

Grundlagen der digitalen Langzeitarchivierung

Eine Handreichung zur digitalen Langzeitarchivierung aus Perspektive der NFDI4Culture Community

Jörg Heseler

Alexandra Büttner

Matthias Arnold



nationale
Forschungsdaten
Infrastruktur
for CULTURE

Grundlagen der digitalen Langzeitarchivierung

Eine Handreichung zur digitalen Langzeitarchivierung aus Perspektive der NFDI4Culture Community

Jörg Heseler, Alexandra Büttner, Matthias Arnold

Version: 1.0.2, 17. Januar 2024

Abstract (de)

Eine wesentliche Aufgabe von NFDI4Culture ist die Bereitstellung einer Infrastruktur, die es ermöglicht, Forschungsdaten langfristig zu speichern und für andere Wissenschaftler:innen zugänglich zu machen. Neben der Entwicklung von Publikationsmöglichkeiten für komplexe Datentypen spielt vor allem der dauerhafte Erhalt von digitalem Kulturgut eine zentrale Rolle. Diese Handreichung widmet sich dem Thema „digitale Langzeitarchivierung“ und richtet sich an Forschende der NFDI4Culture Community sowie an Infrastruktureinrichtungen. Sie vermittelt Grundlagen zur Langzeitarchivierung und stellt die Angebote von NFDI4Culture vor. Zudem werden die Schritte von Materialien in ein digitales Langzeitarchiv beispielhaft erläutert und das OAIS-Modell vorgestellt. Ein weiterer Abschnitt ist den unterschiedlichen Strategien eines digitalen Langzeitarchivs gewidmet. Der umfangreiche Anhang bietet neben Begriffsklärungen eine ausführliche Literaturliste.

Schlagwörter: [NFDI4Culture](#); [Archivierung](#); [digitale Langzeitarchivierung](#); [Forschungsdaten](#); [RADAR4Culture](#); [SLUBArchiv.digital](#); [heiARCHIVE](#); [TIB Preservation-as-a-Service](#); [Datenübernahme](#); [OAIS-Modell](#); [Erhaltungsstrategie](#); [Bitstream Preservation](#); [Content Preservation](#); [Formatmigration](#); [Emulation](#); [Web-Archivierung](#); [Signifikante Eigenschaften](#)

Abstract (en)

A key task of NFDI4Culture is to provide an infrastructure that enables research data to be stored long-term and made accessible to other scientists. In addition to the development of publication options for complex types of data, the long-term preservation of digital cultural assets plays a central role. This guideline is dedicated to the topic of “digital long-term archiving” and is aimed at researchers in the NFDI4Culture community and infrastructure facilities. It conveys the basics of long-term archiving and presents the offerings of NFDI4Culture. In addition, the steps of moving materials into a digital long-term archive are explained using examples and the OAIS model is presented. Another section is dedicated to the different strategies of a digital long-term archive. In addition to clarifications of terms, the extensive appendix offers a detailed list of references.

Keywords: [NFDI4Culture](#); [preservation](#); [digital preservation](#); [research data](#); [RADAR4Culture](#); [SLUBArchiv.digital](#); [heiARCHIVE](#); [TIB Preservation-as-a-Service](#); [ingest](#); [OAIS model](#); [preservation strategy](#); [bitstream preservation](#); [content preservation](#); [format migration](#); [emulation](#); [web archiving](#); [significant properties](#)

Inhalt

Vorwort	5
1 Einführung	8
2 Grundlagen	11
2.1 Langzeitarchivierung von Forschungsdaten	11
Digitale Langzeitarchivierung	11
Langzeitarchivierung vs. Backup	12
Digitale Langzeitarchive und Repositorien	13
Archivieren vs. Publizieren	14
2.2 Vertrauenswürdigkeit digitaler Langzeitarchivierung	16
FAIRe Forschungsdaten	16
Ethische Prinzipien und CARE	18
TRUST-Prinzipien	19
Zertifizierung	20
2.3 LZA-Communities	23
2.4 Angebote von NFDI4Culture	25
NFDI4Culture-Langzeitarchive	25
SLUBArchiv.digital	26
TIB Preservation-as-a-Service	27
heiARCHIVE	28
RADAR4Culture	29
NFDI4Culture Helpdesk	30
3 Wie kommen Materialien in ein digitales Langzeitarchiv?	32
3.1 Datenmanagementplan	34
3.2 Bestimmung der Archivwürdigkeit	36
3.3 Klärung rechtlicher und ethischer Aspekte	39
Rechtliche Aspekte	39
Ethische Aspekte	41
3.4 Identifizierung erhaltenswerter Eigenschaften	42
Dateiformate für die Langzeitarchivierung	49
Empfohlene Dateiformate für die digitale Langzeitarchivierung	49
Auswahlkriterien für Dateiformate	51
3.5 Erfassen von Metadaten	54
Persistente Identifikatoren	58
3.6 Digitalisierung	60
3.7 Sonderfall Webarchivierung	61

4 Forschungsdaten im digitalen Langzeitarchiv	67
4.1 Das OAIS-Modell	67
Erhaltungsplanung	69
Erstellung von Informationspaketen	70
Intellektuelle Einheiten	70
Erstellung von SIPs	70
Einreichung beim Archiv	71
Formaterkennung	71
Formatvalidierung	72
Extraktion technischer Metadaten	74
Formatkorrektur	74
Speicherung im Archiv	75
Abrufen von Arbeitspaketen	77
4.2 Archivinformationssysteme	78
5 Erhaltungsstrategien	79
5.1 Bitstream Preservation	80
5.2 Content Preservation	82
Logical Preservation	83
Formatmigration	83
Emulation	85
Hardware Preservation	86
Semantic Preservation	89
5.3 Exit-Strategie	90
5.4 Auffindbarkeit langzeitarchivierter Objekte	91
6 Nachwort	94
7 Anhang	95
Abkürzungen	95
Synonyme	98
Glossare	100
Datensatz: Übersicht Wikidata–GND	101
Datensatz: Bibliographie	101
8 Referenzen	102
Impressum	108

Vorwort

Im Bereich des kulturellen Erbes wächst die Nachfrage, Forschungsergebnisse zu veröffentlichen, diese langfristig zu erhalten und in der Zukunft verfügbar zu machen. Dabei wachsen die Herausforderungen für Forschende und Archivanbieter mit zunehmender Komplexität ihrer Daten. Speziell den Bereichen Kunstgeschichte, Architektur, Musikwissenschaft, Theater-, Tanz- sowie Film- und Medienwissenschaft widmet sich NFDI4Culture, das „Konsortium für Forschungsdaten materieller und immaterieller Kulturgüter“ (<https://nfdi4culture.de/>). Es ist ein Teilprojekt der von der [Deutschen Forschungsgemeinschaft \(DFG\)](#)¹ geförderten [Nationalen Forschungsdateninfrastruktur \(NDFI\)](#)². Innerhalb von NFDI4Culture ist das Arbeitsprogramm entlang des Forschungsdatenlebenszyklus strukturiert. Der vierte Aufgabenbereich innerhalb des Konsortiums, die [Task Area 4](#)³, widmet sich der Datenpublikation und Langzeitarchivierung.

