

Guide pour un archivage numérique à long terme

Leitfaden für Digitale Langzeitarchivierung

Stand 21.03.2024 (V9 DE)

Mitglieder der AG Sylvie Béguelin (Médiathèque Valais)
Romain Guedj (BCU Fribourg)
Théophile Naito (BCU Lausanne)
Alexis Rivier (Bibliothèque de Genève)
Brigitte Sacker (ZB Zürich)
Tobias Viegner (NB)
Philipp Wiemann (KB Vadana St.Gallen)
Mirjam Zürcher (ZHB Luzern)

Inhaltsverzeichnis

1	Einführung.....	3
2	Allgemeine Fragen.....	4
2.1	Strategie.....	4
2.2	Infrastruktur.....	4
2.3	Umfang.....	4
2.4	Partnerschaften.....	5
2.5	Erhaltung und/oder Vermittlung.....	6
3	Einführung eines digitalen Langzeitarchivs.....	6
3.1	Das OAIS-Referenzmodell.....	6
3.2	Konzeptionelle und technische Details.....	8
4	Richtlinien für den Erwerb.....	10
4.1	Provenienz.....	10
4.2	Darstellung der Daten.....	11
4.3	Datenlieferung.....	12
4.4	Auslesen von Daten.....	13
4.5	Erwerb der Daten.....	15
5	Erhaltungsplanung (Preservation Planning).....	20
5.1	Definition.....	20
5.2	Erwerbungspolitik.....	21
5.3	Erwerbungsprozess digitaler Objekte.....	22
5.4	Ingest (Übernahme).....	23
5.5	Dauerhafte Speicherung.....	24
5.6	Speicher-Strategie.....	25
5.7	Technologie-Monitoring.....	27
5.8	Metadaten.....	27
6	Zugänglichkeit und Nutzbarkeit.....	30
7	Zusammenfassung und Schlussbetrachtung.....	32
8	Anhang.....	34
8.1	Schlüsselfragen.....	34
8.2	Glossar.....	38

1 Einführung

Die digitale Langzeitarchivierung beschäftigt Gedächtnisinstitutionen seit mehreren Jahrzehnten. Bereits 2002 wurde der Standard Open Archival Information System Reference Model (OAIS) veröffentlicht. Dieses Organisationsmodell bildet die massgebliche Grundlage für alle Systeme zur dauerhaften Aufbewahrung digitaler Daten.

Obwohl in der Bundesverwaltung (Nationalbibliothek, Bundesarchiv) oder in den Kantonen (z. B. Wallis und Basel) – bisweilen schon vor mehreren Jahren – Langzeitarchivsysteme in Betrieb genommen wurden und es sehr aktive Anbieter gibt, ist der Prozess zum Aufbau einer solchen Infrastruktur für eine Kantonsbibliothek komplex und es existiert keine betriebsfertige Standardlösung.

Der Kontrast zur Aufbewahrung physischer Dokumentationsquellen, wie sie Bibliotheken und Archive seit Jahrhunderten betreiben und effizient meistern, ist frappierend. Die physischen Quellen innewohnende Einfachheit bestimmt, welche Konservierungsmethoden angewandt werden müssen: Es reicht aus, Dokumente geschützt aufzubewahren, um sie vor dem Verfall zu bewahren und den Zugang zum Inhalt für die Zukunft sicherzustellen.

Die *Kulturbotschaft 2020–2024* des Bundesamtes für Kultur (BAK) beschreibt die Problemstellung der Digitalisierung: «Die Nutzung digitaler Informationen gehört heute zu unserem Alltag. Nicht selbstverständlich sind die langfristige Erhaltung und Nutzbarkeit dieser Daten.»¹ Sie anerkennt, dass «die Gewährleistung der langfristigen Erhaltung und Nutzbarkeit digitaler Daten» komplex ist und die bevorzugte Herangehensweise in der Zusammenarbeit liegt: «Aufgrund der hohen Kosten ist diese Aufgabe nur kooperativ zu bewältigen. Hier sind neben Kantonsbibliotheken und Staatsarchiven auch die NB und das Schweizerische Bundesarchiv als Gedächtnisinstitutionen des Bundes in der Pflicht.»

Die *Strategie der Schweizerischen Nationalbibliothek NB 2020–2028*² enthält die Verpflichtung zur langfristigen Erhaltung digitaler Ressourcen in Zusammenarbeit mit anderen Institutionen des Bundes und der Kantone, um das Gedächtnis der Schweiz zu bewahren. Die von der SKKB/CSBC beauftragte Arbeitsgruppe Digitale Repositorien (AG DigiRep) steht dazu in direktem Zusammenhang.

Der erste Bericht der von der SKKB/CSBC beauftragten Arbeitsgruppe mit dem Titel «*Leitfaden für ein Sammlungsprofil Digitale Information*» (2020)³ beschäftigt sich mit den Herausforderungen der digitalen Langzeitarchivierung im Allgemeinen – insbesondere mit Blick auf Kantonsbibliotheken.

Diese müssen sich bezüglich ihrer Positionierung im Umgang mit digitalen Inhalten eine grundlegende Frage stellen: Hat die Bibliothek den Auftrag, digitale Daten aufzubewahren und muss sie daher über ein digitales Langzeitarchivsystem (DLZA) verfügen?

Fällt die Antwort auf diese Frage positiv aus, sind weitere zentrale Fragen zu klären. Vorliegendes Dokument setzt sich zum Ziel, die notwendigen und zentralen Fragen zu stellen, die bei der kon-

¹ <https://www.fedlex.admin.ch/eli/fga/2020/725/de>

² https://www.nb.admin.ch/dam/snl/de/dokumente/nb_als_flag-einheit/studien_und_berichte/nb_strategie.pdf.download.pdf/nb_strategie.pdf

³³ https://www.bibliosuisse.ch/Portals/0/Inhalte/Sektionen/SKKB/Aktivitaet%20C3%A4ten/Leitfaden-Sammlungsprofil-Digitale-Medien_DE_final_20201006.pdf?ver=5aE0jPMm_jgTJ4rSaRDINQ%3d%3d

kreten Umsetzung einer Infrastruktur zur langfristigen Archivierung digitaler Inhalte berücksichtigt werden müssen.

2 Allgemeine Fragen

2.1 Strategie

Die Einrichtung eines Langzeitarchivierungssystems ist ein Innovationsprozess. In einer Situation mit bestenfalls stabilen finanziellen und personellen Ressourcen und in Anbetracht der Tatsache, dass Basisdienstleistungen nicht aufgegeben werden können (insbesondere diejenigen, die die Bearbeitung und den Zugang zu traditionellen physischen Dokumenten betreffen), erfordert dies eine **klare strategische Ausrichtung** der Bibliothek.

Diese Strategie muss die Definition und Rechtfertigung eines Projekts zur digitalen Langzeitarchivierung rechtfertigen können. Sie muss zudem die dazu nötigen Kompetenzen festlegen und klären, ob diese bereits vorhanden sind. Mittels Neueinstellungen können allfällige Lücken geschlossen und effiziente interdisziplinäre Teams zusammengestellt werden. Die Definition einer gut durchdachten Strategie erfordert den Erwerb eines Minimums an Wissen und Know-how im Bereich digitaler Langzeitarchivierung und kann folglich nicht vollständig ausgelagert werden. Man muss sich zudem bewusst sein, dass der Betrieb eines DLZA wiederkehrende Kosten umfasst, die in die Budgetplanung miteinzubeziehen sind. Da Entscheidungsprozesse in der öffentlichen Verwaltung langsam sind, ist frühzeitig an deren Planung zu denken.

Ein **Pilotprojekt** von begrenztem Umfang stellt eine schlanke Alternative dar. Die in diesem Pilotprojekt gewonnen Erkenntnisse können als Grundlage zur Ausarbeitung der umfassenden Strategie dienen. Für ein Pilotprojekt bieten sich unterschiedliche Gelegenheiten: z. B. die Zusammenarbeit mit einem Datenproduzenten, der die digitale Archivierung erproben möchte, oder die Aufnahme eines wichtigen digitalen Bestandes in die Sammlungen der Bibliothek.

2.2 Infrastruktur

Die Aktivitäten der Bibliothek werden durch eine technische Umgebung unterstützt. Bereits vorhandene IT-Infrastruktur muss berücksichtigt werden und kann als Baustein für eine DLZA-Lösung dienen. Dazu gehören:

- das/die bereits genutzte(n) professionelle(n) Verwaltungssystem(e): integriertes Bibliotheksverwaltungssystem, Archivverwaltungssystem, Museumsverwaltungssystem, Verwaltungssystem für digitale Bibliotheken
- die globale IT-Infrastruktur der jeweiligen übergeordneten Verwaltung (Universität, Kanton, Stadt): Dateiserver, DMS (elektronisches Dokumentenmanagementsystem), Digital Asset Management (DAM) oder sogar ein DLZA, das bereits von einer anderen Verwaltungseinheit (z. B. Archiv) genutzt wird.

2.3 Umfang

Der **Umfang** der betroffenen Daten ist ein entscheidender Faktor, da er Grösse und Wachstum des benötigten Systems bestimmt.

Dieser Umfang hängt vom Sammlungsauftrag der Institution und dessen gesetzlicher Grundlage ab. Die Situation ist in den einzelnen Kantonen sehr unterschiedlich. Einige Westschweizer Kan-

tone (FR, GE, VD) sehen die Abgabe eines Pflichtexemplars vor, die jedoch nicht immer auch explizit digitale Veröffentlichungen umfasst. Andere Kantone haben Sammelaufträge, die auf unterschiedliche Weise in kantonalen Gesetzen oder Verordnungen verankert sind.⁴ Daher muss jede Institution die geltenden Gesetze und Verordnungen sorgfältig prüfen und daraus den Umfang der vom Sammlungsauftrag betroffenen digitalen Ressourcen ableiten.⁵

Sobald der Umfang der Sammlung definiert ist, muss deren Menge

- in Abhängigkeit von den **betroffenen Ressourcentypen** (Text, gescannter Text, Standbild, Ton, Bewegtbild usw.),
- in Abhängigkeit von der Anzahl der zu verarbeitenden Objekte (oder Stunden bei Ton und bewegten Bildern) für jeden Ressourcentyp sowie
- in Abhängigkeit von der erwarteten Zunahme in den nächsten Jahren

quantifiziert werden.

Diese Grössenbestimmung sollte in **Speicherplatz** (TB) ausgedrückt werden. Diese Einheit ermöglicht die Infrastrukturanforderungen des DLZA und deren Entwicklung über die Jahre zu beziffern. Dies gilt auch für ausgelagerte DLZA-Leistungen wie sie das Bundesarchiv anbietet, das die Kosten der zu erhaltenden Daten in Franken pro TB festlegt.⁶

2.4 Partnerschaften

Zur Erreichung der im Rahmen der Strategie festgelegten Innovationsprojekte setzt die NB bevorzugt auf Partnerschaften: «Das Gedächtnis der Schweiz fusst auf enger Kooperation von Bibliotheken, Archiven, Museen und Forschungseinrichtungen. Die Nationalbibliothek setzt sich für eine effektive nationale und internationale Koordination der relevanten Akteure ein.»⁷

In der Vergangenheit führte ein ähnlicher Ansatz zum Aufbau des Verbundes Memoriav. Anstatt eine neue Institution zur Rettung des audiovisuellen Erbes zu schaffen, hat Memoriav die Vernetzung der vielen über das Land verstreuten Fachpersonen und damit deren Kompetenzen bevorzugt.

Für Kantonsbibliotheken ergeben sich Möglichkeiten für eine Partnerschaft auf zwei Ebenen:

- Eine lokale oder vertikale Ebene: Die Trägerschaft (Kanton, Universität, Stadt) verfügt über Schlüsselkompetenzen (IT-Teams) oder über Institutionen mit ähnlichen Anforderungen an die digitale Archivierung (Archive, Museen).
- Eine nationale oder eher horizontale Ebene: Die Kantonsbibliotheken und die NB haben die gleichen Bedürfnisse und bündeln ihre Kompetenzen.

Der Ansatz auf nationaler Ebene ermöglichte den Kantonsbibliotheken die Teilnahme an sehr wichtigen und strukturbildenden Vorhaben im Bereich «Information und Dokumentation»: SLSP, e-rara, e-manuscripta, e-periodica, e-newspaperarchives.

⁴ *Leitfaden für ein Sammlungsprofil Digitale Information* (CSBC/ SSKB 2020), Kap. 5.

⁵ Der *Leitfaden für ein Sammlungsprofil Digitale Information* (CSBC/ SSKB 2020) geht in Kap. 2 «Auswahl» näher auf diese Frage ein.

⁶ *Digitale Archivierung im BAR Dienstleistungen für bundesverwaltungsexterne Kunden*, 01.09.2021. Siehe <https://www.bar.admin.ch/bar/de/home/archivierung/digitale-unterlagen/archivierung-fuer-dritte.html>

⁷ *Strategische Ausrichtung der Schweizerischen Nationalbibliothek NB ab 2024*, 01.03.2024. Siehe https://www.nb.admin.ch/dam/snl/de/dokumente/nb_als_flag-einheit/studien_und_berichte/nb_strategie.pdf.download.pdf/nb_strategie.pdf

Die Einführung digitaler Langzeitarchive erfolgte eher auf der Basis lokaler Partnerschaften. Beispielsweise profitieren das Staatsarchiv Wallis, die Mediathek Wallis und die kantonalen Museen vom Projekt eines DLZA der Walliser Dienststelle für Kultur. Die Bundesverwaltung hat je ein DLZA für die NB und das BAR eingerichtet. Der Beizug von schweizweit auf die digitale Langzeitarchivierung spezialisierten Unternehmen empfiehlt sich, da sie ein gutes Verständnis für die fachlichen Bedürfnisse und die Anforderungen der IT-Dienste mitbringen.

2.5 Erhaltung und/oder Vermittlung

Das OAIS-Modell konzeptualisiert neben der Übernahme digitaler Inhalte (Ingest) und deren Verwaltung im Archiv auch deren Vermittlung für Nutzerinnen und Nutzer (Access).

Das Vermittlungssystem ist von offensichtlichem öffentlichem und politischem Interesse, verleiht dem Projekt jedoch auch eine zusätzliche Komplexität. Der Zugang für die Öffentlichkeit unterliegt möglicherweise Beschränkungen aufgrund von Urheberrecht, Datenschutz oder Sperrfristen, die vom System verwaltet werden müssen.

Zudem muss das Vermittlungssystem mit den bestehenden Verwaltungssystemen verknüpft werden. In einigen Fällen bietet das Verwaltungssystem bereits selbst eine öffentliche Benutzeroberfläche zur Vermittlung. Die NB integriert beispielsweise im DLZA e-Helvetica archivierte digitale Dokumente in ihr Portal HelveticAll (Alma) und unterscheidet zwischen freien Inhalten und eingeschränkten Inhalten, die mittels spezieller Arbeitsplätze konsultiert werden müssen. Andererseits kann es einfacher sein, ein vom bestehenden Verwaltungssystem unabhängiges Vermittlungssystem einzurichten. So nutzt die KUB/F Alma zur Verwaltung und AtoM für die Vermittlung.

