollte gemäß den Standardformaten möglichst über **OAI** erfolgen (vgl. Kapitel 2.3.4).

²² <http://www.zvdd.de/>

²³ <http://www.deutsche-digitale-bibliothek.de/>

²⁴ <http://www.europeana.eu/portal/>

2.3.1 Erschließung, deskriptive Metadaten

Die Erschließung eines Textes und/oder materiellen Objekts in deskriptiven Metadaten erfüllt je nach Material unterschiedliche Funktionen für die Bereitstellung. Zunächst ermöglichen es die deskriptiven Metadaten 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.

Sofern in einem Digitalisierungsprojekt über bereits vorliegende deskriptive Metadaten hinaus eine projektspezifische Erschließung und/oder Informationsbereitstellung geplant ist – und dies wird zumindest im Bereich unikaler Objekte der Regelfall sein – sind die wichtigsten Fragestellungen an das Material zu antizipieren und in einem Kernfeldkatalog zu formulieren. Die systematische Bearbeitung der Metadaten auf der Grundlage eines Kernfeldkatalogs ist Voraussetzung für eine optimale Bereitstellung der digitalisierten Bestände gemäß Kap. 5.2.

Für eine optimale verteilte, auch nachhaltige Nutzbarkeit der Metadaten ist die Erschließung an den **einschlägigen Spartenstandards** und **Referenzmodellen** (CIDOC-CRM²⁵, perspektivisch auch IFLA FRBR/FRBRoo²⁶) zu orientieren und wo immer möglich mit **publizierten Normdaten** zu verknüpfen. Ausdrücklich wird für die Erfassung personenbezogener, biografischer Information die Verwendung der **Gemeinsamen Normdatei (GND)** der Deutschen Nationalbibliothek empfohlen. Darüber hinaus eingesetzte kontrollierte Vokabulare, z.B. **Iconclass²⁷** zur **Bilderschließung**, sollten national und international anschlussfähig sein.

Die Bereitstellung der Metadaten zur weiteren Nutzung gemäß den **materialspezifischen Standards** ist verpflichtend: **METS/MODS** für gedruckte Textwerke (s. [Anhang A](#)), **METS/TEI** für Handschriften (s. Anhang B), **EAD oder SAFTXML** für Archivmaterial²⁸, **LIDO** für (i.d.R. unikale) bildhafte und dreidimensionale Objekte (s. [Anhang C](#)). Die bereitgestellten Metadaten müssen gegen das jeweilige XML Schema valide sein und sind darüber hinaus auf semantische Korrektheit zu überprüfen.

2.3.2 Strukturelle Metadaten für digitale Faksimiles

Wohl zu erwägen ist die Frage der Anwendung von strukturellen Metadaten zur Erschließung von Images, also der Kodierung der strukturellen Elemente eines Dokumentes, wie z.B. Widmung, Vorwort, Kapitel oder Illustration. Die Aufnahme dieser Aspekte nimmt den Gedanken der analytischen Bibliografie auf, die an den Kapitel- und Textstrukturen entlang den Inhalt eines Werkes referiert. In manchen Fällen ist die Erstellung von strukturellen Metadaten von eher nachgeordneter Bedeutung, in anderen Fällen ist die Erstellung eines solchen künstlichen Inhaltsverzeichnisses für die Navigation im Digitalisat unverzichtbar. Beispielsweise ist es keinem Nutzer zuzumuten, in einem 600 Seiten umfassenden digitalen Wörterbuch online zu blättern, um den richtigen Alphabeteinstieg zu finden. Bei manchen Fragestellungen sind strukturelle Metadaten auch für die Recherche interessant. Die

²⁵ CIDOC Conceptual Reference Model: <http://cidoc-crm.org/>

²⁶ IFLA Functional Requirements for Bibliographic Records: <http://www.ifla.org/en/publications/functional-requirements-for-bibliographic-records>

²⁷ www.iconclass.nl/

²⁸ <http://www.archivschule.de/content/462.html>

Entscheidung über die Erstellung struktureller Metadaten ist daher immer eine **material- und projektspezifische**.

Falls strukturelle Metadaten Verwendung finden, wird die Konsultation der über die Website des **DFG-Viewers**²⁹ zugänglichen **Liste von Bezeichnungen** empfohlen. 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 Webseite des DFG-Viewers publizieren, um eine umfassende Nachnutzung zu ermöglichen.

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**³⁰ zu empfehlen. Bei einer Seitenbeschreibung mit einigen qualifizierenden Merkmalen (z.B. Illustrationen oder Annotationen) ist eher der von der Library of Congress gepflegte **Metadata Encoding and Transmission Standard (METS)**³¹ zu empfehlen. Für beide Modelle, die seitenorientierte und die dokumentorientierte, 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 anlangt. 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,³² so dass sich beide Aspekte gegebenenfalls auch sinnvoll verbinden lassen.

Empfohlen wird nach dem derzeitigen Stand bei alten Drucken eine Orientierung an **METS** oder **TEI**. Dessen ungeachtet soll in jedem Fall der **DFG-Viewer** unterstützt werden, der auf METS beruht. Wenn daher TEI für Strukturdaten zum Einsatz kommt, ist im Rahmen des Projekts eine Konversion zu METS notwendig. Da die Basis in beiden Fällen XML ist und die beschriebenen Sachverhalte ähnlich, kann man davon ausgehen, dass diese Konversion keine größere Hürde darstellt.

2.3.3 Sammlungs- und Bestandsbeschreibung

Sammlungs- und Bestandsbeschreibungen kontextualisieren das Einzelobjekt und sind häufig ihrerseits in ihrer historischen Verortung dokumentationswürdig, wenngleich die Abgrenzung einer Sammlung häufig nicht unproblematisch ist.³³ Schon in den klassischen Formen des bibliothekarischen und archivarischen Angebots spielte die Sammlungs- oder Bestandsbeschreibung eine wichtige Rolle, um den Nutzern einen Überblick über die Art und Zusammensetzung der in einer Altbestandsbibliothek oder einem Archiv vorhandenen

²⁹ <http://www.dfg-viewer.de>

³⁰ <http://www.tei-c.org>

³¹ <http://www.loc.gov/standards/mets>

³² http://www.tei-c.org/Sample_Manuals/bestpractice.htm

³³ Unter dem Begriff der Sammlung werde im Folgenden sowohl die in ihrer Zusammensetzung historisch zu verortende analoge Sammlung als auch die genuin digitale Sammlung (z.B. definiert über die in ein DFG gefördertes Digitalisierungsprojekt einbezogenen Bestände) verstanden (vgl. dazu die beiden Beispiele in [Anhang D](#)).

Bestände zu ermöglichen. Museumssammlungen sind per se nach Sammelgebieten, wissenschaftlichen und kuratorischen Gesichtspunkten geordnet. Die Bereitstellung einer Sammlungs- und Bestandsbeschreibung sollte folglich auch in den Online-Angeboten selbstverständlich 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. Ebenfalls empfiehlt sich ein **Wikipedia-Artikel**, sobald erste Projektergebnisse online sind. Erwartet wird darüber hinaus eine **normierte Beschreibung in XML**, um diese Informationen in Zukunft besser in nationalen oder internationalen Portalen zusammenführen und gezielt recherchieren zu können. Diese Beschreibung kann gemäß dem **Dublin Core Collections Application Profile** (s. [Anhang D](#)) oder in demselben Metadatenstandard erfolgen, in dem auch die Objektbeschreibungen verfügbar gemacht werden: **METS**, **MODS** oder **TEI-Header**, **EAD/SAFT** sowie **LIDO** bieten entsprechende Möglichkeiten. 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 ISIL einer Institution aufbaut.