Die vorliegende Handreichung bietet zum einen eine allgemeine und verständliche Einführung in die digitale Langzeitarchivierung mit Perspektive auf die NFDI4Culture Community. Dabei bietet sie einen Überblick über die wichtigsten Archivierungsmodelle und veranschaulicht diese anhand von Beispielen und Anwendungsszenarien aus dem Kulturbereich. Sie enthält Hinweise für Forschende zur Archivierung von Forschungsdaten in einem digitalen Langzeitarchiv, gibt weiterführende Informationen und verweist auf mögliche Anlaufstellen. Zum anderen ist diese Handreichung als „Wegweiser“ für Forschende gedacht, die ihre Forschungsdaten in die digitale Langzeitarchivierung überführen wollen.

Im Bereich digitale Langzeitarchivierung sind nach wie vor Unterschiede in Begrifflichkeiten und Standards zu finden. Zwar ist der technische Ablauf oft ähnlich, in Details werden dennoch Unterschiede in Gewichtung der Schwerpunkte, Bezeichnung von Prozessen oder auch Handhabung der Forschungsdaten deutlich. Daher veranstaltete die Task Area 4 der NFDI4Culture mehrere Foren, um sich gemeinsam mit Betreibern von digitalen Langzeitarchiven auf einheitliche Grundbegriffe und Modelle für die digitale Langzeitarchivierung zu verständigen. Diese Handreichung führt nun die verschiedenen Herangehensweisen aus der Perspektive der NFDI4Culture Community zusammen. Dabei wird immer versucht, die Abläufe aus dem Kontext kulturwissenschaftlicher Daten ([Kailus 2023, Abschnitt 2](#)⁴) darzustellen, mit relevanten Beispielen zu erläutern und auch die weiterführenden Informationen aus dem NFDI4Culture-Bereich besonders zu berücksichtigen.

¹ <https://www.dfg.de/>

² <https://www.nfdi.de/>

³ <https://nfdi4culture.de/id/E1824>

⁴ <https://docs.nfdi4culture.de/ta2-fair-handreichung/besonderheiten-kulturwissenschaftlicher-daten>

Nach einer kurzen →[Einführung](#) werden in Kapitel 2 →[Grundlagen der Langzeitarchivierung](#) dargestellt. Es werden wesentliche Begriffe geklärt und →[Langzeitarchive aus der NFDI4Culture](#) Community vorgestellt. Ein Abschnitt widmet sich der Frage, wie die →[Vertrauenswürdigkeit](#) von Langzeitarchiven festgestellt werden kann. In den Kapiteln 3–4 werden wichtige Schritte entlang des Weges von Materialien ins digitale Langzeitarchiv beschrieben. Diese reichen in Kapitel 3 vom →[Datenmanagementplan](#) über →[erhaltenswerte Eigenschaften](#) zur →[Digitalisierung](#) und Beschreibung mit →[Metadaten](#). In Kapitel 4 wird das →[OAIS-Referenzmodell](#) vorgestellt, anhand dessen aus Forschungsdaten →[Informationspakete](#) erstellt und in das Archiv →[eingeliefert](#) werden. Kapitel 5 zeigt →[Erhaltungsstrategien](#) und -maßnahmen auf, die digitale Langzeitarchive in der NFDI4Culture Community gegenwärtig einsetzen. Im Anhang werden die →[Abkürzungen](#) erläutert, die Begriffe mit ihren →[Synonymen](#) auf Deutsch und Englisch zusammengeführt, Links zu externen →[Glossaren](#) bereitgestellt und auf die einschlägige →[Literatur](#) verwiesen.

Insbesondere die **Kapitel 1 bis 3 richten sich an Forschende**, die in ihrer Arbeit selbst Forschungsdaten generieren. Wesentlich für eine Langzeitarchivierung von Forschungsdaten ist der frühzeitige Kontakt mit potentiell interessanten Anbietern von Langzeitarchiven, insbesondere im Zusammenhang mit der Erstellung eines Datenmanagementplans. Die in **Kapitel 4** beschriebenen Arbeitsprozesse sind **für den engen Austausch mit Mitarbeitenden der Infrastruktureinrichtungen** relevant. Die in **Kapitel 5** vorgestellten Erhaltungsstrategien zielen mehr auf die **Abstimmung von Begrifflichkeiten** und Unterscheidung von Prozessen **auf Seiten der Langzeitarchive** ab. Eine weitere, übergreifende Zielgruppe dieser Handreichung sind **Mediator:innen**, die zwischen Forschung und Archiv vermitteln. Dabei geht die Anwendbarkeit der Handreichung über den Kontext der NFDI4Culture Community hinaus.

Im Zusammenhang mit dieser Publikation wurden wichtige Fachbegriffe zusammengestellt und ihre Verfügbarkeit in Normdatenvokabularen geprüft, insbesondere in der [Gemeinsamen Normdatei der DNB \(GND\)](#)⁵ und [Wikidata](#)⁶. Die dabei festgestellten Lücken werden sukzessive gefüllt.

Bitte beachten Sie, dass die in dieser Handreichung erwähnten Informationen dem Stand von Juli 2023 entsprechen. Soweit nicht anders angegeben, erfolgte der letzte Zugriff auf Online-Quellen am 19.09.2023. Um den Langzeitaspekt auch bei den Weblinks zu gewährleisten, wurden die nicht-persistenten Ressourcen in der [Wayback Machine des Internet Archive](#)⁷ auf Verfügbarkeit geprüft und ggf. dort ergänzt. Zur besseren Benutzbarkeit wird ne-

⁵ <https://swb.bsz-bw.de/DB=2.104/>

⁶ <https://www.wikidata.org/>

⁷ <https://web.archive.org/>

ben dem Link auf die Originalressource mit dem Hinweis „Wayback“ zusätzlich auf die archivierte Version verwiesen.