3 Einführung eines digitalen Langzeitarchivs

3.1 Das OAIS-Referenzmodell

Auf Grundlage einer Klärung der organisatorischen, finanziellen und technischen Rahmenbedingungen und der Erhebung des Bedarfs kann die konkrete Systementwicklung in Angriff genommen werden. Konzeptionelle Basis der meisten heutigen Systeme zur digitalen Langzeitarchivierung ist das OAIS-Referenzmodell⁸. Das *Open Archival Information System* bzw. Offenes Archiv-Informationen-System-Referenzmodell beschreibt Funktionseinheiten, die im Zusammenspiel von Menschen und Systemen dem Zweck dienen, Informationen langfristig – d. h. bis in eine unbestimmte Zukunft – zu erhalten.⁹ Es werden folgende Funktionseinheiten unterschieden:

- Übernahme (Ingest): Nimmt Daten von Produzenten (auch Datenlieferanten genannt) an.
- Administration (Administration): Überprüft den Betrieb der anderen Funktionseinheiten.
- Archivspeicher (Archival Storage): Stellt die Speicherung und das Wiederauffinden von archivierten Daten sicher.
- Datenverwaltung (Data Management): Verwaltet neben Katalogen und Inventaren auch Zugriffe, Sicherheitskontrollen, datenverarbeitende Algorithmen, Statistiken u. v. a. m.
- Erhaltungsplanung (Preservation Planning): Die Erhaltungsplanung stellt sicher, dass die Information im OAIS langfristig zugänglich, verständlich und benutzbar bleibt.
- Zugriff (Access): Macht die Dienste und Informationen im OAIS für Endnutzer sichtbar.

⁸ <http://d-nb.info/104761314X/34>

⁹ http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_183.pdf

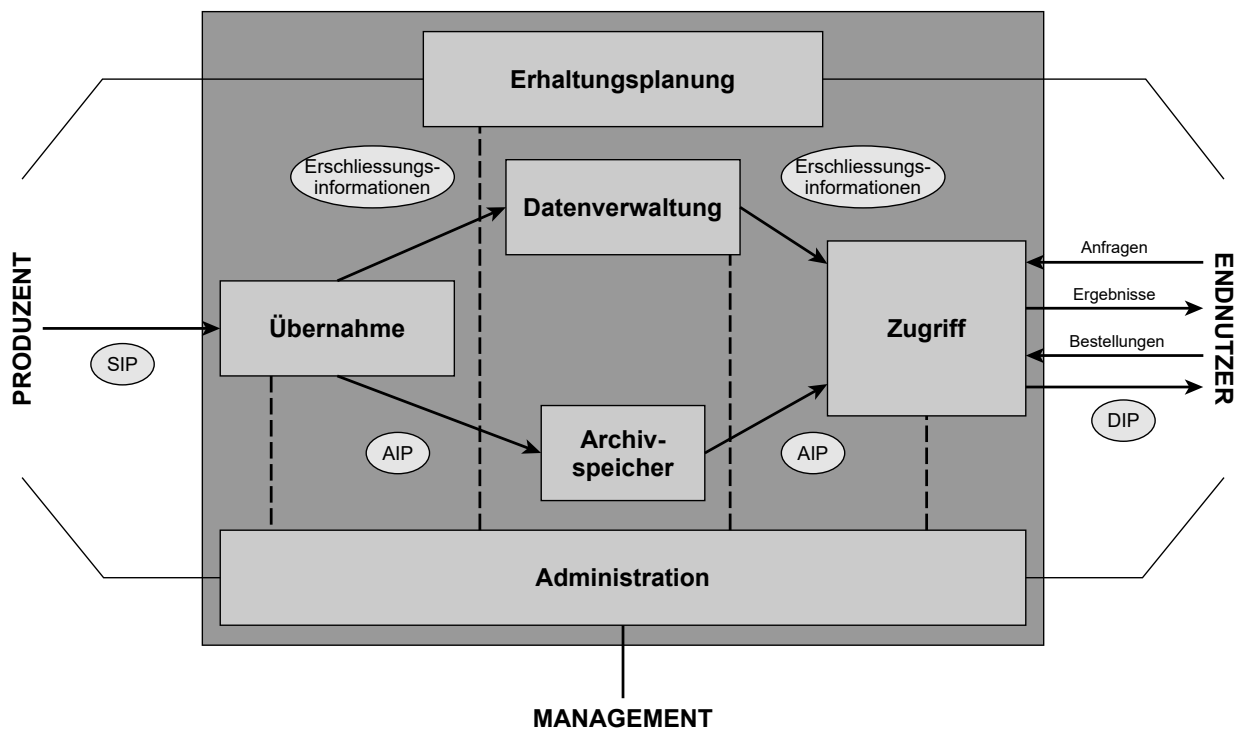


Abbildung 1: OAIS-Funktionseinheiten (nach «Referenzmodell für ein Offenes Archiv-Informationssystem (OAIS)»)

Ein weiteres wesentliches Konzept innerhalb eines OAIS ist dasjenige des Informationspakets, welches aus Inhaltsinformationen, Erhaltungsmetadaten sowie Erschliessungsinformationen besteht. Im Fall eines OAIS wird zwischen Übergabeinformationspaketen (SIP, **Submission Information Packages**), die von Produzenten (auch «Datenlieferanten» genannt) an das System übergeben werden, den im OAIS gespeicherten Archivinformationspaketen (AIP, **Archival Information Packages**) und den an Endnutzer vom System ausgelieferten Auslieferungsinformationspaketen (DIP, **Dissemination Information Packages**) unterschieden. Die Beziehungen zwischen den Paketen können vielfältig sein: So ist es möglich, dass SIP und AIP strukturell gleich aufgebaut sind und sich auch hinsichtlich der Formate der Inhalte marginal unterscheiden – andererseits können sie aber auch strukturell verschiedenartig sein. Allerdings wird das AIP im Vergleich zum SIP zusätzliche Informationen enthalten, die für die Erhaltung der digitalen Objekte benötigt werden. Dasselbe kann für AIP und DIP gelten, wobei hier die Wahrscheinlichkeit für Unterschiede höher ist, da sich die durch die Archivierung gegebenen Anforderungen stärker von denjenigen unterscheiden, die sich für die Anzeige und Auslieferung ergeben.

Ob sich die Pakete strukturell, inhaltlich und hinsichtlich der Formate ähnlich sind oder sich unterscheiden, ist einerseits vom eingesetzten System abhängig und wird andererseits von der Art und Weise bestimmt, wie digitale Objekte durch eine Institution verarbeitet werden. Zum Beispiel könnten TIFF-Bilder in einem Digitalisierungsprojekt laufend vom Scanner als einzelne SIPs ans Archiv übergeben werden. Dort werden sie jedoch in einem einzelnen AIP gespeichert, das die vielen hundert Seiten eines digitalisierten Buches im JPEG2000-Format beinhaltet. Beim Zugriff eines Nutzers werden schliesslich nur die angeforderten Seiten als DIP in Form einer mehrseitigen PDF-Datei ausgeliefert.

3.2 Konzeptionelle und technische Details

3.2.1 Entwicklung

Das OAIS-Referenzmodell ist heute anerkannte ISO-Norm¹⁰ und auch deshalb erfolgreich, weil es dank der Definition von Rollen, Verantwortlichkeiten und Prozessen eine wichtige Grundlage für die Formulierung von Anforderungen an ein System zur Langzeitarchivierung darstellt, aber das System in technischer Hinsicht nicht konkret vorgibt. Dies lässt Raum für die Entwicklung unterschiedlich komplexer technischer Lösungen, die den Bedürfnissen der jeweiligen Organisationseinheiten Rechnung tragen können. Beispielsweise muss nicht unbedingt jeder Prozess in technischer Hinsicht komplett automatisiert ablaufen. Vielmehr kann in manchen Fällen etwa das manuelle Erstellen von SIPs der angemessenere Weg sein, während unter anderen Voraussetzungen eine technische Schnittstelle das Erstellen von SIPs und die Übernahme durch das Langzeitarchiv besser vollautomatisch übernehmen sollte. Diese Flexibilität stellt jedoch gleichzeitig auch eine Schwierigkeit für die Beschaffung eines Systems dar: Angesichts dessen, dass Softwarelösungen zur digitalen Langzeitarchivierung oftmals auf den konkreten Anwendungsfall hin parametrisiert werden, sind häufig aufwändige Vorabklärungen zur Festlegung der genauen Softwarearchitektur nötig. Die weiter oben ermittelten organisatorischen und strukturellen Rahmenbedingungen bilden die Grundlage für das Erstellen eines Pflichtenheftes und die Suche nach einem geeigneten Anbieter. Dank des OAIS-Referenzmodells und der damit gelieferten Terminologie ist es jedoch möglich, dass Anbieter und Auftraggeber sich über die Funktionalitäten verständigen können, ohne detailliert die technische Umsetzung diskutieren zu müssen.

Eine weitere Grundlage für die Systementwicklung sind Angaben zu den erwarteten Quantitäten der Ablieferungen, der abgelieferten Objekte, deren Datenformate und Quellsysteme sowie deren Schnittstellen. Neben der Berücksichtigung der Produzenten sind auch die Datennutzenden (Personen, Institutionen, Systeme) zu beachten. Ein DLZA kann beispielsweise nur für wenige befähigte Personen zugänglich sein oder Nutzende können direkt im Internet über ein internes oder externes Präsentationssystem auf Inhalte zugreifen. Im ersten Fall spricht man auch von einem sogenannten *Dark-Archive*. Im Fall einer direkten Zugriffsmöglichkeit stellt sich die Frage, ob Präsentationssysteme oder Bibliothekskataloge DIPs erst bei einer Bestellung durch einen Nutzenden über ein Bestellsystem generieren und ausliefern oder ob DIPs aller relevanten Informationspakete in den am häufigsten gebräuchlichen Formaten ständig direkt im Präsentationssystem vorgehalten werden. Diese Entscheidung muss in Hinblick auf den jeweiligen Anwendungsfall und in einer Abwägung von Kosten und Nutzen getroffen werden.

3.2.2 Umsetzung

Systeme zur digitalen Langzeitarchivierung basieren oftmals auf *Digital Asset Management* (DAM) Systemen wie *Fedora Commons* und ergänzen diese in einem modularen Aufbau mit weiteren Softwarepaketen, um sämtliche funktionalen Anforderungen abdecken zu können (vgl. Abbildung 2, S. 9). Auf dem Markt existieren sowohl *Open-Source* als auch proprietäre Lösungen von kommerziellen Anbietern. Stehen der Organisation umfangreiche personelle und technische IT-Ressourcen zur Verfügung, kann die Implementation einer Open-Source-Lösung auf eigener Serverinfrastruktur (Anwendungs-, Datenbank- und Fileserver) **On-Premises** aufgrund der guten Anpassung an den Bedarf bei niedrigen bzw. nicht vorhandenen Lizenzkosten eine geeignete Lösung sein. Auf der anderen Seite des Spektrums stehen kleine Organisationen, die entweder eine **cloudbasierte Lösung** oder sogar einen **Preservation-as-a-Service-Dienst** ins Auge fas-

¹⁰ <https://www.iso.org/standard/57284.html>

sen dürften. Bei einer cloudbasierten Lösung liegt das Betreiben der Serverinfrastruktur im Verantwortungsbereich des Dienstleisters. Die Einflussmöglichkeiten des Kunden auf das System sind oftmals begrenzt, da die Hersteller Änderungen zustimmen bzw. Entwicklungen vornehmen müssen, sofern es sich nicht nur um einfache Konfigurationen handelt. Ausserdem ist gesondert zu prüfen, ob bei einer cloudbasierten Lösung einfach nur das System in der Cloud läuft oder auch sämtliche (redundanten) Archivspeicher. Es gibt Angebote, bei denen das System angeboten wird, der redundante Speicher jedoch teilweise in der Verantwortung der Kunden liegt. Somit zeigt sich ein möglicher zusätzlicher Kostenpunkt. Bei *Preservation-as-a-Service* hingegen liegt die Verantwortung für die archivierten Daten wie auch die Kompetenz, etwa im Bereich der Erhaltungsmassnahmen, vollständig beim Dienstleister. Bei beiden Arten von Diensten werden die Kosten zumeist anhand des Speicherbedarfs des archivierten Materials abgerechnet. Dazu-rechnen sind noch die Kosten für Support und Lizenzen.

Architektur

docuteam

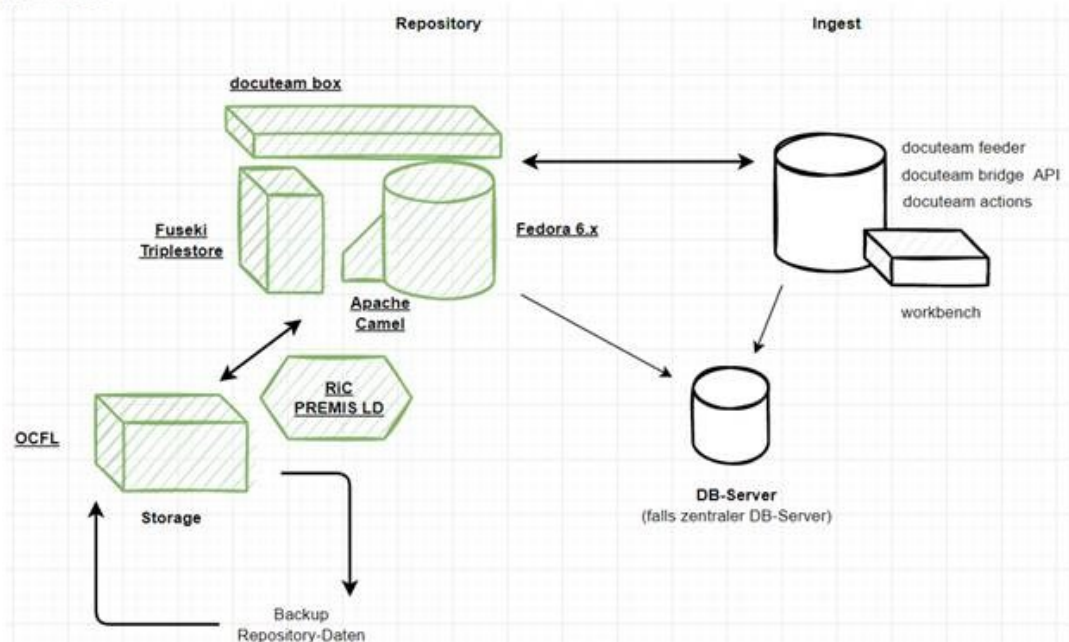


Abbildung 2: Darstellung der modularen Architektur eines Systems zur digitalen Langzeitarchivierung am Beispiel des Systems der Firma docuteam AG (Bildnachweis: docuteam AG)

Im Rahmen der Umsetzung einer Lösung sollte stets auch das Scheitern eingeplant werden. Falls beispielsweise die hauseigene Implementierung einer Open-Source-Lösung aufgrund personeller Veränderungen nicht sinnvoll zu Ende geführt werden kann, könnte der Kauf einer proprietären Lösung in Betracht gezogen werden. Bevor etwaige Gebührenerhöhungen bei einem *Preservation-as-a-Service*-Modell nicht mehr getragen werden können, ist bereits im Vorfeld auf das Vorhandensein einer dokumentierten Exit-Strategie zu achten, die einen Export der Daten und Metadaten in allgemein anerkannten Standardformaten erlaubt (Vorbeugung eines *Cloud-* bzw. *Vendor Lock-in*).