2.3.4 Austausch und Weitergabe der Metadaten

Für den Aufbau einer verteilten digitalen Bibliothek ist die Schaffung eines übergreifenden **Standards zum Austausch von Metadaten** von zentraler Bedeutung (vgl. Kapitel 5). Allerdings können sich Standards nur jeweils innerhalb der jeweiligen Community ausbilden und etablieren. Dabei können ein und dieselben Ressourcen durchaus im Horizont ganz verschiedener Fragestellungen erscheinen und entsprechend divergierende Sets von Metadaten erfordern. Ein generalisiertes Verfahren zum Austausch von Metadaten muss daher flexibel unterschiedliche Metadatenformate bzw. communityabhängige Spezifikationen verwalten können. Dieses lässt sich mit dem Protokoll der **Open Archive Initiative (OAI)**³⁴ 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 Handschriften³⁵ als auch nichttextueller Objekte³⁶ 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 möglich und nötig (MARCXML, MABxml, EAD, SAFT-XML, METS/MODS, METS/TEI-HSS, TEI P5, LIDO etc.).

Die DFG setzt verpflichtend die Bereitstellung von Metadaten über **OAI** voraus. Dabei müssen neben den von OAI vorgeschriebenen **Dublin-Core-Metadaten** materialspezifische Metadaten nach **METS/MODS** für alte Drucke, **METS/TEI** für mittelalterliche Handschriften, **EAD** oder **SAFT-XML** für Archivalien und **LIDO** für bildhafte und dreidimensionale Materialien ausgeliefert werden (vgl. auch Kapitel 5). Eine Abweichung von dieser Verpflichtung muss im Antrag ausdrücklich begründet werden.

³⁴ <http://www.openarchives.org/>

³⁵ 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.

³⁶ 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.

2.4 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 Mobilität 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 digitaler Volltexte abzielen. Von allen Anträgen zur Digitalisierung textueller Materialien wird daher 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.

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.

Die Herstellung von Volltext kann auf mehreren Wegen erfolgen: entweder durch **OCR** oder durch **Abschreiben bzw. Transkription**. Die Frage, welches Verfahren gewählt werden soll, ist u.a. abhängig vom Alter und Zustand der Vorlage und der notwendigen Fehlerfreiheit des Textes für die angestrebten Ziele. Wichtig ist es, am Anfang des Projekts festzulegen, zu welchem Zweck die Volltextgenerierung durchgeführt werden soll. Dieses Ziel sollte auch während des gesamten Digitalisierungsprozesses nicht aus den Augen verloren und immer wieder berücksichtigt werden, insbesondere bei den Entscheidungen über die Frage, ob und wie bestimmte Textmerkmale erfasst werden sollen.

2.4.1 Texterfassung

Textgenauigkeiten

Unabhängig davon, ob man einen Text durch OCR 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. Ein Editionsprojekt wird allerhöchste Maßstäbe anlegen, während ein Massendigitalisierungsprojekt, das Tausende von Titeln umfasst, auf eine vergleichsweise teure manuelle Transkription verzichten muss. Was dabei noch als ausreichend gilt und welche Kosten für welche Qualität akzeptabel sind, muss in Projekten fallweise und materialabhängig 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 Antragsteller 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, müssen **statistische Verfahren** angewendet werden. Ziel ist, an Hand einer Stichprobe überprüfen zu können, 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 so genanntes **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 nicht vereinbart werden.

Bei einer behaupteten Erkennungsquote von über 99 % sollte gegebenenfalls 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

Große Orientierungswerte für Qualität auf dieser Basis sind, dass erst Texte ab einer Genauigkeit von 99,95 % als wissenschaftlich zuverlässig gelten können (also auch negative bzw. ausschließende Suchen zuverlässig möglich sind) und dass Texte je nach Betrachtungsperspektive mit einer Genauigkeit von ca. 80 bis 90 % als schlecht gelten (also zwar keine negativen, aber positive Suchen noch möglich sind). Unterhalb von 80 % scheint der Gesamtnutzen einer Konversion eher fragwürdig, im Bereich zwischen hochwertigem und schlechtem Text kommt es aber immer auf die Art des Projekts an wie auch auf die damit verbundenen Kosten.

OCR

Bisher galt, dass die OCR-Bearbeitung erst bei jüngeren Antiquaschriften ab dem 19. Jahrhundert sowie bei Frakturschriften ab der zweiten Hälfte des 19. Jahrhunderts zu akzeptablen Ergebnissen führt. Da sich der Markt anbieterseitig dynamisch weiterentwickelt und neue Produkte auf den Markt kommen, können die Praxisregeln zur Frage der OCR-Software und ihrer Anwendbarkeit zum jetzigen Zeitpunkt noch keine abschließenden Empfehlungen geben.

Dennoch zeigen erste Pilotprojekte, dass man je nach Art der eingesetzten Software und durch entsprechendes Training bzw. entsprechende Parametrisierung der OCR-Software durchaus brauchbare Resultate auch für Drucke des 16. bis 18. Jahrhunderts erzielen kann. Weitere technische Fortschritte sind zu erwarten. Bei der Beantragung von OCR-Komponenten bei Materialien aus dem genannten Zeitsegment sollten vorab Testläufe vorgenommen und perspektivische Genauigkeiten ermittelt werden, um zu prüfen, ob das Ergebnis dem gewünschten Projektziel entspricht.

Der OCR-Prozess selbst kennt verschiedene Stufen, die aufeinander aufbauen. Üblich ist zunächst die Binarisierung, d.h. ein Bild wird in ein bitonales Format übersetzt. 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-Prozess auswirken. Der Binarisierung folgt die Segmentierung, bei der die Software versucht, die jeweiligen Textbereiche auf einem Image mit Hilfe von Koordinaten zu identifizieren, um die eigentlichen Textbereiche von Illustrationen oder anderen Bildelementen zu trennen. Als besonders problematisch erweisen sich bei diesem Prozess Marginalien in älteren Drucken oder auch Zeitungen mit komplexem Layout. Danach folgt der eigentliche *pattern matching*-Prozess, der durch einen Prozess einer *intelligent character recognition* (ICR, typischerweise durch den Einsatz von Wörterbüchern) unterstützt werden kann. Theoretisch müssen Binarisierung, Segmentierung, *pattern matching* und ICR nicht unbedingt durch dieselbe Software erfolgen. In der Praxis sind diese Schritte jedoch meist in eine Software integriert.

Um Daten aus der OCR nachnutzen zu können (z.B. um Wörter in Images zu *highlighten* oder um Strukturinformationen zu extrahieren), wird die Verwendung des Standards ALTO empfohlen, der von der Library of Congress gepflegt wird.³⁷ Bei der Präsentation von mittels OCR erstelltem Volltext sollten die Nutzer stets die Möglichkeit haben, direkten Einblick in den Text zu gewinnen. Dies gilt gerade bei so genanntem „schmutzigem“ OCR. In diesen Fällen dient der Volltext nur zur Positivsuche, nicht aber als verlässliche Textbasis. Positivsuche bedeutet, dass nur positive Treffer erzielt werden. Ein negatives Ergebnis bedeutet umgekehrt nicht, dass nicht doch ein Treffer vorhanden ist. Ebenso wenig kann man sicher sein, dass mit den gefundenen Treffern alle theoretisch möglichen Treffer gefunden wurden. Ein solches Verfahren kann als Überbrückung bis zur Herstellung besserer Texte gute Dienste leisten. Es sei aber ausdrücklich vor der Gefahr gewarnt, dass Nutzerinnen und Nutzer bei mangelnder Dokumentation zu falschen Ergebnissen geführt werden.

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 99,997 % 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. Anders ausgedrückt, bei 99 % wäre jeder hundertste Buchstabe falsch, pro Zeile gäbe es also ca. einen Fehler.

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 Texts zu überprüfen.

Manuelles Erfassen ist zwar ebenfalls fehleranfällig, allerdings können mit der doppelten Eingabe der Texte und einem anschließenden Vergleichslauf zur Fehlersuche die besten Textqualitäten erreicht werden. Dieses Verfahren ist zurzeit jedoch auch das kostenintensivste. Die eigentliche Texterfassung wird hierbei zumeist im Ausland vorgenommen; der Kontakt mit einer Digitalisierungsfirma sollte jedoch über einen

³⁷ <http://www.loc.gov/standards/alto/>

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.

