

# 7.4 Empfehlungen für größere Einrichtungen

## Auswahl des zu archivierenden Materials

Selektionskriterien bilden die Entscheidungsgrundlage für die Frage, ob ein Objekt langfristig gesichert werden soll. Hierbei spielen neben rechtlichen auch inhaltliche Kriterien eine Rolle. Besondere Bedeutung kommt der Frage nach den Ressourcen zu.

### Empfehlung

*Größere Institutionen sollten anhand ihres Sammlungsprofils und ihres Mandats selektieren:*

- *Ist das Objekt archivwürdig?*
- *Entspricht das Objekt dem Sammlungsprofil bzw. dem Archivauftrag?*
- *Liegen alle benötigten Rechte vor (Langzeitarchivierung, Veränderung, Verbreitung)?*

## Beschreibung der Nutzungsszenarien

Die vorgesehenen Nutzerzielgruppen sind die aktuellen und künftigen Nutzergruppen, die mit dem digitalen Langzeitarchiv oder daran angeschlossenen Nutzungsplattformen interagieren. Unter Nutzerzielgruppen sind an dieser Stelle nicht nur die klassischen Nutzer zu verstehen. Vorgesehene Nutzerzielgruppen sind auch externe Nutzer wie Produzenten, die Objekte abliefern, und interne Nutzer wie die Mitarbeiter der Institution.

Verschiedene vorgesehene Nutzerzielgruppen können unterschiedliche Anforderungen haben. Die vorgesehenen Nutzerzielgruppen entsprechen den klassischen Zielgruppen der Institution. Diese sind bekannt und deren Anforderungen gelegentlich bereits analysiert worden. Zusätzlich kommen als vorgesehene Nutzerzielgruppen hinzu:

- die Datenproduzenten
- die mit der digitalen Langzeitarchivierung befassten Mitarbeiter

Die verschiedenen vorgesehene Nutzerzielgruppen haben bestimmte Anforderungen. Daraus ergeben sich Nutzungsszenarien und signifikante Eigenschaften, die Einfluss auf die Erhaltungsmaßnahmen nehmen.

### Empfehlung

*Die Nutzerzielgruppen müssen der Institution bekannt sein.*

*Eine Analyse der Nutzerzielgruppen sollte durchgeführt werden, die neben den klassischen Zielgruppen der Institution auch die Datenproduzenten und die Mitarbeiter berücksichtigt.*

Die Analyse der Nutzerzielgruppen sollte folgende Punkte beinhalten:

- die Beschreibung und Abgrenzung der verschiedenen Nutzerzielgruppen
- die Beschreibung der Anforderungen der Nutzerzielgruppen
- in den Nutzerzielgruppen verbreitete Nutzerformate

Die Nutzerzielgruppen sollten in wiederkehrenden Abständen erneut analysiert werden, da sich die Anforderungen im Lauf der Zeit ändern können.

## Qualitätsmanagement

Qualitätsmanagement bezeichnet Verfahren und Prozesse, die die Qualität der Objekte sowie der Prozesse im digitalen Langzeitarchiv sicherstellen sollen.

Darunter fallen z.B.:

- Übernahmevereinbarung (ein Dokument, das organisatorische, rechtliche und technische Punkte bei der Übernahme berücksichtigt)
- Eingangskontrolle der Objekte (Überprüfung der Objekte beim Eingang in das digitale Langzeitarchiv)
- Prüfung auf Lesbarkeit und inhaltliche Vollständigkeit
- Prüfung auf inhaltliche Übereinstimmung mit den deskriptiven Metadaten
- technische Fehlerfreiheit
- Standardkonformität
- Dokumentation von Prozessen (Genauere Beschreibung von Prozessen im digitalen Langzeitarchiv: Was wird wann warum wie gemacht? Z.B. schriftlich oder in einer Notation für Prozessdokumentation wie BPMN 2.0.)
- Evaluierung von Prozessen (Dokumentierte Prozesse können nach Kriterien wie Praxistauglichkeit, Zeitaufwand, Ressourcenaufwand, Überschneidungen mit anderen Prozessen usw. untersucht werden. Die Analyse ist Voraussetzung für die Optimierung der Prozesse.)