3.2.3 Vorbereitung

Aus den oben genannten Ausführungen ergeben sich einige Fragen, die im Vorfeld der Entwicklung und Implementierung von Systemen zur digitalen Langzeitarchivierung geklärt werden sollten:

- **On-Premises oder in der Cloud**

Habe ich die Ressourcen, ein DLZA auf eigenen Servern und mit eigenem Personal zu betreiben? Bei Präferenz eines cloudbasierten DLZA: Wieviel Einfluss möchte ich auf die Lösung behalten (z. B. in Bezug zu Erhaltungsmaßnahmen)? Sind der Speicherort der Daten und die gesetzlichen Rahmenbedingungen mit dem Schutzbedürfnis meiner Daten vereinbar?

- **Kostenfolgen**

Wie gross ist der Umfang der Daten, die ich archivieren möchte? Wie viele Datenübernahmen plane ich pro Jahr und wie unterschiedlich sind die Datenquellen? Wie einfach ist es, den Übernahmeprozess an die jeweilige neue Datenquelle anzupassen? Wie hoch ist der Zuwachs pro Jahr?

- **Systemintegration**

Zu welchen anderen Systemen sollen Schnittstellen zur Datenübernahme oder zum Datenzugriff bestehen? Welche Metadaten(formate) sollen dabei ausgetauscht werden? Wie performant sollen diese Prozesse ablaufen?

4 Richtlinien für den Erwerb

4.1 Provenienz

Zur Erweiterung ihrer Sammlungen erwerben Bibliotheken digitale Dokumente von unterschiedlichen Institutionen oder Privatpersonen:

- Eigenproduktion
- interne Digitalisierungsabteilung
- externe Digitalisierungsunternehmen
- Verlage/Buchhandlungen/Vertriebe/etc.
- Privatpersonen (Schriftsteller/Komponisten/Fotografen/etc.)
- Institutionen

Die selbständige Digitalisierung gehört in der Regel nicht zu den Hauptaufgaben einer Bibliothek und erfolgt folglich meist in geringem Umfang oder ist gar nicht vorhanden. Allerdings gibt sie der Bibliothek die umfassende Kontrolle über die mittels eigener Digitalisierung zur Sammlung hinzugefügten digitalen Dokumente.

Von Personen und Institutionen erworbene Daten unterliegen nicht den Produktionsregeln von Bibliotheken und können folglich in verschiedenen Formen vorliegen, sowohl hinsichtlich der Datei- und allenfalls vorhandener Metadatenformate als auch hinsichtlich der Datenträger und der Ablieferungsformen. In diesem Umfeld sind Flexibilität und Anpassungsfähigkeit wichtige Eigenschaften, ebenso wie die Verfügbarkeit von Personal, Geld und erforderlicher technischer Infrastruktur. Zudem helfen Sammlungsrichtlinien, mit denen festgelegt werden kann, dass aufgrund ihrer Menge und oder ihrer technischen Merkmale schwer handhabbare Dokumente nicht aufgenommen werden.

4.2 Darstellung der Daten

4.2.1 Datenebenen

Auf einem Datenträger lassen sich digitale Daten auf mehreren Ebenen betrachten, was folgende Abbildung zeigt:

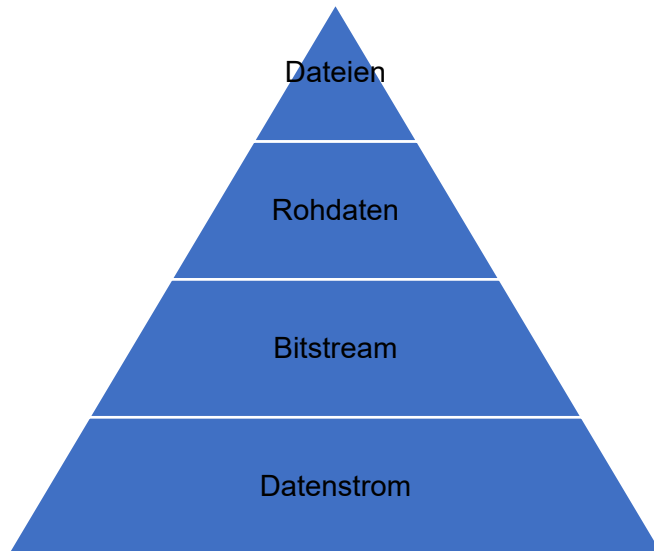


Abbildung 3: Ebenen digitaler Daten auf einem Datenträger

Information wird zunächst als eine Folge von Binärwerten in datenträgerspezifischer Form kodiert (z. B. Vertiefungen und Erhöhungen auf einer CD). Mittels eines geeigneten Lesegeräts wird dieser **Datenstrom** in einen **Bitstream** umgewandelt, der aus **Roh-** und **Prüfdaten** (z. B. Prüfsummen) besteht, während sich die Rohdaten aus den Dateien und technischen Metadaten (z. B. Änderungsdatum) zusammensetzen.

Die Informationsmenge steigt, je tiefer die Ebene ist.

Damit auf Originaldatenträgern abgelieferte Daten und Metadaten angemessen verarbeitet und archiviert werden können, sind die Ebenen digitaler Daten auf dem Datenträger zu berücksichtigen.

4.2.2 Dateiformat

Für von der Bibliothek selbst oder in ihrem Auftrag erstellte Dateien sind die erwarteten Eigenschaften festzulegen. Diese müssen unter Berücksichtigung der Anforderungen der digitalen Archivierung definiert und formalisiert werden.

Auch wenn es keine allgemeingültigen Vorgaben in diesem Bereich gibt, so verfügen doch bereits viele Institutionen über Richtlinien, die als Ausgangspunkt für die Erstellung eigener Vorgaben dienen können.

Oft lässt sich die Ablieferung von Daten in bestimmten Formaten und mit vorgegebenen Eigenschaften nicht durchsetzen. In diesem Fall können folgende Massnahmen ergriffen werden:

- **Informieren**

Informieren des Datenlieferanten über die Vorgaben und Anforderungen der eigenen Institution.

- **Sensibilisieren**

Die Durchführung von Sensibilisierungsmassnahmen bei den Datenlieferanten erlaubt es, «Best Practices» bekannt zu machen, die eine bessere Archivierung der von diesen erstellten und angelieferten Daten gewährleisten. Ein Beispiel hierfür sind die von Memoriav entwickelten Empfehlungen,¹¹ die sich teilweise an ein Publikum ohne besondere Fachkenntnisse richten.

- **Erstellen von Derivaten**

Wenn angelieferte Daten nicht den bestehenden Sammlungsrichtlinien oder nicht den technischen Möglichkeiten eines Archivierungssystems entsprechen, ist die Erstellung von besser für die Archivierung geeigneten Derivaten zu prüfen, damit die Archivierung sichergestellt werden kann.

- **Ablehnen von angelieferten Daten**

Um zu vermeiden, dass Daten angenommen werden müssen, deren Verarbeitung einen unverhältnismässig hohen Arbeitsaufwand erfordert, hilft die Einführung einer Sammlungsrichtlinie, die auch Grenzen dahingehend setzt, welche Arten von Dokumenten und Dateien angenommen werden.

Wenn das DLZA oder die Vorgaben der Bibliothek ein spezifisches Format für SIPs festlegen, so sollen diese Informationen an den Datenlieferanten weitergegeben werden, damit dieser im Idealfall bereits SIPs im korrekten Format übermitteln kann.

4.2.3 Metadatenformat

Den gelieferten digitalen Dokumenten sollten – wenn immer möglich – die entsprechenden Metadaten beiliegen, damit problemlos bibliographische Einträge für diese Dokumente erstellt werden können. Besonders bei Archivmaterial kann dies auch dazu beitragen, Informationen zu erhalten, die auf andere Weise nicht zu beschaffen wären. Beispielsweise kann ein Fotograf den von ihm erstellten Fotografien mittels der von ihm eingesetzten Softwarelösung einen Titel und eine Beschreibung zuweisen. Diese Informationen können verloren gehen, wenn sie nicht automatisch in die Fotografien integriert oder in einer separaten Datei übermittelt werden.

Wie bei den Dateien ist es oft nicht möglich, die Metadaten in einem Format zu erhalten, das einen direkten Import in den Bibliothekskatalog oder das Archivierungssystem ermöglicht. Es ist dann notwendig, das Datenmodell der initial angelieferten Metadaten auf das Datenmodell der Bibliothek abzubilden (Mapping).

Es ist zu beachten, dass neben den separat abgelieferten Metadaten auch viele Metadaten direkt in der Datei selbst abgespeichert sind: Datum der Erstellung, Datum der letzten Änderung, Dateieigentümer usw. Damit für die Archivierung möglichst umfassende Informationen gesammelt, gespeichert und verwendet werden können, sind beide Arten von Metadaten zu nutzen.

4.3 Datenlieferung

4.3.1 Datenlieferung online

Die Datenlieferung online kann auf unterschiedliche Art und Weise erfolgen.

Unabhängig von den einzuführenden Systemen zur Datenübertragung ist frühzeitig die Infrastruktur für die Speicherung der für die Archivierung aufzubereitenden Dateien zu planen. Verfügt die

¹¹ <https://memoriav.ch/de/empfehlungen/>

Institution über ein eigene IT-Abteilung, wird dringend empfohlen, dass diese den Umfang des benötigten temporären Speicherplatzes schätzt. Für die Berechnung sollten die jährlich zu übernehmende Datenmenge und die für den Archivierungsprozess benötigte Zeit berücksichtigt werden, damit ständig ausreichend freier Speicherplatz zur Verfügung steht.

4.3.2 Speicher und Datenaustausch

Zur Übertragung der Daten kann ein vom Datenlieferanten und der Bibliothek gemeinsam genutzter Speicher eingerichtet werden.

Dieser Speicher kann entweder vom Datenlieferanten oder der Bibliothek bereitgestellt werden, z. B. über einen SFTP-Server, oder auch von einem Drittanbieter bezogen werden. Dazu gehören u. a. die folgenden mehr oder weniger bekannten Dienste: Dropbox, Google Drive, kDrive, OneDrive, Proton Drive, Switch Drive. Der gewählte Dienst muss die Anforderungen der Bibliothek und des Datenlieferanten, insbesondere in Hinblick auf den Datenschutz, erfüllen.

4.3.3 Datentransfer

Anstelle eines gemeinsam genutzten Speichers kann auch ein System zur automatischen Datenübertragung genutzt werden wie bspw. SwissTransfer, Switch Filesender, WeTransfer.

Wie beim gemeinsam genutzten Speicher ist zu prüfen, ob der gewählte Dienst die Anforderungen der Bibliothek und des Datenlieferanten erfüllt.

4.3.4 Datenträger

Ein Datentransfer kann auch mittels eines Datenträgers erfolgen, der vom Datenlieferanten an die Bibliothek geliefert wird. Dabei kann es sich z.B. um einen USB-Stick oder eine externe Festplatte handeln, aber auch um einen technologisch veralteten Datenträger.

4.3.5 Automatisierung

Liefert eine Organisation (z. B. ein Verlag) regelmässig Dokumente an die Bibliothek ab, ist zu prüfen, ob sich die Datenübertragung automatisieren lässt.

Bestenfalls kann die Bibliothek auf diese Weise Daten ohne manuelle Eingriffe in ihr DLZA aufnehmen. Dies lässt sich beispielsweise mithilfe einer API lösen, sofern der Datenlieferant seine Daten mittels API bereitstellt und das DLZA diese mittels API entgegennehmen kann.

Verfügt die Bibliothek über eine eigene IT-Abteilung, so kann die Automatisierung eines solchen Prozesses intern erfolgen. Ist dies nicht der Fall, können Angebote von einem oder mehreren IT-Dienstleistern eingeholt werden, um die Machbarkeit der Automatisierung zu prüfen.

4.4 Auslesen von Daten

Die Lieferung von Daten und Metadaten sowie die Kontrolle, ob sie gelesen werden können, ist ein wichtiger Schritt, für den die notwendigen Mittel zur Qualitätssicherung bereitgestellt werden müssen.

Vor jeder Datenlieferung soll der Datenlieferant eine vorgängige Bestandesaufnahme der zu liefernden Daten erstellen. Dies dient der Vorbereitung der Datenübernahme und erlaubt bspw.

- die Prüfung, ob die Daten vollständig abgeliefert wurden,
- die Sicherstellung (bereits vor der Lieferung), dass die Daten verarbeitet werden können, und, wenn dies nicht der Fall ist, das Aushandeln einer Ablieferung alternativer Dateiformate,
- das Anbieten alternativer Ablieferungsmöglichkeiten (z. B. die Abholung der Objekte beim Datenlieferanten, wenn die Datenmenge für eine Onlineübermittlung zu gross ist).

Dies stellt sicher, dass die Daten ungehindert empfangen werden können.

Beim Lesen und Kopieren von Daten von einem Datenträger können zahlreiche Schwierigkeiten auftreten.

Eine externe Festplatte ist aufgrund des Betriebssystems (Mac, Windows oder Linux) und des verwendeten Dateisystems (HFS+, NTFS etc.) möglicherweise nicht mit allen Computern kompatibel. Es kann auch eine Hardware-Inkompatibilität des Anschlusses (USB-A, USB-C, Thunderbolt etc.) vorliegen.

Ist der Datenträger veraltet, kann das Kopieren der Daten einen zusätzlichen Aufwand erfordern. R. François und R. Rochat schlagen einen mehrstufigen Prozess vor und nennen Software- und Hardware-Tools zu dessen Umsetzung.¹² Es kann auch ein spezialisiertes Unternehmen mit dem Auslesen eines veralteten Datenträgers beauftragt werden.

Der vorgeschlagene Prozess umfasst folgende Schritte:

1. Ermitteln und inventarisieren der Datenträger

Speichermedien sind auf Konformität mit der vorläufigen Bestandsaufnahme zu prüfen.

2. Ermittlung des Datenträgertyps und Vorbereitung der Wiedergabe

Beispielsweise kann das für einen bestimmten Datenträger erforderliche Lesegerät fehlen und muss erst erworben werden. Ausserdem kann das Verfahren zum Auslesen der Daten vorgängige Tests nötig machen, sofern es noch nie eingesetzt wurde.

3. Erstellen eines Datenträgerimages

Hier geht es darum, ein Image des Datenstroms (siehe Abbildung 3 auf S. 11) auf Basis der Anforderungen und entsprechend der technischen Möglichkeiten zu erstellen. Der Artikel von R. François und R. Rochat enthält viele Details über Hardware und Software, die für diesen Schritt nützlich sind.

4. Extrahieren der Daten und Metadaten aus dem Datenträgerimage

Zuerst sind die relevanten Dateien und Metadaten zu erkennen und zu extrahieren, bevor mit ihnen gearbeitet werden kann. Dazu kann beispielsweise die Software Aaru¹³ verwendet werden, vorausgesetzt sie unterstützt das Dateisystem des betreffenden Datenträgers.