Dieser Text erscheint gemeinsam mit einer Datenpublikation. Diese enthält u. a. eine Tabelle mit den Fachbegriffen (GND-, Wikidata- und [NFDI4Culture IDs](#)⁸, Synonyme deutsch/englisch) und eine Zotero-Bibliographie.

In dieser Guideline werden unterschiedlich gekennzeichnete Zusatzinformationen angeboten:



Das „i“ verweist auf allgemeine Zusatzinformationen.



Das „Buch“ markiert weiterführende Literatur.



Das „Ausrufezeichen“ signalisiert besonders wichtige Hinweise.



Der diesem Icon folgende Internetlink führt auf die archivierte Version der Ressource in der Wayback Machine des Internet Archive.



Die Checkbox führt Arbeitsschritte zusammen, die Forschende bei der Vorbereitung ihrer Daten für die digitale Langzeitarchivierung beachten sollten und selbst – in Zusammenarbeit mit dem Archiv – durchführen können.

Ein herzlicher Dank für die intensive Mitarbeit geht insbesondere an: Gerald Hübsch, Andreas Romeyke, Micky Lindlar, Grischka Petri und Oliver Vettermann. Für die hilfreichen Hinweise und Anregungen möchten sich die Autor:innen herzlich bedanken bei (in alphabetischer Reihenfolge): Thomas Bähr, Kathrin Berghoff, Eva Bodenschatz, Christoph Eggensgluß, Franziska Fritzsche, Reiner Göldner, Sandra Göller, Peggy Große, Sabrina Herzog, Angela Kailus, Desiree Mayer, Leonhard Maylein, Robert Nasarek, Matthias Razum, Kerstin Soltau, Martha Stellmacher, Dagmar Triebel und Annabel Walz, sowie bei den zahlreichen und hilfreichen Hinweisen anonymer Reviewer:innen.

Die Autor:innen heißen jede Art von Feedback, Kritik, Anregungen, Hinweisen oder Anmerkungen zu dieser Handreichung explizit willkommen und möchten alle Leser:innen ermuntern, uns gerne zu kontaktieren. Bitte wenden Sie sich dazu an den [NFDI4Culture Helpdesk](#)⁹ und wählen Sie dort als Thema den Bereich „Datenpublikation und Langzeitarchivierung“ aus.

Wir freuen uns auf Ihre Rückmeldungen!

⁸ <https://www.wikidata.org/wiki/Property:P11971>

⁹ <https://nfdi4culture.de/de/helpdesk.html>

1 Einführung

Die langfristige Bewahrung und zuverlässige Verfügbarkeit von digitalen Kulturgütern ist eine der **Kernaufgaben** von **NFDI4Culture**. Nur durch eine langfristige Sicherung und Bereitstellung von Forschungsdaten können Nachvollziehbarkeit und Qualität der wissenschaftlichen Arbeit gewährleistet und damit der Zugang für zukünftige Forschung ermöglicht werden. Maßgeblich dafür sind die [FAIR](#)¹⁰-, [CARE](#)¹¹- und [TRUST](#)¹²-Prinzipien ([NFDI 2020](#)¹³; [DFG 2016](#)¹⁴; →siehe Abschnitt [Vertrauenswürdigkeit digitaler Langzeitarchive](#)).

Schon in der Antragsphase von Forschungsprojekten müssen konkrete Absprachen im Hinblick auf eine geeignete Infrastruktur für die Datenarchivierung getroffen werden ([DFG 2021](#)¹⁵ [Wayback](#)¹⁶). So verlangt beispielsweise die Deutsche Forschungsgemeinschaft (DFG) von ihren Zuwendungsempfänger:innen in ihren „Empfehlungen für eine gute wissenschaftliche Praxis“:

„Wenn wissenschaftliche Erkenntnisse öffentlich zugänglich gemacht werden, werden die zugrunde liegenden Forschungsdaten (in der Regel Rohdaten) – abhängig vom jeweiligen Fachgebiet – in der Regel für einen Zeitraum von zehn Jahren zugänglich und nachvollziehbar in der Einrichtung, wo sie entstanden sind, oder in standortübergreifenden Repositorien aufbewahrt.“ ([DFG 2022](#)¹⁷, S. 22).

Im Kontext der Nationalen Forschungsdateninfrastruktur (NFDI) wird von „Forschungsdaten“ im weiteren Sinne gesprochen. Bisher existiert keine einheitliche Definition des Begriffs. In den „Leitlinien zum Umgang mit Forschungsdaten“ schlägt die DFG vor:

„Zu Forschungsdaten zählen u. a. Messdaten, Laborwerte, audiovisuelle Informationen, Texte, Surveydaten, Objekte aus Sammlungen oder Proben, die in der wissenschaftlichen Arbeit entstehen, entwickelt oder ausgewertet werden. Methodische Testverfahren, wie Fragebögen, Software und Simulationen können ebenfalls zentrale Ergebnisse wissenschaftlicher Forschung darstellen und sollten daher ebenfalls unter den Begriff Forschungsdaten gefasst werden.“ ([DFG 2015](#)¹⁸, S.1 [Wayback](#)¹⁹)

¹⁰ <https://www.go-fair.org/fair-principles/>

¹¹ <https://www.gida-global.org/care>

¹² <https://www.nature.com/articles/s41597-020-0486-7>

¹³ <https://www.nfdi.de/nfdi4culture/?msclkid=21e215a8b4ba11ecb3d8c4fdcdafd438>

¹⁴ https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/forschungsdaten/leitlinien_forschungsdaten.pdf

¹⁵ https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/forschungsdaten/forschungsdaten_checkliste_de.pdf

¹⁶ https://web.archive.org/web/20220318091939/https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/forschungsdaten/forschungsdaten_checkliste_de.pdf

¹⁷ <https://doi.org/10.5281/zenodo.6472827>

¹⁸ https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/forschungsdaten/leitlinien_forschungsdaten.pdf

¹⁹ https://web.archive.org/web/20220401090503/https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/forschungsdaten/leitlinien_forschungsdaten.pdf

Innerhalb des DARIAH-DE Working Paper zum Forschungsdatenlebenszyklus erarbeitete das Autor:innen-Team diese Definition:

„Unter digitalen geistes- und kulturwissenschaftlichen Forschungsdaten werden innerhalb von DARIAH-DE all jene Quellen/Materialien und Ergebnisse verstanden, die im Kontext einer geistes- und kulturwissenschaftlichen Forschungsfrage gesammelt, erzeugt, beschrieben und/oder ausgewertet werden und in maschinenlesbarer Form zum Zwecke der Archivierung, Zitierbarkeit und zur weiteren Verarbeitung aufbewahrt werden können.“ (Puhl, et al. 2015²⁰, S.14)

Im Kontext von NFDI4Culture wurde eine allgemein gehaltene Definition als [Handreichung](#)²¹ veröffentlicht.