- Kontrollverfahren (technische und organisatorische Verfahren für die Kontrolle von definierten Regeln und Strukturen oder zur Evaluierung von Ergebnissen nach Änderungen an Objekten)
- Regelkonformität von Transfer- und Archivpaketen
- Qualitätsprüfung nach Formatmigrationen
- Lesbarkeit und Vollständigkeit nach Änderungen am Objekt
- Nutzung von Standards
- Beachtung der Kriterien für die vertrauenswürdige Langzeitarchivierung (nestor-Siegel nach DIN 31644, auch Data Seal of Approval)

## Empfehlung

*Maßnahmen für das Qualitätsmanagement müssen implementiert werden. Es wird empfohlen, eine Eingangskontrolle durchzuführen und Kontrollverfahren einzurichten. Darüber hinaus sollten Standards genutzt und die Kriterien für die vertrauenswürdige Langzeitarchivierung beachtet werden. Prozesse sollten dokumentiert und evaluiert, die Dokumentation sollte regelmäßig auf Aktualität geprüft werden. Die Eingangskontrolle sollte eine (stichprobenartige) Sichtprüfung des Materials und die Prüfung der Übereinstimmung mit den deskriptiven Metadaten beinhalten. Zusätzlich sollten die Objekte identifiziert werden (siehe Tabelle mit Tools, S. 110 ff.).*

*Wenn möglich, sollten die Produzenten die Objekte mit einer Prüfsumme an das Archiv abgeben. Diese Prüfsumme muss nach der Annahme im Archiv geprüft werden. Stimmen die Prüfsummen nicht überein, muss das Objekt abgewiesen und vom Produzenten erneut abgegeben werden, bis ein erfolgreicher Transfer mit übereinstimmenden Prüfsummen erfolgt ist.*

*Die Kontrollverfahren können teilautomatisiert werden. Die Prüfung auf Lesbarkeit und Vollständigkeit nach Formatmigrationen sollte (stichprobenartig) manuell erfolgen.*

*Darüber hinaus muss nach Bestandserhaltungsmaßnahmen geprüft werden, ob die signifikanten Eigenschaften erhalten geblieben sind.*

*Die Prüfung auf Regelkonformität von Transferpaketen kann automatisiert erfolgen, indem ein entsprechendes Prüfverfahren bei der Übernahme implementiert wird.*

*Die Nutzung von Standards bezieht sich auf das Standard-Referenzmodell OAIS, auf Metadatenstandards (z.B. PREMIS, Dublin Core, METS) sowie Dateiformatstandards. Darüber hinaus können darunter auch standardisierte Prozesse und einheitliche Paketstrukturen verstanden werden.*

*Das Langzeitarchivierungssystem muss in der Lage sein, regelkonforme Archivpakete zu erstellen.*

*Einheitlich strukturierte Daten und definierte Prozesse erleichtern nicht nur die digitale Langzeitarchivierung, sondern jegliche Art von Datenverwaltung.*

*Darüber hinaus sollte die Institution den Kriterien für die vertrauenswürdige Langzeitarchivierung folgen und eine Zertifizierung des digitalen Langzeitarchivs anstreben. Zusätzlich zur öffentlichen Sichtbarkeit, dass die digitale Langzeitarchivierung definierten Qualitätsansprüchen genügt, erhält die Institution so auch Gelegenheit, ihre Prozesse, Verfahren und Strukturen von externen Experten überprüfen zu lassen und zu optimieren.*

## Erhaltungsstrategien

Erhaltungsstrategien stellen sicher, dass das Objekt als Datenstrom erhalten und nutzbar bleibt. Dabei sichern Verfahren für die Bitstream-Preservation den Datenstrom und Verfahren für die Content-Preservation die Les- und Interpretierbarkeit des Objekts.