5. Konvertieren, sortieren und dokumentieren

6. Archivieren

¹² Robin François et Rebecca Rochat, «Digital Preservation Pipeline for Data Storage Media at the Cinémathèque Suisse. Imaging and extracting data and metadata from Special Collections media», in: *18th International Conference on Digital Preservation*, 2022.

¹³ <https://www.aaru.app>

4.5 Erwerb der Daten

4.5.1 Grundsätze und Vorkehrungen

Ähnlich wie die Erwerbung von analogen Objekten durchläuft auch die Erwerbung von digitalen Objekten mehrere Schritte. Es wird auf folgende zwei Fälle eingegangen:

- digitale Objekte aus einem Archivbestand
- digitale Objekte aus der Digitalisierung oder einem DMS

Digitale Objekte aus einem Archivbestand

Ein Archivbestand ist eine Sammlung von Dokumenten, die von einem oder mehreren Urhebern erstellt wurden. Die Art dieser Sammlung hängt vom Herstellungskontext ab.

Beispielsweise wird ein Architekturbüro im Laufe seines Bestehens Modelle, Pläne, Fotografien, Filme, Rechnungen und Korrespondenz erstellen. Diese Dokumente werden zu unterschiedlichen Zeiten und für verschiedene Empfänger angefertigt worden sein, weshalb die Vielfalt der digitalen Dokumente gross ist. Wenn das Büro vor den 2000er Jahren gegründet wurde, ist es sehr wahrscheinlich, dass Pläne oder 3D-Modelle mit einer Software erstellt wurden, die es heute nicht mehr gibt.

Darüber hinaus variieren die Arbeitsorganisation, der interne Dokumentenfluss und die Speicherung von Dokumenten von einem Büro zum anderen. In den 1990er Jahren war es eher ungewöhnlich, dass Architekturbüros über ein DAM oder DMS verfügten. Vielmehr organisierte jeder Mitarbeitende die Speicherung seiner Dateien unterschiedlich, selbst dann, wenn mehrere Personen an einem Projekt arbeiteten. So kann ein und dasselbe Dokument an verschiedenen Orten abgelegt, unterschiedlich benannt sein oder in verschiedenen Dateiformaten vorliegen. Beispielsweise kann ein Plan als Maya-Datei vorliegen und gleichzeitig – wurde er an einen externen Mitarbeitenden versandt – als JPEG-, PNG- oder PDF-Datei.

Folglich ist die Vielfalt der Dateien sehr gross, die Anzahl der Duplikate nicht unerheblich, die Benennung der Dateien heterogen und es sind nur sehr wenige Metadaten vorhanden.

Digitale Objekte aus der Digitalisierung oder einem DMS

Im Rahmen der Digitalisierung gibt der Auftraggeber Dateiformat, Bildauflösung, Komprimierungsrate, Metadaten und weitere objektspezifisch eingebettete Metadaten (z. B. Farbräume) vor. Diese Vorgaben normieren die Dateitypen, die erworben und im Archiv erhalten werden.

So kann im Rahmen der Übernahme die Konformität der Objekte geprüft werden. Entsprechen Dateien nicht den für die Digitalisierung festgelegten Spezifikationen, können sie zurückgewiesen werden.

Ebenso können Standards und Prozesse innerhalb eines DMS sicherstellen, dass

- jede Datei einem Format entspricht, das in einer Auswahlliste von Formaten enthalten ist;
- doppelte Dateien vermieden werden;
- die Benennung der Dateien kontrolliert ist;
- die notwendigen Metadaten (Produktionsdatum, Versionierung, Hersteller usw.) vorhanden sind.

In beiden Fällen umfasst der Erwerbsprozess folgende Schritte:

- Erstellung einer vorgängigen Bestandesaufnahme
- Ausarbeitung einer Übernahmevereinbarung zur Klärung der Urheber-, Nutzungs- und Verbreitungsrechte der digitalen Objekte
- Anlieferung
- stichprobenartige Überprüfung des Zustands der digitalen Medien und Dateien
- Auswahl der zu erhaltenden und zu auszuschheidenden Objekte
- Übernahme (Ingest)
- Archivierung

4.5.2 Vorgängige Bestandesaufnahme

Vor dem Erwerb eines Bestandes sollte mit dem Datenlieferanten oder -produzenten eine vorgängige Bestandesaufnahme erstellt werden. Dies erlaubt eine erste Prüfung der Daten nach unterschiedlichen Gesichtspunkten:

- *Unterschiedlichkeit und Anzahl der Datenträger.* Es ist sicherzustellen, dass der Zugriff auf die verschiedenen Datenträger (ZIP-Diskette, Syquest, CD-R usw.) möglich ist.
- *Noch funktionstüchtige oder möglicherweise reparierbare Hardware,* um bestimmte Datenträger zu lesen.
- *Vielfältigkeit der Dateiformate.* Diese Prüfung erlaubt es, den möglichen Bedarf an Software zum Lesen der Dateien zu eruieren.
- *Spezielle, möglicherweise noch im Bestand vorhandene Programme.* Es ist wertvoll, zeitgenössische, bereits vom Produzenten verwendete Software zu erhalten. Dabei ist es wichtig, alle Aufzeichnungen zur Dokumentation der Software (Rechnungen, Korrespondenz) sowie die Schlüssel und Passwörter für die Softwarelizenz aufzubewahren, da diese Software manchmal schwer zu finden ist und nur ihre Weiterverwendung einen originalgetreuen Zugriff auf Dateien aus dem Bestand erlaubt. Allerdings kann dies eine Emulation des Betriebssystems erfordern.
- *Standorte von Datenträgern.* Es ist nicht ungewöhnlich, dass Archivbestände auf mehrere Produktions- oder Lagerorte verteilt sind.
- *Bei einem DMS die Häufigkeit der Uploads pro Datentyp,* sofern es sich um einen kontinuierlichen Datenfluss handelt, da diese Informationen die Planung der für die Verarbeitung dieses Datenstroms benötigten Mittel erlauben.
- *Auswahl der zu erhaltenden und auszuschheidenden Objekte.*
- *Strategie für die Übermittlung der Dokumente an die Institution.* Sie hilft bei der Beantwortung folgender Fragen:
 - Wie werden die Daten kopiert? Erfolgt dies in den Räumlichkeiten des Produzenten durch eine Fachperson des Archivs? So wäre es bspw. möglich, ein Disk-Image der Daten zu erstellen oder mittels rsync eine Kopie der Daten zu erstellen und diese als Bagl-konforme «bags» zu sichern.
 - Kann der Produzent Kopien auf Festplatten erstellen?
 - Wie erfolgt der Transport der Originaldatenträger (CD-R, Digital S, DVD-R, Festplatte, NAS usw.)?
 - Sollen Dateien über einen SFTP-Server versandt werden?
 - Sollen Dateien über eine API versandt werden?
 - Wie kann die Integrität der Daten während der Übertragung sichergestellt werden?

Der Umfang und die Grösse des Archivs bzw. die Komplexität des DMS bestimmen die Genauigkeit und Granularität der vorgängigen Bestandsaufnahme. Je umfangreicher der Bestand, desto gröber ist die Granularität. Normalerweise werden einige Stichproben genommen und detailliert ausgewertet, um aus diesen eine allgemeine Beschreibung des Gesamtbestandes zu extrapolieren.

4.5.3 Ablieferung

Nach Abschluss der vorgängigen Bestandsaufnahme kann die Ablieferung der Daten beginnen. Zu diesem Zeitpunkt werden einige Objekte priorisiert, während andere abgelehnt werden können, da sie bereits in der Sammlung vorhanden sind. Die Ablieferung umfasst entweder die Abholung vor Ort beim Ersteller oder die Lieferung an die Kulturerbe-Institution.

Bei Einsatz eines DMS läuft die Ablieferung mehr oder weniger kontinuierlich und meist online.

Dabei ist die Lieferung der digitalen Objekte an die Institution nicht immer die beste Strategie. So kann der Produzent beispielsweise die digitalen Objekte auf einen temporären Datenträger (z.B. Festplatte) kopieren, wobei wichtige Informationen verloren gehen oder verändert werden können, so z. B. (nicht abschliessend):

- Erstelldatum der Dateien
- Zeichensatzprobleme bei Ordner- und Dateinamen
- ursprüngliche Ordnerstruktur, in der die Dateien abgelegt wurden
- mögliche Abhängigkeiten zu Ordnern an anderen Speicherorten, die nicht Teil der Ablieferung sind
- Dateiintegrität

Abhängig vom Herstellungszusammenhang sollte ein Protokoll mit den physischen und personellen Ressourcen für die In-situ-Erfassung dieser Objekte erstellt werden.

4.5.4 Bestandsaufnahme

Die Bestandsaufnahme von digitalem Archivgut erfolgt in zwei Schritten. Der erste Schritt umfasst die Bewertung des physischen Objekts, der zweite Schritt die Bewertung des digitalen Objekts.

Physisches Objekt

Nach der Ablieferung erfolgt zuerst die Bestandsaufnahme der Datenträger. Jeder Datenträger wird erfasst und identifiziert (CD-R, DVD-RAM, externe Festplatte usw.) und eine grobe Zustandsbeschreibung (Schimmel, Bruch, Kratzer, Staub) wird erstellt.

Digitales Objekt

Die Bestandsaufnahme des digitalen Objekts erfolgt vor dem Ingest, also bevor ein SIP erstellt wird. Sie umfasst eine Reihe von Arbeitsschritten zur Diagnose, Analyse und Überprüfung des Zustands der digitalen Dateien. Um dabei Datenverlust oder -veränderung zu vermeiden, ist eine Arbeitskopie zu erstellen. Für zusätzliche Sicherheit kann eine weitere Kopie als Backup erzeugt werden.

Angeichts der potenziell grossen Datenmengen digitaler Archivbestände ist es nicht immer möglich, die Lesbarkeit aller vorhandenen Dateien zu überprüfen. Die Bestandsaufnahme umfasst

gemeinhin die folgenden Schritte: Extraktion der Metadaten, Dateiformaterkennung, Überprüfung der Gültigkeit von Dateiformaten, Überprüfung der Lesbarkeit der Datei.

- *Extraktion der Metadaten*

Gewisse objektspezifische Metadaten können durch Extraktion geprüft werden. Die Metadaten von PNG- und JPG-Bildern können bspw. mit Hilfe des Dienstprogramms *exiftool* in eine csv-Datei extrahiert werden:

```
exiftool -ext jpg -ext png -r -csv "/Verzeichnis/Quelle/Bilder" > ListeMetadatenBild.csv
```

Eine erfolgreiche Extraktion garantiert nicht die Lesbarkeit des Bildes, liefert aber einen ersten Hinweis auf den Zustand der Datei.

- *Dateiformaterkennung*

Die Dateieindung liefert einen ersten Hinweis auf die Art der Datei aber noch keine eindeutige Identifikation, da die Dateieindung falsch sein (z. B. txt anstelle von tif bei einer TIFF-Datei) oder fehlen kann.

Folgende Open-Source-Tools können zur Erkennung von Dateientypen verwendet werden:

- Fido (Format Identification for Digital Objects) ist ein von der *Open Preservation Foundation*¹⁴ entwickeltes Python-Programm.
- DROID ist ein von *The National Archives* entwickeltes Werkzeug. Sein Vorteil: Es verfügt zusätzlich zur Kommandozeile über eine GUI. Wie Fido nutzt es die Pronom-Datenbank von *The National Archives*.
- FITS ist ein serverbasiertes Java-Programm zur Formaterkennung, das meist über seine API genutzt wird. Es ist sehr schnell und bietet als Mehrwert die Integration mehrerer Werkzeuge zur Dateierkennung wie DROID, JHOVE, Exiftool oder auch Tika u. v. m.
- Siegfried nutzt ebenfalls Pronom als Datenquelle. Seine Installation ist auf der Site «IT for Archivists» beschrieben.¹⁵

- *Dateiformatvalidierung*

Die Dateiformatvalidierung unterscheidet sich von der -erkennung. Sie prüft die Standardkonformität der vorliegenden Dateistruktur. Beispielsweise kann eine Datei gemäss PDF/A-1b les- und verwendbar sein, gleichzeitig aber nicht dem PDF/A-1b-Standard entsprechende Metadaten oder Elemente enthalten. Sie kann also als PDF/A-1b identifiziert werden, ohne alle Spezifikationen des PDF/A-1b-Standards zu erfüllen.

Validierungsprogramme sind meist auf ein spezifisches Format ausgerichtet. Einerseits ist die Validierung in Abhängigkeit vom Dateiformat ein mehr oder weniger einfacher Prozess, andererseits sind die Software und das Repository nicht immer auf dem neuesten Stand. Daher ist es üblich, dass von verschiedenen Validierungsprogrammen gelieferte Ergebnisse nicht identisch sind und von Fall zu Fall interpretiert werden müssen. Die gebräuchlichsten Open-Source-Programme sind:

- Das PDF-Format kann mit Open-Source-Software wie VeraPDF, JHOVE oder mit proprietären Lösungen wie PDFtools validiert werden;
- TIFF-Dateien mit DPFManager, libtiff, JHOVE;
- JPEG-Dateien mit Bad Peggy, JHOVE;
- JPEG2000-Dateien mit jpylyzer.

- *Verlinkungen und Abhängigkeiten*

Einige Dateien enthalten Verlinkungen, die auf andere Dateien verweisen. Deshalb ist es

¹⁴ <https://openpreservation.org>. Siehe <https://github.com/openpreserve/fido#installation> für die Installationsanleitung.

¹⁵ <https://www.itforarchivists.com/siegfried/>

nicht möglich, eine Datei korrekt anzuzeigen oder abzuspielen, wenn die Verlinkungen zu den anderen Dateien defekt sind oder diese sekundären Dateien im Archiv fehlen. In diesem Fall kann die Hauptdatei nicht oder nur teilweise angezeigt oder abgespielt werden: Bspw. erscheinen leere Rahmen bei Bildern oder audiovisuelle Dateien sind teilweise ohne Ton. Es ist deshalb wichtig, den Zustand von Verknüpfungen zu prüfen und das Archiv auf Abhängigkeiten zwischen Dateien hin zu untersuchen.

4.5.5 Submission Information Package (SIP)

Wenn die Bibliothek über ein DLZA verfügt und die Daten nicht bereits in Form eines SIP geliefert werden, muss nach dem Datentransfer ein solches erstellt werden.

Das SIP sollte die relevanten Prüfsummen enthalten, um eine regelmässige Überwachung der Datenintegrität zu ermöglichen.

Da es kein Archivierungssystem und keinen Standard für SIPs gibt, wird empfohlen, das BagIt-Format¹⁶ zu verwenden, das in Zusammenarbeit zwischen der Library of Congress und der California Digital Library entwickelt wurde. Es gibt verschiedene Werkzeuge zur Erstellung und Verwaltung von «Bags», die dem BagIt-Format entsprechen. Die Library of Congress stellt beispielsweise ein Werkzeug namens Bagger¹⁷ zur Verfügung.