Publikationen und die ihnen zugrunde liegenden Roh- bzw. Primärdaten werden innerhalb von NFDI als Forschungsdaten bezeichnet.



Wichtige Links für den Einstieg ins Thema „Forschungsdaten“:

- forschungsdaten.info. (2023, Juli 12). *Forschungsdaten und Forschungsdatenmanagement*. <https://forschungsdaten.info/>
- NFDI4Culture-Handreichung: Albrecht-Hohmaier, Martin, Bergmann, Katharina, Polywka, Andrea, Eggensgluß, Christoph, & Stark, Alexander. (2022). *Kulturwissenschaftliche Forschungsdaten – eine Einführung*. <https://nfdi4culture.de/go/E3508>

Sollen digitale Informationen wie Forschungsdaten langfristig sicher aufbewahrt werden, spricht man von digitaler Langzeitarchivierung (dLZA). Der wichtigste Standard für die elektronische Archivierung ist das Referenzmodell „Open Archival Information System“ (OAIS), das insbesondere für Archivbetreibende von Interesse ist.

In Deutschland existiert mit [nestor](#)²² ein überregionales Kompetenznetzwerk zur digitalen Langzeitarchivierung, das sich in seinen Publikationen und Arbeitsgruppen dem gesamten Spektrum der dLZA widmet.

²⁰ <http://resolver.sub.uni-goettingen.de/purl/?dariah-2015-4>

²¹ <https://nfdi4culture.de/go/E3508>

²² https://www.langzeitarchivierung.de/Webs/nestor/DE/Home/home_node.html



Wichtige Links für den Einstieg ins Thema „digitale Langzeitarchivierung“:

- forschungsdaten.info. (2023, August 11). *Veröffentlichen und Archivieren*. <https://forschungsdaten.info/themen/veroeffentlichen-und-archivieren/>
- nestor. (2019, November 22). *Für den Einstieg*. Nestor - Für den Einstieg. https://www.langzeitarchivierung.de/Webs/nestor/DE/Services/nestor_Einstieg/nestor_Einstieg_faq_node.html  [Wayback](https://web.archive.org/web/20211120121218/https://www.langzeitarchivierung.de/Webs/nestor/DE/Services/nestor_Einstieg/nestor_Einstieg_faq_node.html)²³
- Neuroth, H., Oßwald, A., Scheffel, R., Strathmann, S., & Huth, K. (2010). *nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*. <http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:0008-2010071949>.
- Deutsches Archäologisches Institut. (2023). *IANUS - Forschungsdatenzentrum Archäologie & Altertumswissenschaften*. <https://ianus-fdz.de/it-empfehlungen/>  [Wayback](https://web.archive.org/web/20230324164200/https://ianus-fdz.de/it-empfehlungen/)²⁴
- Deutsche Nationalbibliothek. (2014, August 31). *Einführung in die LZA - nestor—Deutsche Nationalbibliothek—Wiki*. <https://wiki.dnb.de/pages/viewpage.action?pageId=94681401>  [Wayback](https://web.archive.org/web/20221014163248/https://wiki.dnb.de/pages/viewpage.action?pageId=94681401)²⁵



Weiterführende Angebote von NFDI4Culture:

- [Culture Knowledge Base](https://nfdi4culture.de/services/details/knowledge-base.html)²⁶:
Die NFDI4Culture Knowledge Base enthält Handreichungen, Berichte und Spezifikationen aus den Arbeitsbereichen des Konsortiums zu allen Aspekten des Forschungsdatenmanagements im Bereich des materiellen und immateriellen Kulturerbes.
- [Kuratierte Repositorienliste](https://nfdi4culture.de/de/ressourcen/repository.html)²⁷:
Diese Auswahl bietet einen Überblick über fachspezifische und generische Repositorien für die NFDI4Culture Community, in denen Daten zu materiellen und immateriellen Kulturgütern in verschiedenen Medienformaten nachhaltig publiziert oder auch archiviert werden können.

²³ https://web.archive.org/web/20211120121218/https://www.langzeitarchivierung.de/Webs/nestor/DE/Services/nestor_Einstieg/nestor_Einstieg_faq_node.html

²⁴ <https://web.archive.org/web/20230324164200/https://ianus-fdz.de/it-empfehlungen/>

²⁵ <https://web.archive.org/web/20221014163248/https://wiki.dnb.de/pages/viewpage.action?pageId=94681401>

²⁶ <https://nfdi4culture.de/services/details/knowledge-base.html>

²⁷ <https://nfdi4culture.de/de/ressourcen/repository.html>

2 Grundlagen

2.1 Langzeitarchivierung von Forschungsdaten

Digitale Langzeitarchivierung

In Publikationen ist verschiedentlich von „Langzeitarchivierung“ (z. B. forschungsdaten.info²⁸ oder [nestor](https://www.nestor.de/)²⁹), „Langzeitverfügbarkeit“ (DFG 2016) oder auch „Langzeitsicherung und digitaler Bestandserhaltung“ (DFG 2016) bzw. „Langzeitsicherung und Archivierung digitaler Inhalte“ (DFG 2016) die Rede. In dieser Handreichung werden diese Begriffe synonym verstanden und durchweg der Begriff „digitale Langzeitarchivierung“ (dLZA) verwendet. Andere Termini, wie z. B. „Datenarchivierung“ oder „Backup“, die die Verfügbarkeit nur über einen im Vorfeld festgelegten eher kurzen Zeitraum abdecken, zählen nicht dazu. Ein genauere Vergleich hierzu wird im Abschnitt → [Langzeitarchivierung vs. Backup](#) diskutiert.