## Bitstream-Preservation

Digitale Datenträger sind nur von begrenzter Haltbarkeit. Sie sind nicht nur von Verschleiß betroffen, auch Hardwareobsoleszenz ist problematisch. Aus diesem Grund ist ein Speicherkonzept essenziell für die Erhaltung des Bitstroms der digitalen Objekte.

Grundlegende Anforderungen an das Speicherkonzept sind:

- redundante, geografisch verteilte Speicherlösung
- Integritätschecks, z.B. mit Prüfsummen
- Viruscheck
- Datenschutzkonzept
- regelmäßiger Austausch von Datenträgern
- Monitoring
- Skalierbarkeit

Die Sicherung des Bitstroms eines Objekts ist ein stetig wiederkehrender Prozess und Basis jeder Erhaltungsstrategie.

## Empfehlung

Größere Institutionen können allein oder im Verbund ein Rechenzentrum betreiben oder einen Dienstleister mit der Bitstream-Preservation beauftragen. Diese Entscheidung muss im Einzelfall getroffen werden.

Die institutionsinterne Bitstream-Preservation sollte die folgenden Anforderungen erfüllen:

- redundante, räumlich getrennte Speicherung auf verschiedenen Speichermedien
- Integritätschecks mit Prüfsummen

- Viruschecks
- Datenschutzkonzept
- regelmäßiger Austausch von Datenträgern
- Monitoring
- Skalierbarkeit

Prüfsummen können für ein gesamtes Objekt oder mit einem Tool, z.B. FFmpeg, auf Framebasis erstellt werden.

Es sollte eine professionelle Speicherlösung gewählt werden. Es wird zu einer Kombination aus Festplatten- und Bandspeicher geraten.

Darüber hinaus müssen Sicherheitsbestimmungen beachtet werden. Darunter fallen:

- physischer Zugang
- Zugriffsberechtigungen
- Einbruchmeldeanlage
- digitale Zugriffsberechtigungen
- Brandschutz
- Disaster Recovery

Mit einem Dienstleister sollte ein Service-Level-Agreement (SLA) vereinbart werden. Der Umfang der Dienstleistung muss genau evaluiert werden, um sicherzustellen, dass alle Anforderungen an die Bitstream-Preservation erfüllt werden.

## Content-Preservation

Content-Preservation soll sicherstellen, dass der Inhalt eines Objekts auch nach Jahren wiedergegeben und interpretiert werden kann.

Dafür ist es erforderlich:

- das Dateiformat, die -version und den Codec zu identifizieren und zu validieren. Bei AV-Medien wird diese Erkenntnis durch den Codec erschwert, der in dem Containerformat enthalten ist
- technische Metadaten zu extrahieren (siehe Tabelle mit Tools, S. 110ff.)
- Informationen über geeignete Wiedergabesoftware zu sammeln
- Risikofaktoren zu identifizieren
- signifikante Eigenschaften festzulegen
- Kontextinformationen bereitzustellen, die die Wiedergabe des Objekts und die Interpretation des Inhalts ermöglichen

Auf Basis dieser Informationen können Bestandserhaltungsmaßnahmen konzipiert und durchgeführt werden. Dafür gibt es zwei Verfahren: Formatmigration und Emulation.

## Empfehlung

*Große Institutionen sollten eine Erhaltungsstrategie ausarbeiten. Die Erhaltungsstrategie sollte folgende Punkte abdecken:*

- *Formatidentifikation und -validierung. Daraus folgend das Bilden von homogenen Bestandsgruppen.*
- *Risikofaktoren identifizieren*
- *Turnus für Überprüfung der Risikofaktoren festlegen*
- *Erhaltungsmaßnahmen für Bestandsgruppen festlegen*
- *geeignete Tools für die Erhaltungsmaßnahmen definieren*
- *Policies nach dem SCAPE Policy Framework erstellen*

Es wird die Formatmigration empfohlen, da es für die meisten audiovisuellen Objekte nicht erforderlich ist, die ursprüngliche Nutzungsumgebung zu erhalten. Eine Ausnahme stellt Videokunst dar, die im Wissenschaftlichen Film aber nicht als solche vorkommt.