4.5.6 Platzbedarf (Triage)

Bevor SIPs zur Verarbeitung durch das DLZA vorbereitet werden können, sind Möglichkeiten zur temporären Lagerung und Speicherung einzurichten:

- für physische Datenträger: Schrank, Regal
- für die zu extrahierenden Dateien: Datenspeicher

Der Zeitraum, währenddem die Datenträger und Dateien zwischengelagert werden müssen, schwankt. Je nach den Kapazitäten der Institution kann es notwendig sein, Daten über einen längeren Zeitraum vorzuhalten. Daher wird empfohlen, Regalfläche und Datenspeicher auf eine Aufbewahrungsfrist von ein bis zwei Jahren auszurichten.

Die Eigenschaften des Zwischenspeichers für digitale Objekte werden daher wie folgt bestimmt:

- Ermittlung der jährlich zu archivierenden Datenmenge
- Berechnung des Zwischenspeicherumfangs
- Festlegung von Zugriffsrechten auf den Zwischenspeicher für die Mitarbeitenden (Unterscheidung zwischen Schreib- und Leserechten)
- Berücksichtigung des Standorts und der Eigentumsverhältnisse der Speicherinfrastruktur in Hinblick auf das Schutzniveau (DSGVO) und die Sensibilität der Daten. Beispiel: Jedes US-Unternehmen muss nach dem Cloud Act Zugang zu Daten gewähren, wenn ein US-Richter dies verlangt, auch wenn sich der Server in der Schweiz befindet.¹⁸
- Einrichtung eines Backup-Verfahrens für den Zwischenspeicher

¹⁶ <http://www.digitalpreservation.gov/documents/bagit-spec.pdf>

¹⁷ <https://github.com/LibraryOfCongress/bagger>

¹⁸ P. Fischer und S. Pittet, *US Cloud Act - un aperçu*, 08.11.2021. <https://swissprivacy.law/101>

Für die Einrichtung des Zwischenspeichers für die Triage gibt das von der British Library vorgeschlagene «Minimum preservation tool»¹⁹ eine Anregung.

4.5.7 Personalbedarf

Die für die einzelnen Arbeitsschritte benötigten Personalressourcen können anhand der folgenden Rollen bestimmt werden:

- Archivarin oder Konservator für die Bewertung der Relevanz des Bestands oder der Objekte
- Spezialist für Urheberrecht und private Daten
- Restauratorin oder digitaler Archivar für die Bewertung der Medien und Dateien
- IT-Spezialist für die Bereitstellung von Zwischenspeicher, Backups und die Verwaltung von Zugriffsrechten. Bestehen keine internen Ressourcen, kann von externen Anbietern Speicherplatz als Dienstleistung bezogen werden, sofern diese die Anforderungen bezüglich Zugriff, Kosten für die Bandbreite, Backups und Speicherort (in der Schweiz oder im Ausland) erfüllen.
- IT-Spezialistin für die Entwicklung von Anwendungen oder Schnittstellen zum manuellen oder automatischen Transfer digitaler Objekte

In einem Ablaufplan für die Datenübernahme werden die Rollen einer oder mehreren Personen zugewiesen.

5 Erhaltungsplanung (Preservation Planning)

5.1 Definition

Die Erhaltungsplanung für digitale Objekte ist nicht standardisiert und ihr Umfang ist je nach Institution mehr oder weniger klar festgelegt. Ziel dieser Planung ist die Erfassung aller dieser Bereiche in einem Referenzdokument, das allen an der Erhaltung der digitalen Objekte beteiligten Akteuren der Institution als Leitfaden dient.

Es sind mindestens die folgenden Bereiche zu planen:

- Erwerbungspolitik
- Ingest
- Langzeitarchivierung
- Bereitstellung
- Migration und Archivpflege
- Bewertung der Erhaltungsverfahren
- Zuweisung der Hauptaufgaben an das Personal
- Erstellung von Notfallplänen und Risikobewertungen

Diese Arbeit ist zeitaufwendig und relativ komplex, da hierfür unterschiedliche Bereiche der gesamten Institution (IT-Abteilung, Registratur, Urheberrechtsspezialist, Konservatorin usw.) beigezogen werden müssen.

¹⁹ <https://www.dpconline.org/blog/minimum-preservation-tool-mpt>

5.2 Erwerbungspolitik

Der Erhaltungsplanung soll eine Analyse der institutionellen Erwerbungspolitik vorangehen, um die Verschiedenheit und den Umfang des von der Institution zu erhaltenden digitalen Materials einordnen zu können.

Diese Analyse erlaubt es, pro unterschiedlichem Ablieferungskanal die jährlichen Menge an Neuerwerbungen zu ermitteln. Ablieferungskanäle können Ablieferungen aus der Pflichtabgabe (dépôt légal), der Erwerb von Archivbeständen oder Einzelobjekten, die regelmässige Ablieferung auf Grundlage von Vereinbarungen mit Datenlieferanten oder die Ablieferungen anderer Verwaltungsorgane aufgrund gesetzlicher Verpflichtungen sein.

Wurden die Ablieferungskanäle festgelegt, sind die mittels derselben gelieferten Typen digitaler Objekte zu ermitteln. Diese Bewertung sollte im Hinblick auf die digitalen Dateien und Datenträger sowohl qualitativ als auch quantitativ sein.

Bezieht eine Institution beispielsweise Daten über die Pflichtabgabe (dépôt légal), so ist deren Umfang zu prüfen. Handelt es sich nur um Text- und Fotodokumente, so lässt sich anhand dieser Information bereits eine Liste der Dateiformate erstellen, die über diesen Ablieferungskanal aufgenommen werden: unter anderem PDF/A1b oder PDF/A2u, ODT, SVG, TIFF, PNG, JPEG. Bezieht die Institution jedoch auch audiovisuelle Werke, so erweitert sich die Liste der Dateiformate um Formate wie mp4, WAV, MKV, MOV und weitere mehr.

Neben den Dateiformaten sind auch die möglichen Datenträger zu ermitteln. Nimmt eine Institution Daten auf optischen Medien entgegen, benötigt sie einen Arbeitsplatz, der CD-R, DVD-R, Bluray usw. lesen kann. Nimmt eine Bibliothek vornehmlich Fotobestände aus den 1990er Jahren auf, so könnte die Anschaffung eines ZIP- oder JAZZ-Laufwerks (Iomega) für den Zugriff auf Informationsträger in Betracht gezogen werden.

Um das jährlich zu verarbeitende Datenvolumen abschätzen zu können, wird eine quantitative Bewertung empfohlen, idealerweise detailliert für jeden Dokumententyp. Zum Beispiel:

- die Abspielzeit in Stunden für Audio- oder Video
- Anzahl Seiten für Text
- Anzahl Einzelaufnahmen für Fotografien

Damit digitale Objekte während der Datenübernahme identifiziert und verfolgt werden können, sind die pro Objekt verpflichtend vorhandenen beschreibenden Metadaten festzulegen. Dies hat so früh wie möglich zu erfolgen, damit diese Metadaten auch rechtzeitig verfügbar sind.

Die Analyse der Erwerbungspolitik hat in Zusammenarbeit mit folgenden Personen zu erfolgen:

- Sammlungsverantwortliche
- Verantwortliche für die Pflichtabgabe
- Verantwortliche für die Datenübernahme
- Metadatenverantwortliche (wenn vorhanden)
- Verantwortliche für das Urheberrecht (für die Erstellung der Übernahmevereinbarung)

5.3 Erwerbungsprozess digitaler Objekte

Wird der Erwerbungsprozess digitaler Objekte nicht durch ein DMS oder einen automatischen Übermittlungsprozess gesteuert, sollte er im Erhaltungsplan definiert werden. Zwei Szenarien sind möglich (siehe 4.3.54.5.3 Ablieferung):

- Der Datenlieferant übermittelt die digitalen Objekte an die Bibliothek.
- Die Bibliothek holt die Objekte beim Datenlieferanten ab.

Jedes Szenario hat seine Vor- und Nachteile. Die Rahmenbedingungen können sehr unterschiedlich sein und verschiedene Aspekte spielen eine Rolle, wie z. B. die Fachkompetenz des Datenlieferanten, die Art der Speichermedien, auf denen sich die digitalen Objekte befinden, oder die Dateisysteme dieser Speichermedien.

5.3.1 Technische Gegebenheiten beim Datenlieferanten

Der Datenlieferant kann die digitalen Objekte auf unterschiedlichen Medien gespeichert haben:

- NAS
- optisches Laufwerk
- externe Festplatte (mechanisch oder elektronisch)
- magnetisches Wechselmedium auf Diskette (u. a. Diskette, ZIP, Syquest)
- magneto-optisches Wechselmedium (miniDisc)
- interne Festplatte (Computer)
- USB-Stick
- Chipkarte (u. a. SD, SmartMedia, microSD)
- Cloud-Server oder Server mit Webschnittstelle

5.3.2 Erhaltungszustand

Der Erhaltungszustand der Datenträger kann stark variieren. So kann für den Zugriff auf die auf einem Datenträger enthaltene Information eine besondere Bearbeitung nötig sein.

5.3.3 Obsoleszenz

Ältere Speichermedien können spezielle Lesegeräte erfordern, die weder beim Hersteller noch in der Bibliothek vorhanden sind. So benötigt man zum Lesen von ZIP- oder Jazz-Disketten der Marke Iomega spezielle Laufwerke, die der Datenlieferant möglicherweise nicht mehr besitzt. Sind sie vorhanden, benötigt man zudem die passende Hard- und Software, um auf die auf ihnen gespeicherten Informationen zugreifen zu können.

5.3.4 Dateisystem und Zugriff auf das digitale Objekt

Ein Speichermedium allein legt nicht fest, auf welche Art die Information darauf gespeichert ist. Beispielsweise kann Information auf CD-R mit unterschiedlichen Dateisystemen gespeichert werden, so sind u.a. CDDA, iso9660 (mehrere Varianten existieren), CD-i - interaktiv, SACD möglich.

Um die Informationen vollständig zu erhalten, sollten möglichst nicht die einzelnen Dateien, sondern das gesamte Dateisystem eingelesen werden. Dies kann sich mehr oder weniger komplex gestalten, stellt aber sicher, dass neben den Dateien alle Metadaten und Berechtigungen erhalten

bleiben. Zudem wird so das digitale Objekt in seiner ursprünglichen Form erhalten. Werden beispielsweise bei einer DVD nur die Filmdateien übernommen, gehen die DVD-Menüs verloren, während die Übernahme des Dateisystems als Ganzes sämtliche Information (Menüs, Film und weitere Dateien) erhält und damit die Darstellung der DVD in ihrer ursprünglichen Form gestattet.

5.3.5 Vorbereitung der Daten und der Datenträger

Erfolgt die Datenübernahme nicht von einem DMS oder im Rahmen der Digitalisierung nach vorgegebenen Spezifikationen, lassen sich die Dateien kaum direkt an das für den Ingest zuständige System übermitteln. Um den Ingest zu ermöglichen, müssen vorbereitend erfahrungsgemäss zahlreiche, komplexe und zeitaufwändige Arbeiten durchgeführt werden, die im Erhaltungsplan zu berücksichtigen sind.

5.3.6 Pre-Ingest

Alle Operationen zur Vorbereitung der Datenpakete sowie Normalisierungen sind idealerweise in bestehenden Standards wie METS und PREMIS zu dokumentieren. Entsprechend sind Werkzeuge auszuwählen, die die Metadaten von digitalen Objekten lesen und in eine METS-Datei umwandeln können.

Zudem sollten Vorgaben zur Metadatenerfassung definiert werden, die festlegen, welche beschreibenden und spezifischen Metadaten für die Archivierung und Beschreibung benötigt werden.

5.4 Ingest (Übernahme)

5.4.1 Definition

Der Ingest umfasst die Arbeitsschritte zur Dokumentation und Umwandlung digitaler Objekte in ein archivierbares Datenpaket (AIP) sowie in ein Datenpaket für die Verbreitung (DIP). Der Ingest ist einer der zentralen Schritte der Archivierung.

Die folgenden Begrifflichkeiten sind damit verbunden:

- **Produzent (Datenlieferant)**

Der Produzent stellt die zur Archivierung vorgesehenen digitalen Objekte zusammen und übermittelt sie. Dabei ist er nicht zwingend der Urheber des Werks oder eine an dessen Erstellung beteiligte Person.

Im Archivbereich sollte dieser Begriff nicht mit dem Produzenten eines audiovisuellen Werks verwechselt werden, bei dem z. B. der Lichtbestimmer und der Cutter zwei Produzenten des digitalen Filmobjekts sind (der Lichtbestimmer produziert die endgültige Version des Films nach dem Cutter).

- **Konvertierung**

Die Konvertierung ist die Umwandlung des ursprünglichen Formats einer Datei in ein anderes Format, bestenfalls ohne dass der bedeutungstragende Inhalt der Datei verändert wird. Trotzdem führt eine Konvertierung immer zu einer Änderung von Information in der Datei. So ist z. B. die Umwandlung einer Excel-Datei in eine neue CSV-Datei eine Konvertierung. Die Konvertierung, z. B. bei digitalen Bildern, kann einen Informationsverlust bedeuten, muss dies aber nicht.

- **Migration**

Bei der Migration wird das Dateiformat einer Datei in eine neue Version des Dateiformats umgewandelt, was ihre Eigenschaften verändert. Normalerweise wird bei der Migration eine ältere Version eines Dateiformats in eine neuere überführt. Beispiel: Konvertierung eines TIFF 5.0 in ein TIFF 6.0.

5.4.2 Ingest-Arbeitsschritte

Im Folgenden werden – ohne Anspruch auf Vollständigkeit – die wichtigsten Ingest-Arbeitsschritte beschrieben:

- Auflistung der Dateien und der Organisationsstruktur des Archivs oder der Originalablieferung
- Erkennung von Dateiformaten
- Berechnung der Prüfsumme. Verwendet man ein Bag als SIP, kann die neu erstellte Prüfsumme mit derjenigen aus dem SIP verglichen werden. So kann die Integrität der im SIP angelieferten Daten geprüft und sichergestellt werden, dass sie beim Ingest erhalten bleibt.
- Auswahl der Dateien zur Übernahme, Löschen von nicht benötigten Dateien wie «.db» oder bestimmte versteckte Dateien des Dateisystems.
- Einholen der speziellen Metadaten zu den Dateien
- Konvertierung der Quelldateien in Dateien für die Erhaltung und/oder Bereitstellung
- Validierung der konvertierten Dateien. Eine Validierung der Quelldateien kann optional durchgeführt werden.
- Speicherung aller im Rahmen des Ingests durchgeführten Schritte in standardisierter Form in einem Protokoll, das als Datei im AIP abgelegt wird. Diese Datei sollte lesbar, nicht proprietär und wenn möglich standardisiert sein, damit sie am Ende des Ingests validiert werden kann. Der METS-Standard wird empfohlen.
- Durchführen des Ingests. Hierzu gibt es verschiedene proprietäre oder Open-Source-Tools.
- Speicherung des AIP und des DIP im DLZA. Der Vorgang der kurz- oder langfristigen Speicherung schliesst den Ingest ab. Er kann dabei Teil der Ingest-Verarbeitung sein oder auch nach Abschluss des Ingest-Workflows erfolgen. Die Frage der langfristigen Speicherung wird im folgenden Abschnitt behandelt.