Da die meisten Dienstleister den Text nach Zeichenmenge inklusive des Markups berechnen, ist es ratsam, eine zeichenarme Auszeichnungsvariante³⁸ für diese Zwecke zu verwenden.

2.4.2 Zeichenkodierung

Alle verbreiteten Betriebssysteme unterstützen Unicode. Unicode ist auch das Zeichenkodierungsformat von XML, das die Grundlage für die wichtigsten Strukturdatenauszeichnungssysteme darstellt. Daher wird empfohlen, die Texte in **Unicode** abzuspeichern. Zu favorisieren ist das bei europäischen Sprachen sparsamere **UTF-8**. Zeichen, die nicht im Unicode-Standard enthalten sind, können durch Nutzung des *private plane*-Bereichs von Unicode³⁹ abgebildet und durch entsprechende Grafiken oder Fonts repräsentiert werden. 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 in der *encoding description* eines TEI-Header zu dokumentieren sind.

2.4.3 Markup von Volltexten

Wenn nicht triftige Gründe dagegen sprechen, müssen **Volltexte von Drucken und Handschriften** nach dem Modell der **Text Encoding Initiative (TEI)**⁴⁰ kodiert bzw. mit **Markup versehen** werden. Als transparentes XML-Format ist TEI, sofern sorgfältig dokumentiert, auch für die Langzeitarchivierung die prospektiv beste Wahl. Wenngleich mit der ISO-Norm 19005-1:2005 bzw. 19005-2:2011 (PDF/A), die ein langfristiges sicheres Subset von PDF bildet, ein PDF-Format zur Verfügung steht, das den Anspruch erhebt, für die Langzeitarchivierung geeignet zu sein, sollte aus prinzipiellen Erwägungen bei Volltextdigitalisierungsprojekten davon kein Gebrauch gemacht werden. Dies gilt zumindest in den Fällen, in denen die Herstellung des Texts in der Hand der Projektnehmer liegt.⁴¹ PDF ist ein Format, das im Wesentlichen Layoutinformationen speichert und mit dem Ziel entwickelt wurde, Dateien am Bildschirm und auf dem Drucker originalgetreu wiedergeben zu können, wobei originalgetreu meint, dass analoge Quellen auf dem Bildschirm simuliert werden. XML-

³⁸ Z.B. teitite: http://www.tei-c.org/release/doc/tei-p5-exemplars/html/tei_tite.doc.html

³⁹ <http://de.wikipedia.org/wiki/Unicode>

⁴⁰ <http://www.tei-c.org>

⁴¹ In anderen Fällen, in denen Texte bereits in PDF vorliegen, ist eine Umwandlung nach PDF/A erforderlich. Geprüft werden sollte, ob zur besseren Nachnutzbarkeit der Texte eine Wandlung in XML erfolgen kann. Dies ist in der Regel nur mit Abstrichen bei den strukturellen Details möglich.

typische Strukturen lassen sich mit PDF nur bedingt oder nur über Umwege einer Layoutanalyse realisieren und machen PDF daher für viele Funktionen einer rein digitalen Repräsentation von Volltexten ungeeignet. Dies betrifft auch und vor allem Funktionen des *Semantic Web*, das für die aufstrebenden *Digital Humanities* eine zunehmend wichtige Rolle spielt. 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. Kapitel, Unterkapitel, Jahresbände, Aufsätze etc. berücksichtigt. Hinzu kommen weitere denkbare Strukturmerkmale, z.B. Inhaltsverzeichnisse, Register, Zeilenumbruch, Spaltenumbruch, Seitenumbruch, Kopfzeile/Fußzeile/Kolumnentitel, Seitenzahl, Bilder oder bildähnliche Elemente, Bildunterschriften, Marginalien, Schriftwechsel, z.B. der Wechsel von Fraktur zu Antiqua (etwa für fremdsprachige Zitate), der Wechsel der Schriftgröße, Wechsel der Schriftart (recte, kursiv, fett usw.) u.a.m., Formeln, z.B. mathematische (MathML) oder chemische (CML) Formeln, Fortsetzungsmarkierungen (Kustoden) am Fußende von Seiten (für Anschlussbogen) etc.

Die Wahl des Markups 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.

2.4.4 Layout

In manchen Fällen ist es bei der Präsentation eines Volltexts wichtig, das Layout eines Dokumentes langfristig zu sichern. Die Praxisregeln empfehlen für diese Fälle bevorzugt den Einsatz einer **geeigneten Formatierungssprache** (z.B. XSLT, XSL-FO, XQuery, 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 [2.4.2.](#) dargelegt, kein Ersatz für eine Bereitstellung des mit Markup versehenen Volltextes in XML.

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 sind z.B. HTML/XHTML, PDF, ePub oder plain text.

2.5 Langzeitverfügbarkeit

Für die **Langzeitsicherung** und **Archivierung** digitaler Inhalte gibt es zum jetzigen Zeitpunkt keine verallgemeinerbare Lösung, die für alle Objekt- und Materialgattungen geeignet ist. Im Rahmen der Langzeitsicherung werden Dateien in stabilen Formaten auf einem technisch wie organisatorisch sicheren Speichersystem gesichert. Die Archivierung elektronischer Daten setzt auf einem solchen Speicher auf, beinhaltet jedoch noch weitreichendere technische und

organisatorische Festlegungen, die nicht nur eine physische Erhaltung der Daten, sondern Strategien zur Erhaltung der Interpretierbarkeit und Unveränderbarkeit einbeziehen.

Für die Langzeitarchivierung sind besonders technische Informationen und solche über die Veränderungshistorie eines Objekts wichtig.⁴² Insbesondere, wenn es um die Veränderungshistorie eines Objekts geht, hat sich PREMIS⁴³ als Datenmodell etabliert.

Ob die erzeugten digitalen Inhalte langzeitgesichert oder archiviert werden, hängt von der Strategie der jeweiligen Institution ab, die mit einer Antragstellung erläutert werden muss.⁴⁴ Wesentliche Erfolgsfaktoren für eine Langzeitsicherung und Archivierung 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 elektronischer Daten ist das Open Archival Information System (OAIS) als Referenzmodell anzuwenden.⁴⁵ Die „Kriterien für vertrauenswürdige digitale Langzeitarchive“ sind essenziell. Mit ihnen werden u.a. der organisatorische Rahmen, gesetzlichen Rahmenbedingungen, das Qualitätsmanagement sowie die Authentizität für ein vertrauenswürdiges Archiv geschaffen.⁴⁶

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.

Derzeit spielen bei der digitalen Sicherung vor allem zwei Trägermaterialien eine Rolle:

- (1) Bandlaufwerke (*streamer*),
- (2) Festplatten(-systeme)

(Zu 1) Die Sicherung auf Bändern hat noch immer Vorteile (z.B. geringe Stromkosten) aber auch den Nachteil, dass Bänder relativ langsam sind. Wenn häufige Zugriffe auf die digitalen

⁴² http://www.langzeitarchivierung.de/Subsites/nestor/DE/Standardisierung/Metadaten.html;jsessionid=5239B697C2522475FBEC_D446877A114F.prod-worker5

⁴³ http://www.loc.gov/standards/premis/understanding_premis_german.pdf

⁴⁴ 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.

⁴⁵ Das OAIS-Referenzmodell ist als ISO 14721-Standard verabschiedet: <http://public.ccsds.org/publications/archive/650x0m2.pdf>

⁴⁶ 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.

Master benötigt werden, sind Bänder keine gute Lösung. Sie müssen gegebenenfalls durch andere Zwischenspeicher ergänzt werden. Bänder müssen darüber hinaus regelmäßig bewegt werden, um ein Zusammenkleben zu verhindern. Wenn ein Bandarchivsystem genutzt werden soll, wird dringend empfohlen, das System von einem Rechenzentrum betreiben zulassen.