Das Dateiformat sollte identifiziert und validiert werden. Es genügt nicht, ein Objekt über die Dateierdung zu identifizieren, da diese auch irreführend sein kann. Außerdem sollten technische Metadaten extrahiert werden.

Das Dateiformat und der Codec eines Objekts müssen bekannt sein, da alle Erhaltungsmaßnahmen hierauf basieren. Proprietäre Dateiformate sollten, wenn es möglich ist, nach dem Eingang der Originaldatei kopiert werden. Diese Kopie sollte in ein offenes, standardisiertes Archivformat konvertiert werden. Die Originaldatei soll unverändert erhalten bleiben.

Signifikante Eigenschaften sollten definiert werden. Der Fokus sollte auf den für die Institution und deren Nutzer wichtigen Eigenschaften eines Objektes liegen.

Signifikante Eigenschaften hängen u.a. auch davon ab, welchen Zweck die Aufnahme hat. Ist der Zweck der Aufnahme z.B. die Dokumentation eines Ereignisses, eines Experiments usw., so ergibt sich daraus, was besonders erhaltenswert ist. Davon wiederum hängen ggf. weitere technische signifikante Eigenschaften ab.

Diese Anforderungen sollten nach einer Bestandserhaltungsmaßnahme überprüft werden.

Beispiele für technische signifikante Eigenschaften:

Eigenschaft	Komponente	Definition der Eigenschaft
Videokanäle	video	Anzahl der unterschiedlichen Videokanäle im digitalen Objekt
Audiokanäle	audio	Anzahl der unterschiedlichen Audiokanäle im digitalen Objekt
Länge	video	Anzahl der Frames in der Aufnahme
Breite	frame	Breite des Einzelbildes
Höhe	frame	Höhe des Einzelbildes
Bittiefe	pixel	Anzahl der Farbinformationen pro Pixel
Farbmodell	pixel	verwendetes Farbmodell bei der Aufnahme
Farbraum	pixel	verwendeter Farbraum bei der Aufnahme
Seitenverhältnis	pixel	Seitenverhältnis der Pixel
Bildrate	video	Anzahl der Frames in einem definierten Zeitintervall
Interlace	frame	falls es sich um interlaced video handelt
Metadaten		zum Objekt gehörende Metadaten

(Quelle: Archivemata Wiki)

Grundlegende Risikofaktoren sollten definiert werden, z.B.:

- das Format oder der Codec ist von Obsoleszenz bedroht
- das Objekt liegt in einem proprietären oder kaum verbreiteten Format oder Codec vor

Tritt einer der Risikofaktoren auf, sollten alle Objekte im betroffenen Format in ein neues Zielformat migriert werden.

## Authentizität

Die Sicherung der Authentizität bedeutet, nachweisen zu können, dass ein Objekt ist, was es zu sein vorgibt. Da Bestandserhaltungsmaßnahmen wie Formatmigrationen zwangsläufig eine Veränderung eines Objekts verursachen, ist die Sicherung der Authentizität nicht ganz unproblematisch. Folgende Maßnahmen sind zur Sicherung der Authentizität etabliert:

- die Aufbewahrung des Originals und aller Bearbeitungen des Objekts (Versionierung)
- der Nachweis von Veränderungen am Objekt in den Metadaten

## Empfehlung

Größere Institutionen sollten das Originalobjekt sowie alle Versionen des Originalobjekts aufbewahren und Änderungen in den Metadaten dokumentieren. Die Protokollierung von Änderungen muss technisch und organisatorisch sichergestellt sein. Diese beinhaltet z.B.:

- Art und Zeitpunkt der Änderung
- Name des Benutzers, der die Änderung veranlasst hat
- Name des Benutzers oder des Programms, der/das die Änderung durchgeführt hat
- Grund der Änderung

## Metadaten

Metadaten beschreiben ein Objekt. Es wird zwischen folgenden Metadatenkategorien unterschieden:  
**deskriptive Metadaten:** Es handelt sich um klassische Erschließungsmetadaten. Diese beschreiben ein Objekt und gewährleisten, dass es wiederauffindbar ist. Erschließungsmetadaten unterteilen sich in formale Metadaten wie Titel, Urheber, Erscheinungsjahr usw. und inhaltsbeschreibende Metadaten wie Schlagworte, ein Abstract, ein Transkript, Sequenzbeschreibungen usw.