Jeder der oben genannten Arbeitsschritte kann mithilfe von Softwareprogrammen automatisiert werden. Sämtliche der genannten Vorgänge sollten entsprechend den Standards und Empfehlungen der jeweiligen Institution eingerichtet werden.

Die britischen Nationalarchive haben eine nicht abschliessende Liste von Ingest-Softwarelösungen erstellt.²⁰

5.5 Dauerhafte Speicherung

Nach der Erstellung des AIP durch den Ingest muss es dauerhaft gespeichert werden. Dies ist mit unterschiedlichen technischen Lösungen zu erreichen. Deren Evaluation und Auswahl erfolgt gemeinhin durch die IT-Abteilung der Bibliothek oder der Verwaltung, zu der sie gehört, in enger Abstimmung mit der für die Erhaltungsplanung zuständigen Stelle.

²⁰ <https://cdn.nationalarchives.gov.uk/documents/archives/digital-preservation-repository-systems-for-archives.xlsx>.
Fedora ist ein Repository ohne Ingest-Funktionalität.

Der dauerhafte Speicher kann von einem internen oder externen Dienstleister bereitgestellt werden. In beiden Fällen ist sicherzustellen, dass dieser die Vorgaben zu Datenschutz²¹, Sicherheit, Serverstandort und Datensicherung einhält.

5.5.1 Speichertypen

Es lassen sich zwei Speichertypen unterscheiden:

- offline, mehrheitlich auf LTO-Band
- online, auf Festplatte

Online-Speicher ist teuer bei hohem Stromverbrauch, anfällig für Cyberangriffe und hat eine begrenzte Lebensdauer von 5 Jahren. Allerdings ermöglicht er einen schnellen Datenzugriff, z. B. für die Prüfsummenberechnung, das Kopieren oder Löschen.

Offline-Speicher ist langsam, relativ kostengünstig bei sehr geringem Stromverbrauch, vor Angriffen geschützt und hat eine Lebensdauer von bis zu 30 Jahren, wobei die Lesegeräte eine geringere Lebensdauer von 6 bis 15 Jahren aufweisen.

Wichtig ist, dass DIPs nicht auf Offline- sondern auf schnellem Online-Speicher vorgehalten werden, da für die Bereitstellung der DIPs nicht Datenintegrität und dauerhafte Speicherung, sondern schneller Zugriff wichtig sind. Aufgrund ihrer Ausrichtung auf die öffentliche Nutzung sind DIPs von der Entwicklung der Webtechnologien abhängig und müssen daher regelmässig an diese Entwicklung angepasst werden. Diese Anpassung erfolgt in Form einer Migration.

5.5.2 Migration

Datenmigration wird als die Umwandlung von Daten oder deren Struktur in eine neuere Form definiert. Dies bedeutet die Umwandlung des ursprünglichen Dateiformats in ein neueres Format. Zum Beispiel die Umwandlung einer Worddatei von DOCX mit Stand 2014 in ein DOCX der Version 19.0.

In Übereinstimmung mit dem OAIS-Modell ist die Umwandlung zu dokumentieren, wobei folgende Informationen wichtig sind:

- die ursprüngliche Datei
- die zur Umwandlung verwendeten Programme
- die Prüfberichte der Validierungssoftware

Wie bereits erwähnt, können DIPs regelmässiger als AIPs migriert werden.

In der Erhaltungsplanung ist ein Technologie-Monitoring festzulegen, das die Entscheidung erlaubt, wann und nach welchen Standards die Migration durchgeführt werden soll.

5.6 Speicher-Strategie

Die Erhaltungsplanung legt genau fest, wie die Speicherung der Daten zwecks langfristiger Erhaltung im System implementiert wird.

²¹ Verordnung über Datenschutzzertifizierungen (VDSZ). <https://www.fedlex.admin.ch/eli/cc/2007/701/de>

5.6.1 Bitstream-Preservation

Ein Bitstream ist Abfolge von Bits (Informationseinheit mit dem Wert 0 oder 1). Die Integrität der Dateien (digitale Objekte oder Metadaten) wird während verschiedener Verfahrensschritte überprüft, so zum Zeitpunkt der Aufnahme im Archivserver oder der Übernahme in die Backups. Der Prozess der Bitstream-Preservation berechnet die Prüfsumme und vergleicht diese mit der bereits in den Metadaten vorhandenen Prüfsumme.

5.6.2 Redundante Datenhaltung

AIPs, DIPs und temporäre Arbeitskopien werden zur besseren Sicherstellung der Verfügbarkeit auf ein anderes Speichermedium kopiert. Der Kopiervorgang kann synchron oder asynchron sein. Im zweiten Fall ist die Kopie nicht automatisch mit der Quelle identisch.

5.6.3 Backup (Sicherung)

Das Backup ist eine zu einem bestimmten Zeitpunkt erstellte Kopie der im DLZA enthaltenen digitalen Objekte zu Sicherungszwecken. Die Backup-Häufigkeit und die Backup-Art (vollständig oder inkrementell (nur die Unterschiede zur vorherigen Kopie werden gespeichert, um den Speicherplatz zu minimieren) sind festzulegen. Auch ist zu bestimmen, ob die Sicherung auf der Ebene von Dateien, Datenträgern oder virtuellen Maschinen erfolgt.

Das Backup muss auf verschiedenen Speichermedien gespeichert und an verschiedenen Orten aufbewahrt werden. Die Aufbewahrungsdauer des Backups richtet sich nach den Vorgaben der Institution.

Erwähnenswert ist auch der Snapshot (Schnappschuss), der das komplette System zu einem bestimmten Zeitpunkt sichert. Snapshots werden häufiger als Backups erstellt aber nur für einen begrenzten Zeitraum auf derselben oder einer anderen Serverinfrastruktur aufbewahrt.

Bei der Wiederherstellung des Backups wird das System auf einen früheren Zustand zurückgesetzt, z. B. auf jenen des Vorabends. Das bedeutet, dass im schlimmsten Fall die Arbeit eines Tages verloren gehen kann.

Die Kombination aus Backup und Snapshot hingegen ermöglicht die Wiederherstellung des Systems zu einem sehr genauen Zeitpunkt: Stunde, Minute.

5.6.4 Zusammenspiel von Backup und Bitstream-Preservation

Die Backup- und Bitstream-Preservation-Zyklen müssen aufeinander abgestimmt sein.

Angenommen, es findet keine Bitstream-Preservation statt, so werden Fehler in die Sicherung kopiert. Nach Ablauf der Backup-Periode, z. B. nach drei Monaten, ist es dann nicht mehr möglich, die korrekten Daten wiederherzustellen. Um dies zu verhindern, muss der Bitstream-Preservation-Zyklus kürzer als drei Monate sein – einschliesslich der Zeit für die Fehlerbehandlung.

Alternativ kann im Fehlerfall eine Kopie der Backups auf einem schreibgeschützten Datenträger angefordert werden.

5.6.5 Cold-Storage

Der Zugriff auf den Speichertyp *Cold-Storage* (Offline-Speicher) ist sehr langsam. Er dient nur dem sporadischen Zugriff und ist deshalb nicht an eine Anwendung – wie z. B. einen digitalen Lesesaal – angebunden. Dieser Speicher eignet sich nicht zur Bearbeitung von Daten. Dazu muss eine Kopie auf einem leistungsfähigeren Speicher angefordert werden. Gemeinhin sind bei diesem Speicher die Kosten für die Speicherung tief und die Kosten zum Auslesen der Daten hoch, wobei dies natürlich vom Preismodell des Rechenzentrumsanbieters abhängt.

Ein typisches Speichermedium für Cold-Storage ist das LTO-Band, auf das nur geschrieben und von dem nichts gelöscht wird.

Im Rahmen der digitalen Langzeitarchivierung ist festzulegen, ob nur AIPs oder auch DIPs auf dem Cold-Storage gesichert werden. Zudem ist zu klären, wie oft und auf welche Art und Weise die Datenqualität auf dem Cold-Storage geprüft wird.

5.7 Technologie-Monitoring

Das Technologie-Monitoring umfasst die vorausschauende Beobachtung und Bewertung der im Lauf der Zeit auftretenden technologischen Entwicklungen. Es ist von entscheidender Bedeutung, um die kontinuierliche Erhaltung des DLZA zu gewährleisten und das mögliche Risiko eines Zugriffsverlustes auf digitale Objekte zu mindern. Es muss sich dazu mit allen Elementen befassen, die Teil des technischen Ökosystems des DLZA sind. Dies umfasst neben der Hardware besonders Software, Dateiformate und Dateisysteme.

So ist sicherzustellen, dass die eingesetzte Hardware nach wie vor «unterstützt» wird, d. h. leicht repariert werden kann. Beispielsweise findet im Bereich Cold-Storage ein Wechsel vom LTO8- zum LTO9-Standard statt. Noch unterstützen die Hersteller von Bandlaufwerken oder -robotern LTO8. Zukünftig werden sich die Wartungsverträge jedoch verteuern oder die Hersteller werden die Wartung von LTO8 nicht mehr anbieten. Es ist daher ratsam, diese technologische Entwicklung zu beobachten, damit die Umstellung auf LTO9 geplant werden kann, bevor der kritische Fall auftritt, dass die eingesetzten LTO8-Laufwerke nicht mehr funktionieren und sich nicht mehr reparieren lassen.

Dasselbe gilt für die Überwachung von Dateiformaten. So unterstützen beispielsweise die wichtigsten Webbrowser allmählich den H265-Codec. Da er bei gleicher Qualität stärker komprimiert, kann die Beobachtung seiner Entwicklung interessant sein, um ihn rechtzeitig in DIPs einzusetzen. Die Analyse der Unterschiede zwischen den Codecs H264 und H265-Codec ist ein typisches Beispiel des Technologie-Monitorings: Bei H265 reduziert sich die Datenmenge im Vergleich zu H264 um die Hälfte.

5.8 Metadaten

Bibliographische bzw. deskriptive Metadaten haben im DLZA die Funktion, ein im AIP gespeichertes digitales Objekt eindeutig identifizieren zu können. Dies ist unter zwei Gesichtspunkten wichtig. Einerseits werden zur Identifikation von digitalen Objekten zur Verwaltung und Nutzung des DLZA bibliographische Metadaten benötigt und andererseits ist es eine gute Praxis, dem AIP sämtliche deskriptiven Metadaten mitzuliefern, die der Identifikation des im AIP gespeicherten digitalen Objekts dienen – und nicht nur diejenigen, die vom Langzeitarchiv zur Verwaltung benötigt werden. Dies erhöht die Chance, dass das im AIP gespeicherte digitale Objekt auch

nach einem Export aus dem DLZA (oder anhand von Backups im Katastrophenfall) eindeutig identifiziert werden kann (selbsttragendes AIP).

5.8.1 Wahl eines Standards

Die Wahl des Metadatenstandards ist abhängig von der Art des Einsatzes des DLZA in der Gedächtnisinstitution. Normalerweise besteht neben dem DLZA ein weiteres Sammlungsverwaltungssystem, in dem die Gedächtnisinstitution ihre analogen, physisch vorhandenen Objekte verwaltet (z. B. Bibliothekssystem). Dabei orientiert sie sich an den in ihrem Bereich üblichen Standards, die sowohl den Metadatenstrukturstandard als auch das Regelwerk zur Erschliessung von Objekten (z. B. RDA, ISAD(G), SPECTRUM) vorgeben. Da ein solches Sammlungsverwaltungssystem bereits besteht, bietet es sich an, die digitalen Objekte ebenfalls in diesem System zu erschliessen, damit die Sammlung integral in einem System beschrieben ist. Dies führt dazu, dass Metadaten digitaler Objekte bereits im Metadatenformat der Sammlungsverwaltung vorliegen und entsprechend im digitalen Langzeitarchiv verwendet werden können und sollen.

Dieses Vorgehen ist besonders mit Blick auf eine möglichst einfache Organisation der Systeme von Vorteil. Es bringt allerdings auch Probleme mit sich. Werden bspw. Archivalien in einem LMS erschlossen, reduziert das viele Möglichkeiten der Beschreibung. Teilweise müssten auch deskriptive Informationen erfasst werden können, die nicht jede Nutzerin und jeder Nutzer sehen sollte, die aber langfristig gesehen wichtig sind (bspw. bestimmte Erwerbungsinformationen oder inhaltliche Beschreibungen von sensiblen Daten). Eine Einschränkung der Sichtbarkeit bieten viele LMS nicht an.

Hinzu kommt, dass allenfalls bereits mehrere Datengebersysteme vorhanden sind (bspw. Repositorien). Dann würde es sich empfehlen, eher diese verschiedenen Systeme anzubinden und ihre jeweiligen Metadaten mittels API ins DLZA zu laden und dabei zu mappen.

Die Wahl des Metadatenstandards ist zudem abhängig von der Verwendung der bibliographischen Metadaten im DLZA. So ist es sinnvoll, für die zur Identifikation im DLZA benötigten Metadaten einen Standard zu verwenden, der von diesem unterstützt wird. Im Rahmen des Ingests sind dann die für die Verwaltung des DLZA benötigten Informationen aus den Metadaten der Sammlungsverwaltung in das DLZA zu überführen. Macht die Gedächtnisinstitution in diesem Bereich Vorgaben, so kann dies negative Auswirkungen haben. Sie schränkt die Wahl des DLZA ein, was bei einer Beschaffung in einem kleinen Anbietermarkt zu fehlenden Angeboten führt. Dies kann umgangen werden: Da die meisten modernen Metadatenstandards XML-basiert sind (wie weiter unten beschrieben), lassen sich in der Entwicklung meist einfach Mappings entwickeln und für die Suche indexieren. Containerformate wie z. B. METS sind zudem recht offen für verschiedene Standards deskriptiver Metadaten, die darin integriert werden können.

Geht es aber um die Speicherung der bibliographischen Metadaten als Teil des selbsttragenden AIP, so sind die bibliographischen Metadaten vollständig im von der Gedächtnisinstitution vorgegebenen Format zu erhalten, da sie dieses Format bereits kennt.

Es folgen allgemeine Überlegungen zu Metadatenformatstandards, die die Maschinenlesbarkeit von Metadatenstrukturen sicherstellen. Im Umfeld der Gedächtnisinstitutionen existieren verschiedene domänenspezifische Metadatenformatstandards (z. B. MARC21, MODS, EAD, RiC, LIDO). Es besteht kein domänenübergreifender Standard, der die Strukturierung von Metadaten verschiedener Typen von Gedächtnisinstitutionen ohne Informationsverlust erlaubt.

Ein gegenüber den domänenspezifischen Standards verlustbehafteter Metadatenformatstandard ist Dublin Core. Domänenspezifische Metadaten können zwar nach Dublin Core übertragen wer-

den, aber eine Rückübertragung ohne Bedeutungsverlust ist nicht mehr möglich. Der Einsatz von Dublin Core für die Ablage der bibliographischen Metadaten im selbsttragenden AIP ist also nicht empfehlenswert. Vielmehr empfiehlt es sich, die bibliographischen Metadaten dem AIP in dem Metadatenformat beizulegen, das bei der jeweiligen Gedächtnisinstitution verwendet wird.