(Zu 2) Besonders zu empfehlen im Sinne einer Migrationsstrategie ist die redundante Datenhaltung auf Festplattensystemen. Die redundante Datenhaltung erfolgt in Form sogenannter *network attached storage* (NAS)-Systeme, des *storage attached network* (SAN) oder der *content addressed storage* (CAS) im Rechenzentrumsbetrieb. Dabei ist eine redundante Sicherung auf zwei unabhängigen Systemen notwendig.

In allen Fällen ist eine Replikation auf zwei physikalisch unabhängigen Systemen notwendig, wobei durchaus das gleiche Trägermaterial für beide Sicherungen zum Einsatz kommen kann. Mögliche Varianten sind festplattenbasierte Systeme, ein Bandsystem oder auch die Ausbelichtung auf hochauflösendem Farbmikrofilm. Festplattensysteme erlauben den schnellen und unkomplizierten Zugriff auf die Daten und erleichtern nötige Migrationen.

Es sei darauf hingewiesen, dass Digitalisierungsprojekte aus Sicht der DFG stets **Projekte der gesamten Einrichtung** sind: Die Unterstützung der das Projekt abwickelnden 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.

Da die Langzeitarchivierung digitaler Medien ein bisher noch nicht abschließend diskutierter Bereich ist und die Konzepte sowie die angewandte Technik stetig voranschreiten, ist es nötig, sich jeweils einen aktuellen Überblick über dieses Thema zu verschaffen. Als Einstiegspunkt in das Thema eignet sich das **Kompetenznetzwerk Nestor**⁴⁷.

3. Organisatorische Fragen

Die Digitalisierung kann *inhouse* oder durch Beauftragung eines Dienstleisters erfolgen. Im ersten Fall können Geräte, aber auch projektbedingte Personalkosten, bei der DFG beantragt werden. Gerade bei größeren Mengen oder bei konservatorisch problematischen Materialien ist es manchmal aus Sach- und Finanzgründen günstiger, die Arbeiten im eigenen Haus mit eigenen Geräten und eigenem Personal durchzuführen. In anderen Fällen ist die Beauftragung eines Dienstleisters wieder die bessere und kostengünstigere Wahl. Schließlich kann es in Einzelfällen sinnvoll sein, Personal- oder Sachmittel zu beantragen, die dann aber an einem anderen Standort, beispielsweise einer großen Bibliothek mit einschlägiger Digitalisierungskompetenz, zum Einsatz kommen.

Die **Entscheidung** darüber, ob ein Vorhaben als **Inhouse-Projekt** konzipiert oder ob auf das **Angebot eines Dienstleisters** zurückgegriffen wird, ist daher immer eine **projektspezifische Entscheidung**, die allein in die Verantwortung der Antragsteller fällt. Schließlich ist das

⁴⁷ <http://www.langzeitarchivierung.de/>

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 Prüfung der von einem Dienstleister gelieferten Qualität ist meist nicht einfach. Bei der Filmdigitalisierung ist oft nicht klar festzustellen, ob eine mangelhafte Digitalisierung auf einen schlechten Film oder ungenügende Scantechnik des Dienstleisters zurückzuführen ist. Um eine angemessene Bewertung der gelieferten Farbtreue zu erleichtern, ist insbesondere Wert darauf zu legen, dass bei einer Digitalisierung vom Original möglichst Farbkeile (i.d.R. einmal repräsentativ für eine Bildserie mit identischen Einstellungen) mit fotografiert werden.

Bisherige Erfahrungen haben gezeigt, dass die Vergabe von Teilaufgaben eines Digitalisierungsprojektes an externe Dienstleister sehr effektiv sein kann, aber keineswegs immer ist. Grundsätzlich ist festzuhalten, dass auch der konsequente Einsatz von Dienstleistern die bestandshaltende Einrichtung nicht der Aufgabe enthebt, Kompetenzen im Bereich der Digitalisierung im Hause bereitzuhalten – allein schon deshalb, weil nur dann kompetente Vertragsverhandlungen mit Dienstleistern möglich sind. Schließlich sei nochmals hervorgehoben, dass die DFG davon ausgeht, dass die Digitalisate langfristig innerhalb der Infrastruktur der Häuser gepflegt werden. In der Antrag stellenden Institution selbst müssen daher ausreichende Personal- und Sachmittel für die Projektleitung und die Projektkontrolle sowie die Auswahl und Vorbereitung der Materialien zur Digitalisierung bereitstehen. Des Weiteren werden Arbeiten zur spezifischen Erschließung und Aufbereitung der digitalisierten Materialien nahezu immer in den antragstellenden Einrichtungen selbst durchgeführt.

Liegen diese Bedingungen vor, wird die Beauftragung externer Dienstleister häufig vorteilhaft sein. Die Arbeit mit einem Dienstleister ist auch dann möglich, wenn die zu digitalisierenden Materialien nicht außer Haus gegeben werden können, da bei größeren Projekten die Dienstleister mit eigenen Geräten und Personal vor Ort arbeiten. Auch nachgelagerte Produktionsstufen, wie die Bereitstellung der Daten im Netz, Vertrieb und Vermarktung von Offlineprodukten, spezialisierte Druckzentrums-dienstleistungen oder die Migration zur Langzeitarchivierung, können durch Fremdfirmen manchmal wirtschaftlicher erfolgen als durch die antragstellenden Einrichtungen selbst.

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 durchgeföhrten 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.

4. 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 selbstständiges Objekt in einem sich dynamisch entwickelnden integralen Forschungsraum. Während also die klassische Fotokopie als Kopie zu vernachlässigen ist, bedarf die digitale Kopie als Kopie eigener Zitationsregeln, die im Übrigen nur für das Internet gelten. Versionen auf CD oder anderen nicht vernetzten Speichermedien lassen sich wie Filme oder Papierkopien behandeln und zitieren. Wird aber eine Kopie ins Internet gestellt, muss sie, um von anderen Dokumenten 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 klassische Form weitergenutzt werden kann und sollte, die Festlegung und die online zugängliche Dokumentation von **Adressierungstechniken** erforderlich.

Ein positiver Effekt der netzbasierten Zitierform, die sich in der Regel an der physischen Bildfolge orientieren wird, ist, dass das **Referenzieren eindeutig** wird, was man wegen der vielen Fehler in alten Drucken oder auch aufgrund des Mangels an Paginierungen oder Folierungen 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 Images eindeutig zitieren, die nicht zum eigentlichen Korpus gehören (Deckel, Spiegel, zusätzliche digitalisierte Wasserzeichen, Teilaufnahmen von Illustrationen etc.). Einzige Voraussetzung dafür ist, dass sich ein bestimmtes Bild in einer alphanumerisch aufsteigenden Sequenz eindeutig verorten lässt (z.B. 00001, 00002, 00002a, 00003 etc.; hier wurde ein Bild 00002a eingefügt). Zitierziel ist hierbei ein Image.

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).

Im Grundsatz sollte immer die größtmögliche Granularität angestrebt werden. Vor allem zwei Funktionalitäten sind bei der Netzpräsentation zu nennen. Zum einen die Erreichbarkeit des Werks als Ganzem, zum anderen die Erreichbarkeit einzelner Seiten oder Doppelseiten aus diesem Werk. Die Struktur eines „Zitats“ könnte wie folgt aussehen (fingiert):

<http://digitalebibliothek.ubique.de?titelid=234&image=0002>

oder aber

<http://digitalebibliothek.ubique.de/titel234/image0002/>

Sichergestellt werden muss die Erreichbarkeit und Zitierbarkeit eines Werkes als Ganzem und die Erreichbarkeit und Zitierbarkeit von einzelnen physischen Seiten von 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.