In dieser Handreichung ist „Archiv“ oder „archivieren“ immer in Bezug auf die digitale Langzeitarchivierung gemeint. ([Owens, 2014](#)) erläutert eine Reihe von Problemen beim Umgang mit diesem Begriff.

„**Digitale Langzeitarchivierung**“ bedeutet die sichere Aufbewahrung von Daten ab zehn Jahren. Das **digitale Langzeitarchiv** stellt dabei die „Organisation (bestehend aus Personen und technischen Systemen), die die Verantwortung für den Langzeiterhalt und die Langzeitverfügbarkeit von Information in digitaler Form sowie die Bereitstellung für eine bestimmte Zielgruppe übernommen hat“ (DIN 31644, S. 6).



Das Ziel der digitalen Langzeitarchivierung ist es, archivierte Informationen zum einen **unverfälscht und vollständig**, sowie zum anderen **benutzbar** zu erhalten.

Einen festgelegten **Aufbewahrungszeitraum** der digitalen Langzeitarchivierung gibt es aktuell nicht. Die **DFG** verlangt bei der Förderung von Forschungsprojekten oder -vorhaben, Forschungsdaten oder „Primärdaten als Grundlage für Veröffentlichungen“ (DFG 2009, S. 2) in der Institution, in der sie entstanden sind, **mindestens zehn Jahre** aufzubewahren (DFG 2019, S. 24). Die reine Aufbewahrung von Daten ist jedoch keine digitale Langzeitarchivierung.

²⁸ <https://forschungsdaten.info/>

²⁹ <https://www.langzeitarchivierung.de/>

Laut [nestor-Handbuch](#)³⁰ (S. 18) bedeutet digitale Langzeitarchivierung **keine Festlegung** auf fünf oder 50 Jahre, sondern die Übernahme von Verantwortung für einen stetigen Überwachungsprozess, der permanent angepasst werden muss. Einige digitale Langzeitarchive offerieren eine befristete Archivierung über einen kürzeren, festen Zeitraum zwischen fünf und 20 Jahren, während andere sogar eine unbefristete Archivierung über einen längeren, nicht definierten Zeitraum anbieten. Dabei reicht bei vielen existierenden Langzeitarchivierungsangeboten, wie bei [RADAR4Culture](#)³¹ (Haltefrist mindestens 25 Jahre), ein Erhalt der Datenintegrität (→siehe Abschnitt [Bitstream Preservation](#)) aus. Bei Angeboten mit unbefristeten Archivierungszeiträumen, wie etwa dem [SLUBArchiv.digital](#)³², gewinnt hingegen ein Erhalt der Nutzbarkeit (→siehe Abschnitt [Content Preservation](#)) an Geltung.

Langzeitarchivierung vs. Backup

Viele Forschende sichern ihre Projektdaten bereits regelmäßig durch Backups auf Festplatten oder anderen Datenträgern. Das ist auch dringend zu empfehlen, um Datenverlust vorzubeugen und bei Bedarf eine schnelle Wiederherstellung der zumeist noch in Arbeit befindlichen Daten zu gewährleisten. Dennoch handelt es sich bei einem **Backup** lediglich um die Sicherung eines Zustands von Forschungsdaten zu einem bestimmten Zeitpunkt. **Langzeitarchivierung** ist hingegen der Prozess zum dauerhaften Erhalt einer Bitstream-Archivierung (→siehe Kapitel [5 Erhaltungsstrategien](#)). Die für die Aufbewahrung in einem digitalen Langzeitarchiv ausgewählten Forschungsdaten sind nicht für den alltäglichen Gebrauch bestimmt und werden in der Regel vom Langzeitarchiv auf Magnetbändern gespeichert. Wichtig ist in diesem Zusammenhang auf den Gebrauch des Wortes „Archiv“ zu achten. Der Begriff „Archiv“ wird gerne in vielfältigem Kontext genannt und meint oft nicht die Langzeitarchivierung, sondern lediglich eine (temporäre) Speicherung von Daten (Owens 2014).

Unabhängig davon, ob Forschungsdaten später in Langzeitarchiven aufbewahrt werden, sollten Forschende eine [geeignete Backup-Strategie](#)³³  [Wayback](#)³⁴ verfolgen und sämtliche [Daten regelmäßig in Backups sichern](#)³⁵  [Wayback](#)³⁶.

³⁰ <http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:0008-2010071949>

³¹ <https://radar.products.fiz-karlsruhe.de/de/radarabout/radar4culture>

³² <https://slubarchiv.slub-dresden.de/>

³³ <https://www.ahd.de/was-ist-eine-backup-strategie-methoden-und-tipps/>

³⁴ <https://web.archive.org/web/20230606093030/https://www.ahd.de/was-ist-eine-backup-strategie-methoden-und-tipps/>

³⁵ <https://www.forschungsdaten.info/themen/speichern-und-rechnen/datensicherheit-und-backup/>

³⁶ <https://web.archive.org/web/20230616141413/https://forschungsdaten.info/themen/speichern-und-rechnen/datensicherheit-und-backup/>



Weiterführende Informationen zur lokalen Datensicherung:

- forschungsdaten.info. (2023, Mai 5). *Datensicherheit und Backup | Speichern und Rechnen*. <https://forschungsdaten.info/themen/speichern-und-rechnen/datensicherheit-und-backup/>  Wayback³⁷

Zur Verwendung des Begriffs „Archiv“:

- Owens, T. (2014, Februar 27). What Do you Mean by Archive? Genres of Usage for Digital Preservers. *The Signal: Digital Preservation*. <http://blogs.loc.gov/digitalpreservation/2014/02/what-do-you-mean-by-archive-genres-of-usage-for-digital-preservers/>  Wayback³⁸

Digitale Langzeitarchive und Repositorien

Repositorien und Archive sind Speicherorte für digitale Forschungsergebnisse. Im Wesentlichen ist ein **digitales Archiv** ein allgemeinerer Begriff, der sich auf die langfristige Aufbewahrung und Zugänglichkeit digitaler Objekte bezieht. Ein **Repositorium** ist eine Plattform für die Organisation und Bereitstellung von Inhalten, die öffentlich oder für eine bestimmte Nutzer:innengruppe zugänglich ist. Mit Repositorien wird oft eine digitale Publikation der Daten in Verbindung gesetzt, wobei das Repositorium hier zugleich sowohl als Publikationsplattform als auch als digitaler Speicherort fungiert (→siehe auch Abschnitt [Archivieren vs. Publizieren](#)).