**administrative Metadaten:** Administrative Metadaten sind für die Verwaltung der Objekte im Langzeitarchiv und den Betrieb des Systems erforderlich, z.B. systeminterne Identifier, Nutzerdaten, Rollen- und Rechtevergabe.

**rechtliche Metadaten:** Rechtliche Metadaten beschreiben den Nutzerzugriff auf die Objekte. Sie legen auch fest, ob die Objekte bearbeitet und wie sie bereitgestellt werden dürfen.

**technische Metadaten:** Technische Metadaten beschreiben z.B. Dateiformat und -version, den Codec,

Laufzeit, Dateigröße, Bitrate usw. Technische Metadaten können aus der Datei extrahiert werden.  
**strukturelle Metadaten:** Strukturelle Metadaten beschreiben die Zusammensetzung eines Objekts und Zusammenhänge mit anderen, z.B. wenn ein Objekt aus mehreren Dateien in einer festgelegten Reihenfolge besteht oder wenn Objekte Teil einer Serie sind.  
**Provenienzmetadaten:** Diese Metadaten beschreiben die Objektgeschichte, um zu dokumentieren, welcher Nutzer wann mit welchen Tools welche Bearbeitung vorgenommen hat. Diese Metadaten sind von großer Wichtigkeit für den Nachweis der Authentizität.

## **Empfehlung**

*Größere Institutionen müssen Metadaten aller Kategorien erfassen. Die Erschließung der Objekte wird für gewöhnlich in einem Nachweissystem erfolgen. Dort können neben deskriptiven Metadaten abhängig vom System auch technische, rechtliche und strukturelle Metadaten erfasst werden. Diese Metadaten sollten vollständig oder in Auswahl in das Langzeitarchivierungssystem überführt werden, das in der Lage ist, administrative und Provenienzmetadaten zu erzeugen und zu verwalten.*

## **Erschließung**

Neben den formalen Metadaten sollten auch inhaltsbeschreibende Schlagworte verwendet werden. Die Erschließungstiefe ist abhängig von den Vorgaben der Institution und den Anforderungen der Nutzer. Wenn das Nachweissystem der Institution für die Erschließung und Verwaltung von Multimediaelementen geeignet ist, können die audiovisuellen Objekte mit dem Nachweissystem verwaltet und beschrieben werden. Ist das vorhandene Nachweissystem nicht geeignet, sollte frei verfügbare Datenbanksoftware verwendet werden, die für Multimediaobjekte geeignet ist. Nachweissysteme sind in der Regel nicht für die Langzeitarchivierung geeignet. Es kommt auch ein Digital Asset Management System (DAM)/Media Asset Management System (MAM) in Betracht. Ein DAM/MAM ist allerdings i.d.R. nicht für die Langzeitarchivierung geeignet. Ob das System für die Langzeitarchivierung geeignet ist, muss im Einzelfall untersucht werden. Für den Nachweis der Objekte und die Erfassung von deskriptiven und strukturellen Metadaten sind Nachweissysteme oder DAM/MAM-Systeme jedoch gut geeignet. Ob rechtliche und technische Metadaten erfasst werden können, muss im Einzelfall geprüft werden.

## **Provenienz- und Erhaltungsmetadaten**

Es muss ein OAIS-konformes Langzeitarchivierungssystem eingesetzt werden, das in der Lage ist, alle erforderlichen Metadaten zu verwalten. Bereits vorhandene Metadaten, z.B. deskriptive und rechtliche Metadaten, können aus dem bestehenden Nachweissystem vollständig oder in Auswahl in das digitale Langzeitarchivierungssystem über eine Schnittstelle importiert werden. Im Langzeitarchivierungssystem sollten alle relevanten Metadaten in einer METS/MODS-Datei gespeichert werden, um eine schnellere Verarbeitung in Fällen der Formatmigration oder Metadaten-Updates zu gewährleisten.