Die Erzeugung von **URNs** über die Deutsche Nationalbibliothek mindestens auf Werkebene wird nachdrücklich empfohlen.⁴⁸

5. Bereitstellung der Digitalisate für die Öffentlichkeit

5.1 Rechteklärung und Open Access

Die Deutsche Forschungsgemeinschaft ist Mitunterzeichnerin der „Berlin Declaration“ zum *open access*. Die Ergebnisse der durch sie geförderten Digitalisierungsprojekte stehen der weltweiten Forschung daher unentgeltlich und im Geiste dieser Erklärung zur Verfügung.⁴⁹ Insofern wird grundsätzlich davon ausgegangen, dass Digitalisate in einer Qualität, die die Verfolgung der großen Mehrzahl der üblichen Forschungszwecke erlaubt, kostenfrei über das Internet bereitgestellt werden. Sollte in Projekten davon abgewichen werden, ist dies im Antrag ausdrücklich zu begründen.

Die Digitalisierung von wissenschaftlichen 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.

Bereits mit Beginn der Projektplanung, spätestens aber zum Zeitpunkt der Antragstellung muss eine **Rechteklärung** bezüglich der zu digitalisierenden Materialien erfolgt sein. Insbesondere eventuell vorliegende **Urheber-, Persönlichkeits- und Leistungsschutzrechte** müssen dabei berücksichtigt werden. Die Rechteklärung gestaltet sich häufig aufwendiger als zunächst erwartet. Für die entsprechenden Aufwände ist daher ausreichend Zeit und personeller Einsatz einzuplanen. Stellt sich die Rechteklärung als schwierig dar, wird empfohlen, professionelle Unterstützung, beispielsweise durch das Justiziariat der bestandshaltenden Einrichtung zu suchen.

Im Sinne des **open access** und der **open source**⁵⁰ wird erwartet, dass alle Metadaten (deskriptive und strukturelle) und maschinenlesbaren Volltexte bzw. das den Volltexten zu Grunde liegende XML sowie XSLT-Skripte und DTDs oder XML-Schema-Dateien unter einer **Creative Commons-Lizenz** angeboten werden. Maßgebliches Ziel ist es, Möglichkeiten zu schaffen, Metadaten und Volltexte auch in anderen Nutzungskontexten als der unmittelbaren Projektumgebung (z.B. zum Zwecke des *data mining* oder der Datenaggregation z.B. im ZVDD) auswerten zu können; dies erfordert das Herunterladen, die Neuindexierung und das Anbieten in eigenen Forschungs- und Präsentationskontexten. Um die Leistung des Urhebers angemessen zu würdigen, wird empfohlen diese Texte unter einer CC BY-SA-Lizenz zur

⁴⁸ <http://www.persistent-identifier.de/>

⁴⁹ <http://oa.mpg.de/openaccess-berlin/berlindeclaration.html>

⁵⁰ Vgl. Bodard, Gabriel/Juan Garcés: Open Source Critical Editions: A Rationale, in: Marilyn Deegan/Kathryn Sutherland (Hrsg.): Text editing, print and the digital world., Farnham: Ashton Publishing 2009, S.83-98.

Verfügung zu stellen. Bei Metadaten ist zu prüfen, ob sie unter einer CC0-Lizenz zur Verfügung gestellt werden können.

Darüber hinaus sollten Images in einer Form bereitgestellt werden, die eine **vollumfängliche wissenschaftliche Nachnutzung** in anderen Forschungskontexten erlaubt (z.B. durch Nutzung von einzelnen Images „in Frames“ nach dem Modell des DFG-Viewers bzw. Googles iFrame-Einbettung oder durch generelle Erteilung einer Erlaubnis, Vorschaubilder auf fremden Servern anzeigen zu dürfen). Die Erhebung von Reproduktionsgebühren oder Entgelten für Kopien in darüber hinausgehenden 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 **einheitliche Herkunftsangaben sowie gegebenenfalls den Hinweis auf die Förderung durch die DFG** anbringen. Dies erfolgt im Fall von Bilddigitalisaten in der Regel durch das **Hinzurechnen einer Nachweisleiste in die veröffentlichte Nutzerkopie** (zum Beispiel im Grafikformat JPEG). 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 muss die Herkunftsangabe in den begleitenden Metadaten enthalten sein und im Bereitstellungssystem auch ausgegeben werden.

Bei Projekten, in denen nicht nur gemeinfreie Materialien digitalisiert werden und daher mit kommerziellen Partnern oder Verlagen zusammengearbeitet wird, kann eine **verzögerte Publikation (moving wall) von bis zu einem Jahr nach Projektende** vereinbart werden.

5.2 Mindestanforderungen für die Bereitstellungssysteme von Digitalisaten

5.2.1 Funktionalitätsanforderungen

Unabhängig von der gewählten Architektur müssen mindestens folgende Funktionalitäten bereitgestellt werden:

Sammlungen/Bestände sind in der Regel über eine Vielzahl von Wegen zugänglich:

- über die **Website** der anbietenden Institution
- über eine **OAI-Schnittstelle**
- über den lokal implementierten oder an einem anderen Ort betriebenen „**DFG-Viewer**“, sofern für das angebotene Material zutreffend (vgl. [Kapitel 6.3](#))
- über eine Suchanfrage an den **lokalen und regionalen Bibliothekskatalog/an die jeweilige materialspezifische Online-Anwendung**
- 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 Portal der „Virtuellen Fachbibliotheken“
- über **Internet-Suchmaschinen**.

Über die genannte **OAI-Funktionalität** ist auch der **DFG-Viewer** zu bedienen (sofern das jeweilige Format unterstützt wird). Eine Ablieferung lediglich durch XML-Dateien sollte zugunsten der OAI- Funktionalitäten nicht mehr erfolgen. Neu eingerichtete OAI-Schnittstellen

sollten in Frage kommenden Portalen und Fachbibliotheken gemeldet werden. Ferner sollten geeignete Maßnahmen ergriffen werden, die dazu führen, dass die Metadaten von Suchmaschinen gefunden werden (z.B. mit dem *sitemap protocol*⁵¹).

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 einfache, googleorientierte Suchinstrumente einer größeren Benutzergemeinschaft dienen 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:

- Downloadfunktion⁵²
- Druckfunktion für die ausgegebene Dokumentansicht⁵³
- Aus zentralen DFG-geförderten Informationssystemen (VD16, VD17, VD18, Fachportalen etc.) sollte zuerst auf eine Ansicht im Style „DFG-Viewer“ verlinkt werden.
- Für sämtliche DFG-geförderten Bereitstellungssysteme wird erwartet, dass Nutzerinnen und Nutzer die **automatisierte Möglichkeit zum Feedback** bezüglich des digitalen Angebots erhalten. Eine entsprechende Möglichkeit ist auf der Projektseite bzw. im digitalen Bereitstellungssystem an zentraler Stelle einzurichten.

5.2.2 Technische Mindestanforderungen

Server sind so anzulegen, dass sie, soweit anwendbar⁵⁴,

- alle Materialien in einer hinreichend guten Qualität anbieten, 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 aber auch in einer Qualität anbieten, die über DSL-Anschlüsse ohne unzumutbare Verzögerungen bearbeitbar sind.
- zum wissenschaftlichen Gebrauch den kostenlosen Download kompletter Einheiten in einer einzigen Datei (beispielsweise einzelne gedruckte Werke) ermöglichen.

⁵¹ <http://www.sitemaps.org/>

⁵² Ein Download nach Abschnitten oder einzelnen Seiten ist dann vorzusehen, wenn die Größe der Gesamdatei nicht mehr handhabbar wäre.

⁵³ Eine Druckfunktion nach Abschnitten oder einzelnen Seiten ist dann vorzusehen, wenn die Größe der Gesamdatei nicht mehr handhabbar wäre.

⁵⁴ 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.

- alle jeweils gängigen Browser unterstützen, soweit dies der Natur der Sache nach nicht unmöglich ist.⁵⁵

6. 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 Dokumenten haben, gelten zusätzlich folgende Mindestanforderungen. Diese betreffen einige Basisanforderungen und die mindestens zu realisierende Funktionalität.