Dabei ist darauf zu achten, dass sämtliche zur späteren maschinellen und intellektuellen Auswertung des Metadatenformats benötigten Informationen ebenfalls bereitgehalten werden, da sich sonst langfristig nicht garantieren lässt, dass die dem AIP beigelegten Metadaten noch ausgelesen und verstanden werden können. Bietet das Verwaltungssystem (z. B. Bibliothekssystem, Archivsystem, Museumssystem) mehrere Exportmöglichkeiten für bibliographische Metadaten an, so ist bevorzugt ein international verbreiteter, gut dokumentierter Datenformatstandard zu wählen. Der Einsatz eines solchen Standards erhöht aufgrund bestehender und gepflegter Dokumentation die Wahrscheinlichkeit, dass die bibliographischen Metadaten längerfristig noch ausgelesen und verstanden werden können.

Hat die Gedächtnisinstitution ihre Metadatenstruktur und die von ihr verwendeten Felder eigenständig und ohne Bezug zu einem international verbreiteten Standard festgelegt, muss sie zwingend sämtliche Informationen bereithalten, um die Maschinenlesbarkeit und die intellektuelle Auswertung über einen langen Zeitraum zu sicherzustellen. Die laufende Pflege und Dokumentation des Metadatenformats ist in diesem Fall zwingend durch die Gedächtnisinstitution sicherzustellen. Wenn die AIPs selbsttragend sein sollen, müsste konsequenterweise in diesem Fall die Dokumentation des Formats sogar zwingend in jedem AIP mit abgelegt werden, nicht nur an zentraler Stelle bei der Institution.

Da zum aktuellen Zeitpunkt die Containerformate von AIPs normalerweise auf dem XML-Datenformat aufbauen, empfiehlt es sich, entweder direkt einen XML-basierten Metadatenformatstandard (z. B. EAD, LIDO oder MODS) zu verwenden oder einen Metadatenformatstandard einzusetzen, für den eine XML-Version existiert (z. B. MARCXML). Dies erlaubt es, die bibliographischen Metadaten direkt in die Datenstruktur des AIP einzubinden, da der XML-Standard dies vorsieht (Kombinieren von XML-Schemata).

Ob eine solche Einbindung im Kontext des individuellen DLZA sinnvoll ist, muss mit dem Anbieter des Systems geklärt werden. Integriert in das Container-XML können deskriptive Metadaten indexiert und für die Suche im Archiv genutzt werden. Ist die direkte Einbindung in das Containerformat im DLZA nicht möglich, so sind die bibliographischen Metadaten als eigenständige Datei dem AIP hinzuzufügen. Liegen sie als eigene Datei im AIP, ist aber die Indexierung der Metadaten oftmals nicht oder nur mit zusätzlichem Entwicklungsaufwand möglich. Dessen sollte man sich bewusst sein und allenfalls mit dem Anbieter eine Lösung finden, die die Suche nach deskriptiven Metadaten ermöglicht.

5.8.2 Technische oder objektspezifische Metadaten

Technische Metadaten unterscheiden sich von beschreibenden Metadaten. Letztere beziehen sich auf allgemeine Elemente wie Titel, Autor, Dauer, Grösse, Datum usw. und geben Auskunft über das Objekt in seiner intellektuellen Dimension. Technische Metadaten beziehen sich dagegen auf Informationen, die eher von den Eigenschaften des Objekts abhängen. Die technischen Metadaten für ein bewegtes Bild sind beispielsweise:

- die Auflösung
- der Kompressionsalgorithmus für Audio und Video (Codec)
- der Farbraum

- die Farbtiefe
- die Bildwiederholrate pro Sekunde

Diese Metadaten sind in der Regel Teil der Dateien selbst. Ihre Organisation und Beschreibung innerhalb der Datei ist standardisiert, damit sie von verschiedenen Programmen gelesen werden können.

Für technische Metadaten der unterschiedlichen Objektkategorien existieren jeweils eigene Standards. So gibt es eigene Standards für Standbilder, audiovisuelle Dateien, 3D usw.

Technische Metadaten in Standbildern können beispielsweise in den Standards EXIF, IPTC oder XMP vorliegen, während für technische Metadaten in audiovisuellen Dateien die Standards EBU-core, MXF, MPEG7, Quicktime usw. verwendet werden.

5.8.3 Erhaltungsmetadaten

Alle Informationen über die während des Ingests und der Archivierung durchgeführten Verarbeitungsschritte sind für die Erhaltung zu verwendende Metadaten. Im weitesten Sinn sind sämtliche für die Erhaltung digitaler Objekte des DLZA benötigten Metadaten als Erhaltungsmetadaten zu betrachten. Die Grenze zwischen technischen Metadaten und Erhaltungsmetadaten ist fließend. Sie unterscheiden sich nicht in ihrer Art, sondern hinsichtlich ihrer Verwendung.

So ist beispielsweise die Version des für die Kodierung einer Videodatei verwendeten Codecs eine zu den technischen Metadaten gehörende Information, die einerseits für einen Forscher von Interesse ist, der die Hintergründe eines Videofilms studiert, und andererseits eine für die Erhaltung des digitalen Objekts relevante Information. Dies weil die Information zum Codec es erlaubt, im Rahmen einer Erhaltungsmassnahme Entscheidungen über die Konvertierung oder Migration des Videos zu treffen. Damit unterscheidet sich der Einsatzzweck dieser technischen Metadaten im Rahmen der Erhaltung von der ursprünglich vom Produzenten vorgesehenen Nutzung (EBUcore). Aus diesem Grund wurden für Erhaltungsmetadaten eigene Standards entwickelt, die ihrem Einsatzzweck besser gerecht werden. Drei Standards werden allgemein eingesetzt: METS, PREMIS, SEDA (Frankreich).

6 Zugänglichkeit und Nutzbarkeit

Digital archivierte Objekte können auf verschiedene Weise zugänglich gemacht werden. Grundsätzlich wird in diesem Kapitel – insbesondere bei Film- und Tondateien – nicht zwischen einem Zugang via Download oder Stream unterschieden.

Eine Möglichkeit, Dateien in einem Langzeitarchiv zugänglich zu machen, ist die Nutzung einer vom DLZA bereitgestellten Oberfläche. In diesem Fall verfügt das DLZA über ein Modul, wie z. B. einen Reader oder einen Player, das speziell dafür entwickelt wurde, um die Zugänglichmachung der archivierten Objekte zu erleichtern.

Eine weitere Möglichkeit ist die Nutzung eines eigenständigen Portals der Institution, das mit einer Schnittstelle ans DLZA angebunden ist. Über dieses Portal können die archivierten Objekte dann von Benutzerinnen und Benutzern eingesehen und heruntergeladen werden. Die Übergabe der Dateien vom Archiv ans Portal kann automatisiert oder kuratiert erfolgen, je nach den Anforderungen des Archivs.

Eine dritte Möglichkeit besteht darin, die archivierten Objekte über externe Portale zugänglich zu machen. Dies setzt jedoch voraus, dass eine Nutzungserlaubnis vorliegt und die eigenen Dateitypen mit den Vorgaben des Portals kompatibel sind. Es gibt zahlreiche Plattformen, die oft nach Medien oder Themen (Manuskripte, Plakate, Audiovisuelles usw.) geordnet sind. Sie erleichtern die wissenschaftliche Suche und vereinheitlichen die Beschreibungen.

Meist bleibt jedoch ein «Rest» von Dateien, der nirgends zugänglich gemacht und nachgenutzt werden kann. Der Zugang kann beispielsweise durch das Format der Dateien erschwert werden. So können Dateien in Formaten, die für bestimmte Berufe oder Bereiche spezifisch sind, online nur schwer aufgerufen bzw. geöffnet werden. Beispiele dafür sind digital erstellte Partituren, Dateien, die mithilfe eines Desktop-Publishing-Tools (z. B. InDesign) erstellt wurden oder Architektenpläne, die Anzeigewerkzeuge erfordern, die sich nicht in die Präsentationsoberfläche des digitalen Langzeitarchivs integrieren lassen. Eine mögliche Lösung besteht darin, dass die Dateien in ein gängiges Format wie PDF migriert werden, entweder beim Aufrufen im DLZA oder allenfalls während eines Downloadvorgangs.

Auch Dateien unter rechtlichen Zugangsbeschränkungen wie Sperrfristen können nicht ohne weiteres nachgenutzt werden. Dies gilt auch für Dateien, die aus urheberrechtlichen Gründen nicht frei im Internet einsehbar sein dürfen, sondern deren Zugriff strikt kontrolliert und restriktiv erfolgen muss. In diesem Fall ist es sinnvoll, die Bereitstellung in einem gesicherten Online-Raum zu prüfen. Eine weitere Möglichkeit bietet die Konsultation vor Ort über einen dedizierten Arbeitsplatz in der Bibliothek. Es gibt eine Reihe von Möglichkeiten, Zugangsbeschränkungen technisch umzusetzen, indem z. B. kein Reader öffnet oder die Datei selbst nicht an ein Präsentationsportal weitergegeben wird.

Dateien, deren Zugang und Nachnutzung restriktiv gehandhabt werden muss, stellen spezielle Anforderungen an ein Zugangssystem. Im Archivbereich hat sich dafür die Bezeichnung *digitaler Lesesaal* durchgesetzt. Dort wird der archivierte Gesamtbestand verzeichnet, auch die unter einer Sperrfrist stehenden Objekte. Wird eine Datei zur Nutzung angefragt, wird sie in ein Format migriert, in dem sie angesehen und heruntergeladen werden kann.

Als letztes gibt es *Eaasi – Emulation as a service infrastructure*²². Hier wird im Browser eine Emulation der (alten) Systemumgebung erstellt und die Datei angezeigt. Derzeit ist Eaasi noch ein separates Projekt und nicht in einem Archivsystem integriert. Vorstellbar ist, dass Eaasi als Modul eines digitalen Lesesaals eingerichtet werden kann. Anhand der dem AIP beigelegten Metadaten wird klar, welche Art von Emulation mitgeliefert werden muss. Im digitalen Lesesaal wird dann angezeigt, dass es eine solche Emulation gibt. Mit Klick auf den Link öffnet sich im Browser dann die passende Emulation, die auch die zugehörigen Originaldaten des jeweiligen AIP enthält.

Durch die Weiterentwicklung und Adaption von Eaasi für die Langzeitarchivierung könnte die Migration von Dateiformaten unnötig werden, was die rein technische Zugänglichkeit der zu archivierenden Dateien über die Zeit hinweg enorm erleichtern würde. Die rechtlichen Vorgaben für den Zugang und die Nachnutzung digitalen Archivguts gilt es in jedem Fall einzuhalten. Ausserdem macht auch Emulation-as-a-service-Infrastructure die Migration von Dateien nicht obsolet. Sie bleiben als Backup-Plan notwendig, um den Erhalt der Daten zu gewährleisten.

²² <https://www.softwarepreservationnetwork.org/emulation-as-a-service-infrastructure/>

7 Zusammenfassung und Schlussbetrachtung

Im Juni 2023 veröffentlichte der Bund die *Kulturbotschaft 2025–2028*²³. In ihr wird der Wille bekundet, die Aufgaben der Nationalbibliothek im digitalen Bereich zu stärken. Mittels Revision des Nationalbibliotheksgesetzes soll der Umfang ihrer Tätigkeit im digitalen Bereich geklärt werden, nämlich das Sammeln von Informationen über die Schweiz, unabhängig davon, ob sie mit oder ohne physischen Träger vorliegen. Zentral ist dabei die Einführung eines digitalen Pflichtexemplars.

Die Aufgabe der NB zur Koordination mit den Kantonsbibliotheken und anderen Gedächtnisinstitutionen (Archive, Museen) im Bereich der digitalen Archivierung wird bekräftigt. Für Gedächtnisinstitutionen sind die Rahmenbedingungen zur Konkretisierung oder Vertiefung von DLZA-Projekten daher besonders günstig und vielversprechend.

Digitale Inhalte herrschen heute klar vor. Nicht nur weil die meisten von ihnen unabhängig von der Form (Text, Ton, Standbild und Bewegtbild) direkt so hergestellt werden («born digital»), sondern auch weil die Digitaltechnik zu einem bevorzugten Medium für die Aufbewahrung bestimmter analoger – insbesondere audiovisueller – Medien geworden ist, die fragil und schlecht haltbar sind («Retrodigitalisate»).

Alle diese Inhalte zusammen sind daher wichtige historische Zeugnisse aus dem 20. und 21. Jahrhundert. Die Kantonsbibliotheken sind sich ihrer Verantwortung für dieses Erbe bewusst.

In gewisser Weise ähnelt die heutige Situation bei der digitalen Archivierung der Informatisierung der bibliothekarischen Aufgaben vor 30 bis 40 Jahren: Nur wenige Institutionen – insbesondere Universitäten – hatten ein IT-System und erst nach und nach setzten sich Bibliotheksverwaltungssysteme bei den unterschiedlichen Bibliotheken flächendeckend durch.

Dieser Leitfaden zeigt, dass die Institutionen bei der Einführung eines digitalen Archivs nicht auf sich allein gestellt sind: Werkzeuge, Methoden und Standards sind vorhanden. Die Erfahrung mit der digitalen Archivierung nimmt dank der Ausbildung von Fachleuten und erworbener Kenntnisse zu. Unterstützen können zudem unsere Hochschulen, spezialisierte Organisationen wie KOST/CECO oder Memoria v sowie mehrere Unternehmen, von denen einige auf Bibliotheken und Archive spezialisiert sind.

Es gibt jedoch einen nicht mehr zu vernachlässigenden exogenen Faktor, der die gesamte Gesellschaft beeinflusst: die Energiekrise und damit die Notwendigkeit, über einen ressourcenschonenden Betrieb nachzudenken. So wie es zur Aufbewahrung physischer Dokumente neue Gebäude gibt, die die klimatischen Bedingungen mit einem Minimum an Energie regulieren können, müssen auch DLZA ressourcenschonend sein. Dies bedeutet, dass Entscheidungen in verschiedenen Phasen des Archivierungsprozesses getroffen werden müssen:

- Der Umfang der zu archivierenden elektronischen Inhalte macht eine Auswahl oder Triage nötig.
- Der Speicherbedarf der Dateiformate ist vorrangig zu berücksichtigen, besonders bei audiovisuellen Inhalten.
- Das Gleichgewicht zwischen Online-, Nearline- und Offline-Archivierungsmethoden ist zu wahren und die Wahl einer angemessenen Anzahl von Kopien zu bedenken.

²³ <https://www.bak.admin.ch/bak/de/home/themen/kulturbotschaft.html>

Die aus Kulturerbe-Bibliotheken des ganzen Landes stammenden Redaktorinnen und Redaktoren dieses Leitfadens sind ferner davon überzeugt, dass Gedächtnisinstitutionen von einer Zusammenarbeit auf dem Gebiet der digitalen Archivierung profitieren können. Durch Erfahrungsaustausch aber auch durch Vernetzung ihrer Kapazitäten: So ist es beispielsweise durchaus denkbar, dass eine Bandkopie von St. Galler Archivdaten im Wallis aufbewahrt wird und umgekehrt.

Wir können davon ausgehen, dass in den nächsten Jahren kollektive Lösungen und Organisationsmodelle entwickelt werden, um die Archivierung des digitalen Erbes aller Kantone auf rationale und effiziente Weise zu gewährleisten.