## **Präsentation/Zugriff**

Das Format der Nutzungskopien sollte die Anforderungen der vorgesehenen Nutzerzielgruppen berücksichtigen. Weitverbreitete, offene Formate sollten bevorzugt werden; proprietäre Formate sind häufig an eine spezielle Wiedergabesoftware gebunden, binden an einen Hersteller und begrenzen die Nutzung. Im Idealfall soll die Nutzungskopie ohne die Installation zusätzlicher Software lesbar sein. Für verschiedene Nutzungsszenarien können verschiedene Nutzungsformate erforderlich sein.

## **Empfehlung**

*Größere Institutionen sollten Nutzungskopien erzeugen und für die Nutzung vorhalten. Das Format der Nutzungskopie sollte den Anforderungen der Nutzerzielgruppen entsprechen. Es muss beachtet werden, wie die Objekte für die Nutzung zur Verfügung gestellt werden dürfen. Folgende Nutzungsszenarien sind denkbar:*

- lokale Nutzung
- der Versand von Datenträgern
- Onlinezugriff mit und ohne Registrierung

*Für die lokale Nutzung muss ein Format gewählt werden, das in der betreffenden Umgebung wiedergegeben werden kann. Beim Versand von Datenträgern ist die Größe des Nutzungsformats ggf. durch die Speicherkapazität begrenzt. In diesem Fall kann eine Kompression erforderlich sein. Den Webzugriff muss die Institution technisch bewältigen können: Da die Datei-größe die Übertragungsdauer beeinflusst, ist für den Webzugriff die Kompression des Objekts erforderlich. Außerdem kann es erforderlich sein, dass der Zugriff auf angemeldete Benutzer oder nur bestimmte Nutzergruppen beschränkt werden muss.*

## **Dateiformate und Codecs**

Für den Archivmaster sind offene, standardisierte und weitverbreitete Formate zu bevorzugen, die nicht oder verlustfrei komprimieren, um einem späteren Qualitätsverlust vorzubeugen. Proprietäre Dateiformate sind oft an eine spezifische Software eines Herstellers gebunden und stellen eine potenzielle Gefahr für die langfristige Verfügbarkeit dar, da der Code nicht öffentlich zugänglich ist und Formatmigrationen so erschwert oder sogar unmöglich werden. Zu beachten ist bei der Wahl des Archivmasters auch die Kompatibilität eines Formats und Codecs. Darüber hinaus muss basierend auf den Anforderungen der vorgesehenen Nutzerzielgruppen auch eine Entscheidung über die Formate für Nutzungskopien getroffen werden. Eine Zusammenstellung empfohlener Formate und Codecs finden Sie auf den Seiten der netor AG media unter [AG Media](#)

## **Empfehlung**

*Die große Vielfalt potenzieller Ablieferungsformate stellt die Institutionen vor eine große Herausforderung. Je höher die Anzahl der Dateiformate, desto größer ist der Aufwand für die Bestandserhaltung (Preservation Planning, Preservation Watch, Preservation Actions, Risikomanagement usw.).*

*Es ist deshalb sinnvoll, die Anzahl der Dateiformate im digitalen Langzeitarchiv möglichst gering zu halten, sofern die Institution Einfluss darauf nehmen kann. Dies kann durch folgende Verfahren geschehen:*

- *die Veröffentlichung einer Liste bevorzugter Dateiformate*
- *das Erzeugen einer neuen Version durch Formatmigration nach dem Eingang proprietärer Objekte. Die Originaldateien müssen dabei unverändert aufbewahrt werden.*
- *eine Kombination beider Verfahren*

*Die Vorgehensweise sollte in einer öffentlich verfügbaren Policy dokumentiert werden, um den Prozess für Produzenten und Nutzer transparent zu machen.*

*Als Format empfiehlt sich Matroska (.mkv), ffv1-codiert, Audio: Wave Linear PCM 48 kHz. Die empfohlene Kombination aus Dateiformat und Codec kann mit FFmpeg erzeugt werden.*