6.1 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 analog zu einer Freihandbibliothek in sachlicher Aufstellung zu bieten. 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 Bibliotheken und Archive sollten sowohl bei der Navigation in digitalen Sammlungen/Beständen als auch bei der Recherche über Indizes und Register einen institutionenübergreifenden Zugriff ermöglichen. Darüber hinaus ist eine transparente Verknüpfung der Bereitstellungssysteme mit lokalen Katalogsystemen und Verbunddatenbanken wünschenswert.

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

- Die Metadaten werden zentral in einem Katalogsystem (z. B. dem lokalen OPAC oder einem Bibliotheksverbundkatalog) gehalten, die digitalen Dokumentdateien (inkl. elektronischem Inhaltsverzeichnis und Register) werden in einem hierarchisch gegliederten Dateisystem 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.

⁵⁵ 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.

6.2 Funktionalitätsanforderungen

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

- Ansteuern eines beliebigen Images
- Anfang: Springen an den Anfang eines Dokuments
- Ende: Springen an das Ende eines Dokuments
- Vor: Eine Seite vorgehen
- Zurück: Eine Seite zurückgehen
- Volltextrecherche (für Bücher ab 1850 verpflichtend)⁵⁶
- Metadaten-Info: Hier können Nutzerinnen und Nutzer die Informationen aus den im DMS 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 Server Materialien, die konzeptuell in der Benutzung üblicherweise zu übergeordneten Einheiten zusammengefasst werden (mehrständige Werke), so sollen diese übergeordneten Einheiten als solche sichtbar werden.

6.3 DFG-Viewer

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-gefördeter 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“)
- Schaffung einer OAI-Schnittstelle mit Unterstützung des METS-Standards: Diese Schnittstelle dient zunächst der einheitlichen Anzeige von Images und ihren Metadaten in DFG-geförderten Projekten und transportiert materialabhängig METS/MODS für Drucke, METS/TEI für Handschriften und METS/EAD für Archivmaterialien. Über diese Schnittstelle werden Funktionen des Weiterblätterns, der Anzeige von Metadaten und andere grundsätzliche Funktionen beschrieben und damit einheitliche Anzeige- und Blätterfunktionen geschaffen, die auch innerhalb von zentralen Nachweisportalen einen homogenen Zugriff auf dezentrale Ressourcen (*harvesting*) erlauben.

Zum Zweck der überregionalen Erstpräsentation müssen DFG-geförderte Digitalisierungsvorhaben die genannten Schnittstellen bedienen und/oder den Viewer im DFG-Style an ihrer Einrichtung selbst implementieren. Die Einrichtung einer OAI-Schnittstelle bleibt

⁵⁶ Nach dem Stand der gegenwärtigen Technik ist eine OCR-Erkennung bei Drucken der Maschinenpressenzeit ab 1850 verpflichtend.

davon unberührt. [Anhang A](#) definiert das METS/MODS-Format der Schnittstelle sowie das Viewer-Design.

7. Prüfliste für Antragsteller und Gutachter

Wie aus den bisherigen Kapiteln ersichtlich 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 folgende Prüfliste ist daher folgendermaßen zu verstehen:

- Bei allen Anträgen mit Digitalisierungskomponenten werden unabhängig von ihrer inhaltlichen Begutachtung die vorgeschlagenen technischen Konzepte geprüft.
- Darüber hinaus wird geprüft, ob die gegebenenfalls notwendigen Rechteklärungen erfolgt sind, die eine vollumfängliche Nachnutzbarkeit der Digitalisate erlauben.
- Antragsteller müssen in ihren Anträgen nachvollziehbar belegen, dass die Durchführung des Projekts die im Folgenden aufgelisteten Standards einhält.
- Jede Abweichung von diesen Standards muss ausführlich begründet werden.
- Zur besseren Vergleichbarkeit sollen Anträge (bzw. Arbeitsberichte) Auskünfte über die geschätzten (bzw. tatsächlichen) Kosten für das Scannen inklusive Strukturdatenerfassung pro Image, Kosten für die Volltexterfassung sowie Kosten für projektspezifische Speicherkapazitäten enthalten.⁵⁷
- Sieht ein Digitalisierungsprojekt vor, die im Folgenden angeführten Standards zu überschreiten, so ist, soweit dadurch erhöhte Kosten entstehen, deren Notwendigkeit im Einzelnen zu begründen.

⁵⁷ Vgl. dazu das Statistikblatt „Digitalisierung“ unter <http://www.dfg.de/lis>, das in der Rubrik „Formulare und Merkblätter“ unter „Ergänzende Formulare und Formblätter LIS“ zugänglich ist. Zur Berechnung der realen Kosten gibt es zurzeit noch keine allgemein akzeptierten Standards. Die Praxisregeln empfehlen daher, die Kosten des Scanvorgangs inklusive Strukturdatenerfassung folgendermaßen zu errechnen:

- (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, DVDs für den 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 Arbeitsverlauf (Workflow) erst bestimmt wird. Dazu nötige Pilotstudien sind vor der Antragstellung abzuschließen.
- Bei der Digitalisierung von Material der kulturellen Überlieferung findet eine insgesamt gut verstandene Technik Verwendung. Es ist daher davon auszugehen, dass die dabei auftretenden Kosten laufend sinken.

7.1 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:

- Ob die vorgesehene Personalausstattung hinreichend und notwendig ist. 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.
- Ob die vorgesehenen Bearbeitungszeiten realistisch sind. Soweit die vorgesehenen Bearbeitungszeiten nicht unmittelbar plausibel sind, sind sie entweder durch die Erfahrungen in abgeschlossenen, vergleichbaren Projekten oder durch die Ergebnisse selbst vorgenommener Vorabtests zu belegen.

7.2 Technische Parameter der digitalen Reproduktion

Ziel der Digitalisierung ist die möglichst originalgetreue Wiedergabe des Materials nach Maßgabe der wissenschaftlichen Erfordernisse. Für alle im Rahmen der Projekte zu digitalisierenden Materialien sind die beantragten Qualitäten so zu wählen, dass Darstellungen, die unmittelbar öffentlich zugänglich gemacht werden, ohne menschliche Intervention im Rahmen des Projekts durch Batchprozesse realisiert werden können. Die folgenden Richtlinien für die Mindestanforderungen an die Digitalisierung beziehen sich ausschließlich auf den **digitalen Master** und stellen materialübergreifende Empfehlungen dar. Zu materialspezifischen Empfehlungen, die über diese Mindestanforderungen hinausgehen, vergleiche Kapitel [2.2.2](#).

- Die Auflösung ist so zu wählen, dass in der Auflösung der Archivkopien die kleinsten relevanten Details klar zu sehen sind, wenn die Datei auf ein Viertel der Ausgangsgröße verkleinert wird. Grundsätzlich wird eine Auflösung von 300 dpi empfohlen. (→ [2.2.1.1](#))
- Für die Sicherung des digitalen Master ist eine Farbtiefe von 8 Bit pro Kanal, d.h. 24 Bit ausreichend. (→ [2.2.1.2](#))
- 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 JPEG2000 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 lizenzenfreien Bereiche

von JPEG2000 Verwendung finden. Zu den mit JPEG2000 einhergehenden Risiken bezüglich der Langzeitarchivierung vergleiche Kapitel 2.2.1.4. (→ [2.2.1.4](#))

7.3 Metadaten

Metadaten sind grundsätzlich in einer von der Software unabhängigen und standardkonformen Form bereit zu stellen, in aller Regel in einer XML-Codierung.