Bei vielen digitalen Langzeitarchiven handelt es sich um sogenannte **Dark Archives**. Dies sind Archive, bei denen archivierte Forschungsdaten nicht unmittelbar öffentlich zugänglich sind, sondern in der Regel der Zugriff auf Archivdaten beantragt werden muss. Nutzer:innen können also nicht direkt auf die archivierten Daten zugreifen, sondern sind auf Vermittler bzw. Stellvertreter (z. B. Präsentationssysteme oder Fachabteilungen) angewiesen, die diese Daten auf Anfrage bereitstellen.

Das [SLUBArchiv.digital](#)³⁹, [TIB Preservation-as-a-Service](#)⁴⁰ und [heiARCHIVE](#)⁴¹ sind Beispiele für Dark Archives. Vorteile liegen im sicheren Betrieb, einer vom Publikationssystem unabhängigen Wartung, sowie der Möglichkeit eines einfachen Wechsels zwischen Archiv-Dienstleistern (→siehe Abschnitt [Exit-Strategie](#)).

Gleichzeitig gibt es Repositorien, an die eine digitale Langzeitarchivierung angeschlossen ist. Ein Beispiel hierfür ist [RADAR4Culture](#)⁴². Charakteristisch für Repositorien sind die niedrige

³⁷ <https://web.archive.org/web/20230616141413/https://forschungsdaten.info/themen/speichern-und-rechnen/datensicherheit-und-backup/>

³⁸ <https://web.archive.org/web/20220522104757/https://blogs.loc.gov/thesignal/2014/02/what-do-you-mean-by-archive-genres-of-usage-for-digital-preservers/>

³⁹ <https://slubarchiv.slub-dresden.de/>

⁴⁰ <https://www.tib.eu/de/publizieren-archivieren/digitale-langzeitarchivierung>

⁴¹ <https://heiarchive.uni-heidelberg.de/>

⁴² <https://www.radar-service.eu/>

Zugangsschwelle für Datengebende und -nehmende, die dauerhafte Auffindbarkeit und hohe Sichtbarkeit der archivierten Daten. Zudem machen Repositorien Informationen über gespeicherte Daten verfügbar, so dass z. B. Bibliotheks- oder Forschungskataloge mittels sogenannter Harvester automatisiert diese Informationen extrahieren und indizieren können.

NFDI4Culture stellt eine aktuelle, [kuratierte Repositorienliste vertretener Fachdisziplinen auf nationaler Ebene](#)⁴³ bereit. Eine Zusammenstellung von Forschungsdatenrepositorien beinhaltet das Registry of Research Data Repositories, [re3data](#)⁴⁴. Es dient als Ressource für Forscher:innen, die nach geeigneten Datenarchiven suchen, um ihre Forschungsdaten zu speichern, zu teilen oder zu finden. Auf re3data können [Langzeitarchive und Repositorien aus Kulturwissenschaften auf internationaler Ebene](#)⁴⁵ (19 Datenbanken, Stand: 18.08.2023) eingesehen werden.



Weiterführende Informationen zu re3data:

- Pampel, H., Weisweiler, N. L., Strecker, D., Witt, M., Vierkant, P., Elger, K., Bertelmann, R., Buys, M., Ferguson, L. M., Kindling, M., Kotarski, R., & Petras, V. (2023). Re3data – Indexing the Global Research Data Repository Landscape Since 2012. *Scientific Data*, 10(1), 571. <https://doi.org/10.1038/s41597-023-02462-y>

Archivieren vs. Publizieren

Die digitale Langzeitarchivierung und das digitale Publizieren von Forschungsdaten sind zwei unterschiedliche Prozesse.

Die **Archivierung** zielt auf langfristiges Aufbewahren und zuverlässiges Verfügbarmachen ab. Oft geht mit ihr ein eingeschränkter Zugang einher. Das SLUBArchiv.digital konzentriert sich ausschließlich auf die Archivierung von Forschungsdaten und ist losgelöst vom Publikationsprozess. heiARCHIVE ist ein Archivinformationssystem, das eng mit den Publikationsplattformen der Universität Heidelberg verzahnt ist.

Beim digitalen **Publizieren** steht der unmittelbare Zugriff auf Forschungsdaten im Vordergrund. Im Idealfall haben Publikationen → [persistente Identifikatoren](#), die zusätzlich auf eine längerfristige Aufbewahrung hinweisen. Beim Publizieren werden Forschungsdaten oft schnell abrufbar gemacht, indem man eine komprimierte Speicherung der Daten auf Festplatten akzeptiert. Im Gegensatz dazu wird digitales Archivgut meist verlustfrei, möglichst unkomprimiert und mit zusätzlichen Metadaten gespeichert. Es beansprucht deutlich mehr

⁴³ <https://nfdi4culture.de/de/ressourcen/repositorien.html>

⁴⁴ <https://www.re3data.org/>

⁴⁵ <https://www.re3data.org/search?query=&contentTypes%5B%5D=Archived%20data&subjects%5B%5D=103%20Fine%20Arts%2C%20Music%2C%20Theatre%20and%20Media%20Studies>

Speicherplatz als eine Publikation. In diesem Zusammenhang stellt sich daher die Frage, ob alle publizierten Daten unmittelbar archiviert werden sollten (→siehe Abschnitt [Bestimmung der Archivwürdigkeit](#)).

Einige Repositorien achten bereits bei der Publikation der Forschungsdaten darauf, dass diese in einem →[archivfähigen Format](#) vorliegen. Bei RADAR4Culture werden die Daten für mindestens 25 Jahre archiviert. Andere Repositorien transferieren langzeitarchivfähige Daten in separate Langzeitarchive. Ein Beispiel hierfür ist der Workflow von [heiDATA](#)⁴⁶ bzw. [art-historicum.net@heiDATA](#)⁴⁷ nach heiARCHIVE.