8 Anhang

8.1 Schlüsselfragen

8.1.1 Kapitel 2: Allgemeine Fragen

Resources (human and materials) inside, financial funds

- Do you have IT resources in your institution from Human Resources and hardware perspectives? (Have you already planned or evaluated how many IT resources could be available for the digital long-term archive?)
- Do you already collaborate with some a network of institutions or mutualize resources for other project (storage facilities, IT resources, conservation lab, exhibition production ...)? (may be placed somewhere else)
- Do you have the funds - and subsequently the personnel - not only to ingest digital resources but to preserve them?
- Is your institution already capable of long-term preservation of digital objects in terms of personnel? If not - are you capable of allocating the personnel?

Scheduling

- Do you have already a deadline or have you scheduled a plan to setup a long-term archival system?

evaluation of the archivable material

- Have you evaluated approximatively the volume of data your institution would preserve? Have you evaluated the increase of your archive per year?
- What kind of data your institution would preserve? Born digital material from your institution; born digital material created outside of the institution; digital material from digitization.
- Do you already have a collection and preservation strategy for your digital data? (What you want to collect and how you want it to preserve)
- Do you have a collection management system for your analogue collection?

collaboration

- Is the data you want to preserve already part of another's institutions collection strategy?

legal framework of your institution

- Does your legal framework allow the long-term preservation of digital objects?
- What is the scope of your long-term preservation, do you want simply to preserve your collection or additionally open it to the public? (I think we can recommend people with limited resources to focus on preservation project first; then when ingest-archiving process is done, implement the access part)

diversity of platform: management part and the access part

- If you open your collection to the public, how do want it to be presented - separately as a digital collection or together with your main collection of analogue materials?

management of the collection (system and human resources)

- How should the preservation of digital objects be integrated in your institution's workflows? Is a separate team doing it or do you want it to be integrated with your current workflows and teams?
- continuing education or new team?

8.1.2 Kapitel 3: Systementwicklung

Core question: How to build a digital long-term archive from scratch?

Includes aspects such as the conception of a digital long-term archive, the development of cross-system workflows and project planning, an overview of vendors, architectures and standards, important principles, cost issues and maintenance...

- Which organizational questions should be answered, if you plan to design a long-term archive?
 - What is our goal, our mission statement? What do we want to archive, for what and for whom? Who (department, team) is responsible for this?
 - Priorities: Are priorities set, with which datatype you want to begin? Do you have a strategy begin with the most urgent, the biggest heap, the most complex use case?
 - Are there external factors that influence the roadmap, the available budget or the choice of systems?
 - Which legal and contractual regulations are relevant for our institution?
 - What form of access and usage are we aiming for?
 - Who are our internal and external stakeholders? What are their needs?
 - What is the internal perception of the topic? Do we need to promote awareness?
 - What could our exit strategy look like? Is there an institution that could take over services and data in an emergency?
- Which questions regarding resources and cooperation should be answered, if you plan to design a long-term archive?
 - What are the skills / is the job description of the person (to be) in charge of digital long-term preservation within a specific organisation?
 - Do we have the financial and human resources to accompany the development and operation of a long-term archive?
 - Who should work with the software? Librarians, archivists or IT staff? It could be necessary to do some tasks manually, e.g., build structure of a SIP.
 - Do we need or can we use an existing service instead of building our own long-term archive? How do we deal with such issues in principle (outsourcing versus self-determination)?
 - With which partners will we necessarily or usefully have to work with (cantonal or university IT, providers of storage infrastructures, providers of library and archive systems ...)?
 - Is your future digital long-term digital archive to be integrated in an existing IT architecture?
- Which technical-conceptual questions should be answered, if you plan to design a long-term archive?
 - Who are our data providers (persons, institutions, systems)? Which file formats and data sources have to be integrated (e.g., data carriers in archives; digital publications like e-books, journals, databases, websites; digitized cultural heritage)? If you have a DAM, do you think it is relevant for the LTA?

- What else do you want to integrate in your AIP apart from the digital object and its administrative metadata? Bibliographic metadata? If so, do you want this data to be up-to-date? If it has to be up-to-date in which system, do you plan to keep it up-to-date?
- What volume of data - initially and in estimated annual growth - will be ingested into the long-term archive?
- Who are our data users (persons, institutions, systems)?
- Which level of security is required for your stored data? Are the objects that are going to make up your collection legally sensitive? If so in which sense and what are the consequences? (Localisation of the storage)
- Which financial questions should be answered, if you plan to design a long-term archive?
 - What are the initial costs (points) for setting up and developing a long-term archive?
 - What are the annual costs (points) of running a long-term archive?
 - What hidden costs should be taken into account (e.g., internal project support, knowledge development, costs in cooperation with partner ITs ...)?
- What are the characteristics that a minimal (maybe temporary, until a complete solution is available) digital preservation system should have (see also: <https://www.scimeeting.cn/m/video/play/2?vid=172068>)?

8.1.3 Kapitel 4: Richtlinien für den Erwerb

Core question: What are our guidelines, ideas and wishes for acquisition and preservation of digital objects?

- Do you have already set channel for producer to deposit digital material? Do these producers have their own wishes and ideas that are of relevance for preservation and accessibility?
- Which entities (persons, departments, systems) should deliver data in the future? Can these be connected automatically via interfaces or do delivery tools have to be included in the planning?
- Which wishes do we already have for accessibility during acquisition?

Subchapter 4.1: Born digital material and acquisition through APIs

- Are there existing standards for data delivery (APIs, data formats, file formats, SIP)?
- How can you ensure the integrity of the data?
- Which possibilities have data providers which we are cooperating with to deliver the file formats and metadata we are demanding?

Subchapter 4.2: Digital historical material or material from analogue carriers

- Do we have the necessary hardware for reading data carriers or do we need them? Or do we need partner organizations which can fulfil this task for us?
- How can you ensure the integrity of the data if it isn't ingested immediately in the long-term archive?
- how to handle original carriers or media before and after ingest?

8.1.4 Kapitel 5: Erhaltungsplanung (Preservation Planning)

Core question: How do we take care of the objects in a long-term archive?

Includes aspects such as setting up ingest workflows (I would say that ingest may be describe in chapter 2 - better not, because the ingest is a detailed process with many aspects. Chapter 2 is more focused on the pre-ingest and the fact, how data comes from a producer to our infrastructure), migrating objects of different natures (text, image, AV ...), tool recommendations, storage strategies, recommendations for formats suitable for archiving or where to find them ...

Subchapter 5.1: Ingest workflow creation and optimization

- Which work steps in an ingest workflow have to be carried out and to what extent can they be automated? Which steps are optional or recommendable? (may be placed in Chapter 2)
- To which extent can I or do I want to influence my workflow?
- Potential for optimization (workflows, hardware, parallelization, ...)
- Risks of optimization

Quality control and data curation

Subchapter 5.2: Strategy for storage

- How do you ensure bit preservation?
- What is the strategy for redundant copies? Which partners are involved?
- What is the backup strategy for the LTA (AIP-only or with surrounding (meta-)data -> file structure)?

Subchapter 5.3: File formats and recommendations

- What recommendations or sources of recommendations are there regarding file formats suitable for archiving?

Subchapter 5.4: Metadata

- What strategy do we pursue with regard to metadata: should they be archived as completely and up-to-date as possible in the AIPs or do they primarily serve to correctly identify an AIP and link it to the single source of truth in a library system, for example?
- Which type of identifier do you need for your long-term archive? Should it be linked to the persistent identifier in other systems or of your DIP?

8.1.5 Kapitel 6: Zugänglichkeit und Nutzbarkeit

- Which access levels do we need to support?
 - public on internet, public in the library, authenticated users, special authorization for researchers, inaccessible
- Which legal regulations are relevant for us?
 - Copyright law, protection of personality, archive law ...
- What access platforms already exist that we can use for accessibility? For what does no platform exist yet?
- Is it necessary to create consultation copies (example: derive jpg from tiff)?
- Which type of persistent identifier is possible for you and which do you prefer? Do you need identifiers for a whole DIP (Dissemination information package, Zugangskopie) or also for parts of it (pages, parts of a file, annotations)? Is the PID of the DIP linked to the internal identifier of the AIP (Archival information package)?

- Which level of security is needed for the access of your data? Are the objects that are going to make up your collection legally sensitive? If so in which sense and what are the consequences (localization of the storage)?

8.2 Glossar

AARU	Aaru Data Preservation Suite Werkzeug zur Erhaltung der Inhalte verschiedener Datenträger. https://www.aaru.app
AIP	Archival Information Package Archivinformationspaket, durch das OAIS-Referenzmodell definierter Begriff.
API	Application Programming Interface Die Schnittstelle zur Anwendungsprogrammierung ist eine Schnittstelle, über die eine Software Dienste für andere Software anbietet.
AtoM	Access to Memory Ein webbasiertes Archivverwaltungssystem, das unter anderem vom International Council on Archives (ICA) unterstützt wird. https://www.accesstomemory.org
BagIt	Spezifikation, die eine Methode zur Übertragung oder Sicherung digitaler Inhalte beschreibt. Wird zur Erstellung von SIPs oder AIPs verwendet.
BAK	Bundesamt für Kultur https://www.bak.admin.ch
BAR	Schweizerisches Bundesarchiv https://www.bar.admin.ch
BCU/F	Bibliothèque cantonale et universitaire, Fribourg https://www.fr.ch
CCCA	Compact Disc Digital Audio
DAM	Digital Asset Management Ein System zur Speicherung, Organisation und gemeinsamen Nutzung der digitalen Ressourcen einer Organisation.
Datenträgerimage	Ein Datenträgerimage ist eine exakte Kopie eines optischen oder magnetischen Datenträgers in Form einer Datei.

DIP	Dissemination Information Package Auslieferungsinformationpaket gemäss OAIS-Standard.
DLZA	Digitale Langzeitarchivierung Dauerhafte Erhaltung und Bereitstellung von digitalen Informationen. Die Informationen können durch Digitalisierung einer physischen Quelle oder bereits in digitaler Form («born digital») erworben werden.
DMS	Dokumentenmanagementsystem
DROID	Digital Record and Object Identification Von den National Archives of the United Kingdom entwickeltes Tool zur automatischen Identifizierung von Dateiformaten.
DSGVO	Datenschutz-Grundverordnung Von der Europäischen Union implementierter Referenztext für den Schutz von Personen-Daten.
EAASI	Emulation as a service infrastructure
EAD	Encoded Archival Description XML-Standard für die Kodierung von Archivinventaren, der von der University of California, Berkeley entwickelt wurde und von der Society of American Archivists gemeinsam mit der Library of Congress gepflegt wird.
EBUCore	Ein von der European Broadcasting Union (EBU) entwickeltes Set von beschreibenden und technischen Metadaten für audiovisuelle Medien.
EXIF	Exchangeable Image File Format Spezifikation für die Speicherung von beschreibenden und technischen Metadaten in Bildformaten wie JPEG oder TIFF.
FITS	Flexible Image Transport System Dateiformat für Bilder im naturwissenschaftlichen Umfeld.
HFS+	Hierarchical File System (extended) Von Apple für MacOS entwickeltes Dateisystem.
Image	Siehe Datenträgerimage
IPTC	International Press Telecommunications Council Set von beschreibenden und technischen Metadaten für Bilder.
ISAD(G)	International Standard Archival Description (General) Internationaler Standard zur Verzeichnung archivischer Unterlagen.

JHOVE	JSTOR/Harvard Object Validation Environment Werkzeug zur Identifizierung digitaler Objekte.
KOST/CECO	Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen. Centre de coordination pour l'archivage à long terme de documents électroniques http://kost-ceco.ch
LIDO	Lightweight Information Describing Objects Vom Internationalen Museumsrat (ICOM) entwickeltes, standardisiertes XML-Format für die Beschreibung von Museumsobjekten.
LMS	Library Management System Bibliotheksverwaltungssystem oder auch Bibliothekssystem. Heute meist ein integriertes System mit zentraler Datenhaltung.
LTO	Linear Tape-Open Herstellerunabhängige Datenspeichertechnik auf Magnetband in einem offenen Format.
MARC21	Machine-Readable Cataloging Formatstandard für die bibliografische Beschreibung. Er wurde in den 1960er Jahren entwickelt und wird weltweit von Bibliotheken verwendet.
MARXML	Schema zur Darstellung von MARC21 in XML.
METS	Metadata Encoding and Transmission Standard Standard der die Zusammenfassung sämtlicher für die Beschreibung eines Dokuments erforderlichen Metadaten in einer einzigen XML-Datei erlaubt.
MKV	MatrosKa Video Offenes Containerformat für audiovisuelle Daten (Videos, Bilder, Audios, Untertitel)
MODS	Metadata Object Description Schema Oft in Verbindung mit METS verwendeter XML-Standard für die bibliografische Beschreibung.
MOV	Von Apple entwickeltes Videoformat.
MP4	Moving Picture (Experts Group), Norm 4 Containerformat für audiovisuelle Daten (Videos, Bilder, Audios, Untertitel).

MPEG7	Moving Picture (Experts Group), Norm 7 Beschreibung multimedialer Daten mithilfe von Metainformationen.
MXF	Material eXchange Format Containerformat für Video- und Audiodaten.
NAS	Network-Attached Storage Netzwerkbasierter Dateiserver für die zentrale Speicherung von Daten.
NTFS	New Technology File System Von Microsoft für Windows entwickeltes Dateisystem.
OAIS	Open Archival Information System Reference Model Referenzmodell, das Funktionseinheiten und Funktionen eines DLZA beschreibt (ISO 14721).
ODT	OpenDocument Text Offenes Format für Textdokumente.
PREMIS	PREservation Metadata: Implementation Strategies Metadatenstandard für die Erhaltung digitaler Dokumente.
RDA	Resource Description and Access Weltweit von Bibliotheken verwendetes Regelwerk zur Katalogisierung.
RiC	Records in Context Standard für die Beschreibung von Archivmaterial, herausgegeben vom Internationalen Archivrat (ICA).
rsync	Remote Synchronization Freie Software zum Synchronisieren/Kopieren von Dateien.
SACD	Super Audio CD Optischer Datenträger für Audiodateien.
SEDA	Standard d'échange de données pour l'archivage Standard zum Austausch von Archivdaten.
SFTP	Secure File Transfer Protocol – SSH File Transfer Protocol Datentransferprotokoll für den gesicherten Datentransfer über das Internet.
SIP	Submission Information Package Übergabeinformationspaket gemäss OAIS, das dem DLZA zum Ingest übergeben wird.

SLSP	Swiss Library Service Platform Verbund der wissenschaftlichen und akademischen Bibliotheken der Schweiz.
SPECTRUM	Standard zur Verwaltung von Museumssammlungen.
VDSZ	Verordnung über die Datenschutzzertifizierungen https://www.fedlex.admin.ch/eli/cc/2007/701/de
WAV	Waveform Audio File Format Von IBM und Microsoft entwickeltes Audiodateiformat.
XMP	eXtensible Metadata Platform XML-basiertes Format zur Speicherung von Metadaten in Bilddateien.
ZIP	Proprietärer Datenträger der Firma Iomega, lanciert 1994.