- 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. Ausdrücklich wird für die Erfassung personenbezogener, biografischer Information die Verwendung der Gemeinsamen Normdatei (GND) der Deutschen Nationalbibliothek empfohlen. Darüber hinaus eingesetzte kontrollierte Vokabulare, z.B. Iconclass zur Bilderschließung, sollten national und international anschlussfähig sein. (→ [2.3.1](#))
 - Die Bereitstellung der deskriptiven Metadaten zur weiteren Nutzung gemäß den materialspezifischen Standards ist verpflichtend (→ [2.3.1](#) sowie Anhänge A und B):
 - METS/MODS für gedruckte Textwerke
 - METS/TEI für Handschriften
 - EAD oder SAFTXML für Archivmaterial
 - LIDO für (i.d.R. unikale) bildhafte und dreidimensionale Objekte
- Die bereitgestellten Metadaten müssen gegen das jeweilige XML Schema valide sein und sind darüber hinaus auf semantische Korrektheit zu überprüfen.
- Die Publikation der deskriptiven Metadaten über eine OAI-Schnittstelle ist verpflichtend, wahlweise im eigenen System oder über ein geeignetes Zielportal. (→ [2.3.1](#))
 - Der Nachweis der Digitalisate und Metadaten muss bei Bibliotheksgut entweder durch Katalogisierung der elektronischen Ausgabe oder durch Angabe der PURL der Bilddateien bzw. Angabe einer persistenten Verknüpfung im Katalog (OPAC, Verbundsystem) erreicht werden. Von Antragstellern aus universitären Einrichtungen wird erwartet, dass sie die Frage der Katalogisierung mit ihren örtlichen Bibliotheken abstimmen bzw. durch diese durchführen lassen. Der Nachweis digitalisierter Drucke im Zentralen Verzeichnis digitalisierter Drucke (ZVDD)⁵⁸ wird erwartet. Materialien, die nicht in bibliothekarische Nachweisinstrumente eingebracht werden können, sollten in geeigneten fachlichen oder fachübergreifenden Online-Anwendungen präsentiert werden. Die Einbringung der Daten in die Deutsche Digitale Bibliothek (DDB) sowie die Europeana wird allen Projekten empfohlen. (→ [2.3.1](#))
 - Aus sämtlichen in übergeordnete Datenbanksysteme eingepflegten Daten oder Metadaten eines Projekts muss die Projektzugehörigkeit eindeutig hervorgehen. Es obliegt dem Datenbanksystem gegebenenfalls sicherzustellen, dass diese Informationen durch einen Suchlink oder andere geeignete Filter bzw. Suchbedingungen selektiert werden können, um das Arbeitsergebnis überprüfen und dem von der DFG geförderten Projekt eindeutig zuordnen zu können. (→ [2.3](#))
 - Die Ablieferung an die genannten Portale sollte gemäß den Standardformaten möglichst über OAI erfolgen. (→ [2.3.1](#))

⁵⁸ <http://www.zvdd.de/>

- Die Entscheidung über die Erstellung struktureller Metadaten ist immer eine material- und projektspezifische. Falls strukturelle Metadaten Verwendung finden, wird die Konsultation der über die Website des DFG-Viewers zugänglichen Liste von Bezeichnungen erwartet. Empfohlen wird nach dem derzeitigen Stand eine Orientierung an METS oder TEI, wobei der DFG-Viewer bei textuellen Materialien auf jeden Fall unterstützt werden muss. (→ [2.3.2](#))
- Es wird erwartet, dass Digitalisierungsprojekte mindestens Gegenstand und Umfang der jeweiligen Material- bzw. Objektauswahl auf einer Seite im Netz, möglichst auch in Englisch, darstellen. Ebenfalls empfiehlt sich ein Wikipedia-Artikel, sobald erste Projektergebnisse online sind. Erwartet wird darüber hinaus eine normierte Sammlungsbeschreibung (*collection level description*) in XML, um diese Informationen in Zukunft besser in nationalen oder internationalen Portalen zusammenführen und gezielt recherchieren zu können. Diese Beschreibung kann gemäß dem Dublin Core Collections Application Profile oder in demselben Metadatenstandard erfolgen, in dem auch die Objektbeschreibungen verfügbar gemacht werden: METS, MODS oder TEI-Header, EAD/SAFT sowie LIDO bieten entsprechende Möglichkeiten. (→ [2.3.3](#) sowie [Anhang D](#))

7.4 Volltextgenerierung

Von allen Anträgen zur Digitalisierung textueller Materialien 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. (→ [2.4](#))

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. (→ [2.4](#))

- Die Herstellung von Volltext kann auf zwei Wegen erfolgen: entweder durch OCR oder durch Abschreiben bzw. Transkription. (→ [2.4.1](#))
- Um eine gewisse Einheitlichkeit bei der Beurteilung der Textgenauigkeit zugrunde legen zu können, werden Antragsteller gebeten, diesbezügliche Angaben hinsichtlich der Buchstabengenauigkeit zu machen, d.h. mangelhafte Worttrennungen und Layoutfehler unberücksichtigt zu lassen. Stichproben sollten sich an dem in 2.4.1 beschriebenen statistischen Verfahren orientieren. (→ [2.4.1](#))
- Bezuglich der Zeichencodierung wird empfohlen, die Texte in Unicode abzuspeichern. Zu favorisieren ist UTF-8. (→ [2.4.2](#))
- Wenn nicht triftige Gründe dagegen sprechen, müssen Volltexte von Drucken und Handschriften nach dem Modell der Text Encoding Initiative (TEI) kodiert bzw. mit Markup versehen werden. (→ [2.4.3](#))
- In manchen Fällen ist es bei der Präsentation eines Volltextes wichtig, das Layout eines Dokumentes 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 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

- 2.4.2. dargelegt, kein Ersatz für eine Bereitstellung der mit Markup versehene Volltext in XML. (→ [2.4.4](#))
- 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 unter einer Creative Commons-Lizenz angeboten werden. (→ [5.1](#))

7.5 Langzeitverfügbarkeit

Für die Langzeitsicherung und Archivierung digitaler Inhalte gibt es zum jetzigen Zeitpunkt keine verallgemeinerbare Lösung, die für alle Objekt- und Materialgattungen geeignet ist. Im Rahmen der Langzeitsicherung werden Dateien in stabilen Formaten auf einem technisch wie organisatorisch sicheren Speichersystem gesichert. Die Archivierung elektronischer Daten setzt auf einem solchen Speicher auf, beinhaltet jedoch noch weitreichendere technische und organisatorische Festlegungen, die nicht nur eine physische Erhaltung der Daten, sondern Strategien zur Erhaltung der Interpretierbarkeit und Unveränderbarkeit einbeziehen. (→ [2.5](#))

Die Langfristverfügbarkeit 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. (→ [2.5](#))

Es sei darauf hingewiesen, dass Digitalisierungsprojekte aus Sicht der DFG stets Projekte der gesamten Einrichtung sind: Die Unterstützung der das Projekt abwickelnden Fachabteilung durch die IT-Infrastruktur des Hauses wird vorausgesetzt. Dabei wird begrüßt, wenn sich kleinere Einrichtungen der Kompetenz und der Dienstleistung größerer Einrichtungen bedienen. (→ [2.5](#))

Anträge müssen nachvollziehbare Aussagen zur institutionellen Langzeitsicherung und Archivierung enthalten. Digitalisate sind in redundanter Form zu archivieren. (→ [2.5](#))

Da die Langzeitarchivierung digitaler Medien ein bisher noch nicht abschließend diskutierter Bereich ist und die Konzepte sowie die angewandte Technik stetig voranschreiten, ist es nötig, sich jeweils einen aktuellen Überblick über dieses Thema zu verschaffen. Als Einstiegspunkt in das Thema eignet sich das Kompetenznetzwerk Nestor. (→ [2.5](#))

7.6 Organisatorische Fragen

Die Entscheidung darüber, ob ein Vorhaben als *Inhouse*-Projekt konzipiert oder ob auf das Angebot eines Dienstleisters zurückgegriffen wird, ist immer eine projektspezifische Entscheidung, die allein in die Verantwortung der Antragsteller fällt. Bei der Zusammenarbeit mit Dienstleistern muss die projektnehmende Einrichtung deutlich machen, dass sie ein Projekt kompetent begleiten kann. In Verträgen müssen Leistungen exakt festgelegt 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. (→ [3](#))

7.7 Zitieren, persistente Adressierung

Digitalisate müssen, um von anderen Dokumenten 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 klassische Form weitergenutzt werden kann und sollte, die Festlegung und die online zugängliche Dokumentation von Adressierungstechniken erforderlich. Sichergestellt werden muss die Erreichbarkeit und Zitierbarkeit eines Werkes als Ganzem und die Erreichbarkeit und Zitierbarkeit von einzelnen physischen Seiten von 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. Die Erzeugung von URNs über die Deutsche Nationalbibliothek mindestens auf Werkebene wird nachdrücklich empfohlen. (→ [4](#)).