⁴⁶ <https://heidata.uni-heidelberg.de/>

⁴⁷ <https://heidata.uni-heidelberg.de/dataverse/arthistoricum>

2.2 Vertrauenswürdigkeit digitaler Langzeitarchivierung

FAIRe Forschungsdaten

Forschungsdaten sollen laut der [GO FAIR-Initiative](#)⁴⁸ „FAIR“ sein. FAIR steht für „Findable, Accessible, Interoperable und Reusable“ und zielt darauf ab, die Auffindbarkeit, Zugänglichkeit, Interoperabilität und Wiederverwendbarkeit von Forschungsdaten und digitalen Publikationen sicherzustellen. Die Herausforderung liegt in der konsequenten Umsetzung der „Maschinentauglichkeit“ der Daten, denn es gilt, diese auch ohne oder mit nur minimalem menschlichem Eingriff zu finden, darauf zuzugreifen, sie zu interpretieren, zu integrieren und in computergestützten Analysen wiederzuverwenden.

Ziel der Anwendung der FAIR-Prinzipien ist es, Forschungsdaten und digitale Publikationen effizienter und nachhaltiger nutzbar zu machen. Dies fördert die Zusammenarbeit zwischen verschiedenen Forschungsgruppen, ermöglicht die Verifizierung wissenschaftlicher Ergebnisse und erleichtert den Wissensaustausch innerhalb der Forschungsgemeinschaft und darüber hinaus.

FAIRe Forschungsdaten sollten mit eindeutigen, persistenten Identifikatoren (→siehe Abschnitt [Persistente Identifikatoren](#)) versehen und in einer durchsuchbaren Ressource auffindbar sein. Bei einem Dark Archive mit eingeschränktem Benutzerzugang (→siehe Abschnitt [Digitale Langzeitarchive und Repositorien](#)) würde es ausreichen zu wissen, dass sich Forschungsdaten in einem Verzeichnis im Archiv befinden. Dann können Nutzende zumindest Kontakt aufnehmen. Dies gilt auch für sensible Daten, die vielleicht nicht unmittelbar zugänglich sind. Metadaten, die Informationen über die Daten enthalten, sollten in standardisierter Form bereitgestellt werden, um ihre Auffindbarkeit zu verbessern. Durch die Anwendung offener Standards und Protokolle werden die Interoperabilität und Integration von Daten aus verschiedenen Quellen gefördert. Schließlich sollten Forschungsdaten und digitale Publikationen mit möglichst offenen Lizenzen versehen werden.

Für die Archive ist eine regelmäßige Überprüfung und Aktualisierung der Dateiformate notwendig, um sicherzustellen, dass die Daten auch in Zukunft zugänglich bleiben. Dafür müssen geeignete technische Infrastrukturen und Systeme bereitgestellt werden. Metadaten spielen auch hier eine entscheidende Rolle, da sie Informationen über den Kontext und die Interpretation der Daten liefern und somit ihre Wiederverwendbarkeit unterstützen. Darüber hinaus sollten geeignete Sicherheits- und Datenschutzmaßnahmen getroffen werden, um die Integrität und – bei Bedarf – Vertraulichkeit der archivierten Daten zu gewährleisten.

⁴⁸ <https://www.go-fair.org/>

Als Hilfestellung für die Einschätzung der FAIRness eigener Daten bzw. der unterstützenden Leistungen eines Repositoriums bei ihrer Verfügbarmachung wurde das [FAIR Maturity Model](#)⁴⁹ entwickelt.

Es gibt erste Tools zur Bewertung der FAIRness von Forschungsdaten. Darunter **Selbstevaluierungswerkzeuge**, mit dem Benutzer ihre Forschungsdaten eher grundlegend anhand eines kurzen Fragenkataloges bewerten können. Ein bekanntes Beispiel ist das [ARDC FAIR Data Self Assessment Tool](#)⁵⁰, bei dem Benutzer eine Punktzahl anhand ihrer gemachten Multiple-Choice-Angaben erhalten. Ein neueres Beispiel ist der [NFDI4Culture FAIR Check](#)⁵¹, bei dem Nutzer zusätzlich auch differenzierte Angaben über Kommentarfelder geben und anschließend optional in einem individuellen Gespräch mit NFDI4Culture-Mitarbeitern Fragen klären können. Eine andere Art sind **automatisierte Evaluierungswerkzeuge** wie z. B. [F-UJI](#)⁵², bei dem Nutzer nach Eingabe einer URL eine Bewertung in Form einer Punktzahl erhalten. Diese sind jedoch oft auf spezifische Plattformen abgestimmt, wobei es für Forschungsdaten aus kulturspezifischen Domänen daher meist noch schwierig ist, aussagekräftige Bewertungsergebnisse zu erhalten.



Übersicht über die FAIR-Prinzipien:

- GO FAIR Initiative. (2022, Januar 21). *FAIR Principles*. GO FAIR. <https://www.go-fair.org/fair-principles/>

Weiterführende Informationen zu FAIR im Kontext von NFDI4Culture:

- Kailus, Angela. (2023). *Handreichung für ein FAIRes Management kulturwissenschaftlicher Forschungsdaten*. (NFDI4Culture-Handreichung) <https://doi.org/10.5281/ZENODO.7716941>
- Vettermann, O., & Petri, G. (2023). Should I CARE about FAIR? – Ein juristischer Blick auf Prinzipien des Forschungsdatenmanagements. *RuZ - Recht und Zugang*, 4(1), 5–29. <https://doi.org/10.5771/2699-1284-2023-1-5>
- forschungsdaten.info. (2023, Juli 21). *FAIRe Daten | Veröffentlichen und Archivieren*. <https://forschungsdaten.info/themen/veroeffentlichen-und-archivieren/faire-daten/>

Forschungsdaten und Lizenzen:

- FIZ Karlsruhe GmbH. (2023). *Lizenzen für Forschungsdaten | RADAR*. <https://radar.products.fiz-karlsruhe.de/de/radarfeatures/lizenzen-fuer-forschungsdaten>
- forschungsdaten.info. (2023, Mai 12). *Forschungsdaten veröffentlichen | Rechte und Pflichten*. <https://forschungsdaten.info/themen/rechte-und-pflichten/forschungsdaten-veroeffentlichen/>

⁴⁹ <https://zenodo.org/record/5834115>

⁵⁰ <https://ardc.edu.au/resource/fair-data-self-assessment-tool/>

⁵¹ <https://nfdi4culture.de/id/E5080>

⁵² <https://www.f-ujii.net/>

Überblick von FAIR-Tools:

- Lang, K., Assmann, C., Neute, N., Gerlach, R., & Rex, J. (2023). *FAIR Assessment Tools Overview*. <https://doi.org/10.5281/ZENODO.7701941>

Ethische Prinzipien und CARE

Ethische Prinzipien bei der Einlieferung von Materialien in ein Archiv sind von individuellen Archivvorgaben und den Hochschulgesetzen der jeweiligen Bundesländer abhängig, wie von [Vettermann \(2023\)](#)⁵³ erläutert.