7.8 Bereitstellung der Digitalisate für die Öffentlichkeit

Die Digitalisierung von wissenschaftlichen Materialien wird durch die DFG gefördert, um diese Materialien der Forschung in Deutschland und weltweit zugänglich zu machen. Alle Projekte sind so anzulegen, dass die Ergebnisse der Forschung frühzeitig und nachhaltig zugänglich werden.

- Bereits mit Beginn der Projektplanung, spätestens aber zum Zeitpunkt der Antragstellung muss eine Rechteklärung bezüglich der zu digitalisierenden Materialien erfolgt sein. Insbesondere eventuell vorliegende Urheber-, Persönlichkeits- und Leistungsschutzrechte müssen dabei berücksichtigt werden. (→ [5.1](#))
- Im Sinne des *open access* und der *open source* wird erwartet, dass alle Metadaten (deskriptive und strukturelle) und maschinenlesbaren Volltexte bzw. das den Volltexten zu Grunde liegende XML sowie XSLT-Skripte und DTDs oder XML-Schema-Dateien unter einer Creative Commons-Lizenz angeboten werden. (→ [5.1](#))
- Darüber hinaus sollten Images in einer Form bereit gestellt werden, die eine voluminöse wissenschaftliche Nachnutzung in anderen Forschungskontexten erlaubt. (→ [5.1](#))
- Bei Projekten, in denen nicht nur gemeinfreie Materialien digitalisiert werden und daher mit kommerziellen Partnern oder Verlagen zusammengearbeitet wird, kann eine verzögerte Publikation (*moving wall*) von bis zu einem Jahr nach Projektende vereinbart werden. (→ [5.1](#))

Digitalisate sollen im Regelfall über unterschiedliche Wege zugänglich sein: (→ [5.2.1](#))

- über die Website der anbietenden Institution
- über eine OAI-Schnittstelle
- über den lokal implementierten oder an einem anderen Ort betriebenen „DFG-Viewer“, sofern für das angebotene Material zutreffend
- über eine Suchanfrage an den lokalen und regionalen Bibliothekskatalog / an die jeweilige materialspezifische Online-Anwendung
- 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 Portal der „Virtuellen Fachbibliotheken“

- über Internet-Suchmaschinen

Alle Materialien sind in einer für wissenschaftliche Zwecke hinreichend guten Qualität anzubieten und mit intuitiv zu bedienenden Navigationsmöglichkeiten zu versehen, so dass die Benutzung für die fragliche Zielgruppe und auf an den Hochschulen üblicherweise vorhandenem Equipment mühelos möglich ist. Alle jeweils gängigen Browser müssen unterstützt werden, soweit dies der Natur der Sache nach nicht unmöglich ist. (→ [5.2.2](#) und [6.2](#))

Für sämtliche DFG geförderte Bereitstellungssysteme wird erwartet, dass Nutzer die automatisierte Möglichkeit zum Feedback bezüglich des digitalen Angebots erhalten. Eine entsprechende Möglichkeit ist auf der Projektseite bzw. im digitalen Bereitstellungssystem an zentraler Stelle einzurichten. (→ [5.2.1](#))

Die DFG erwartet, dass DFG geförderte Projekte auf ihren im Internet angebotenen Daten eindeutige Herkunftsangaben sowie gegebenenfalls den Hinweis auf die Förderung durch die DFG anbringen. Dies erfolgt im Fall von Bilddigitalisaten in der Regel durch das Hinzurechnen einer Nachweisleiste in die veröffentlichte Nutzerkopie (zum Beispiel im Grafikformat JPEG). (→ [5.1](#))

Im Regelfall – wenn die Natur der Digitalisate einen Teil dieses Leistungsspektrums nicht implizit ausschließt – wird erwartet, dass ein Digitalisierungsprojekt Pläne für folgende öffentlich zugängliche Schnittstellen bereitstellt:

- Einen selbständigen Server, der das digitalisierte Material mit zu seiner Nutzung notwendigen Instrumenten bereitstellt.
- Alle Digitalisate sind so anzubieten, dass langfristig zitierfähige URLs möglichst niedriger Granularität entstehen. Die Form, in der zitiert werden soll, ist deutlich anzuzeigen.
- Eine Schnittstelle im technischen Sinn, die einem für den jeweiligen Materialbereich von der DFG geförderten Portal den Zugriff auf alle im Rahmen des Projektes erstellten Metadaten ermöglicht, insbesondere eine OAI-Schnittstelle, die neben Dublin Core METS/MODS-, METS/TEI- oder METS/EAD-Metadaten an geeignete Harvester liefert.
- Geeignete Maßnahmen, die dazu führen, dass die Metadaten von Suchmaschinen gefunden werden.

8. Regeln für die Durchführung von Digitalisierungsprojekten

Zwischen der Förderzusage der DFG und dem tatsächlichen Arbeitsbeginn vergeht üblicherweise eine gewisse Zeit, während der die Hardware gekauft, Verträge mit den Dienstleistern geschlossen und die Kräfte des Projekts rekrutiert werden. Die in den Projektanträgen vorgesehenen Arbeitspläne gehen dagegen üblicherweise davon aus, dass das Projektteam vollständig und arbeitsfähig ist. Um den Verlauf der Projekte besser begleiten und Fehlentwicklungen besser gegensteuern zu können, wird daher erwartet, dass die Projektnehmer die beschriebenen vorbereitenden Tätigkeiten in einer Vorbereitungsphase durchführen, während derer wenig oder keine Personalmittel ausgegeben werden. Sind diese Vorbereitungen abgeschlossen, ist der DFG der tatsächliche Projektbeginn schriftlich

mitzuteilen. Alle weiteren Fristen zählen ab diesem Zeitpunkt. Zwischen der Förderzusage und dem tatsächlichen Projektbeginn darf in der Regel nicht mehr als ein Jahr verstreichen.

Alle Projekte sind verpflichtet, zu einem Zeitpunkt, da eine Gegensteuerung gegen Fehlentwicklungen noch möglich ist, ein funktionsfähiges Modell des Internetangebotes zu präsentieren (in der Regel nach Ablauf des ersten Jahres der Förderung). Dieses Modell muss den Nachweis erbringen, dass die oben angeführten Mindestansprüche eingehalten werden. Alle Komponenten des technischen Angebotes müssen zu diesem Zeitpunkt prinzipiell funktionsfähig sein. Entstehen Rückfragen oder Zweifel, ist das Internetangebot gegebenenfalls vor Ort zu demonstrieren.

Erst nachdem der Server in seinen Grundfunktionalitäten abgenommen ist und monierte Mängel abgestellt sind, sind Verlängerungsanträge möglich.

Als Bestandteil aller Berichte ist mit dem Statistikblatt Digitalisierung anzugeben:

- Ein wie großer Bestandteil des Mengengerüsts tatsächlich digitalisiert wurde.
- Ein wie großer Bestandteil des Mengengerüsts im Internet bereitsteht.
- Welche aussagekräftigen Zugriffszahlen sich aus den Logdateien des Servers ablesen lassen.

Wesentliche Unterschreitungen der im Projektantrag beschriebenen Arbeitsgeschwindigkeit sind häufig ein Zeichen ungenügender Projektorganisation und stellen keine erfolgversprechende Begründung für Verlängerungsanträge dar. Sind sie signifikant, ist auch die Förderung bereits ursprünglich angekündigter Projektphasen kritisch zu überprüfen.