Um sicherzustellen, dass Datenrichtlinien historische Kontexte und Machtunterschiede berücksichtigen, wurden die CARE-Prinzipien entwickelt. Diese Prinzipien, die für „**C**ollective **B**enefit, **A**uthority to **C**ontrol, **R**esponsibility and **E**thics“ stehen, betonen die Bedeutung kollektiven Nutzens, des Rechts auf Datenkontrolle, der Verantwortung und Ethik. Sie decken insbesondere den Kontext indigener Gemeinschaften ab und setzen sich für Datensouveränität und die Einbeziehung indigener Werte und Interessen ein.

Obwohl die CARE-Prinzipien besonders auf indigene Kontexte abzielen, geben sie auch darüber hinaus reichenden ethischen Prinzipien Raum. In öffentlichen Archiven werden so Inhalte, die Gewalt verherrlichen, zur Volksverhetzung aufrufen oder die Ehre verletzen, in der Regel nicht aufgenommen.

In sozialgeschichtlichen Archiven wie dem [International Institute of Social History \(IISH\)](#)⁵⁴ kann unter bestimmten Umständen eine Ausnahmeregelung gefunden werden, wenn wissenschaftliches Interesse nachgewiesen wird.



Weitere Informationen zur Klärung ethischer Aspekte und den CARE-Prinzipien:

- GIDA - Global Indigenous Data Alliance. (2018). *Die CARE-Prinzipien für indigene Data Governance*. <https://www.gida-global.org/s/Die-CARE-Prinzipien-fur-indigene-Data-Governance.pdf>  [Wayback](#)⁵⁵
- Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J. D., Anderson, J., & Hudson, M. (2019). *Die CARE-Prinzipien für indigene Data Governance* (deutsche Übersetzung). <https://doi.org/10.5281/zenodo.5995059>

⁵³ <https://doi.org/10.5771/2699-1284-2023-1-5>

⁵⁴ <https://iisg.amsterdam/en>

⁵⁵ <https://web.archive.org/web/20231027025750/https://static1.squarespace.com/static/5d3799de84560400199cd24/t/637acb53881a0973324d18bf/1668991830292/Die+CARE-Prinzipien+f%C3%BCr+indigene+Data+Governance.pdf>

- Reimer Witt. (1997). *Kodex ethischer Grundsätze für Archivarinnen und Archivare*. 3–8. https://www.ica.org/sites/default/files/ICA_1996-09-06_code%20of%20ethics_DE.pdf  Wayback⁵⁶
- Rizzolli, M., & Imeri, S. (2022). CARE Principles for Indigenous Data Governance. *o-bib. Das offene Bibliotheksjournal / Herausgeber VDB*, 1-14 Seiten. <https://doi.org/10.5282/O-BIB/5815>
- Vettermann, O., & Petri, G. (2023). Should I CARE about FAIR? – Ein juristischer Blick auf Prinzipien des Forschungsdatenmanagements. *RuZ - Recht und Zugang*, 4(1), 5–29. <https://doi.org/10.5771/2699-1284-2023-1-5>

TRUST-Prinzipien

Neben den erwähnten FAIR- und CARE-Prinzipien haben die TRUST-Prinzipien eine bedeutende Rolle bei der Gewährleistung einer vertrauenswürdigen und nachhaltigen Langzeitarchivierung digitaler Inhalte gewonnen. Sie gelten für digitale Repositorien sowie digitale Langzeitarchive. TRUST ist das Akronym für „**T**ransparency, **R**esponsibility, **U**ser focus, **S**ustainability und **T**echnology“.

Transparency (Transparenz) ist dabei das grundlegende Prinzip. Es zielt darauf ab, dass Prozesse und Entscheidungen der digitalen Langzeitarchivierung nachvollziehbar sein sollen. Dazu gehört, transparent zu dokumentieren, welche Archivierungsrichtlinien und -verfahren angewendet wurden, um das Vertrauen in die Integrität und Verfügbarkeit der gespeicherten Daten zu stärken.

Responsibility (Verantwortung) bedeutet, dass das digitale Archiv für den Schutz der aufzubewahrenden Inhalte während des gesamten Archivierungszeitraumes zuständig ist.

User Focus (Benutzerorientierung) meint, dass Bedürfnisse der Benutzer im Mittelpunkt stehen sollten, d. h. Archivinhalte sollten einfach zugänglich sein und nutzerfreundlich präsentiert werden.

Sustainability (Nachhaltigkeit) bedeutet, dass das Archiv gespeicherte Inhalte langfristig erhalten und zugänglich machen muss. Dies erfordert neben einer kontinuierlichen Überwachung von Speichermedien auch eine regelmäßige Datenmigration veralteter Formate (→siehe Abschnitt [Erhaltungsplanung](#)) sowie eine Finanzierungsplanung, um Archivdienste langfristig zu gewährleisten.

Technology (Technologie) zielt auf die Auswahl und den Einsatz geeigneter Technologien, um die Sicherheit und die Integrität gespeicherter Daten zu gewährleisten sowie dabei auch eine Anpassung an zukünftige technische Anforderungen zu ermöglichen.

⁵⁶ https://web.archive.org/web/20230914180521/https://www.ica.org/sites/default/files/ICA_1996-09-06_code%20of%20ethics_DE.pdf



Weiterführende Informationen zu TRUST:

- TRUST, N. A.-H. W. (2020). *Nestor endorsement of TRUST Principles*. <https://doi.org/10.5281/ZENODO.4120927>
- Lin, D., Crabtree, J., Dillo, I., Downs, R. R., Edmunds, R., Giaretta, D., De Giusti, M., L'Hours, H., Hugo, W., Jenkyns, R., Khodiyar, V., Martone, M. E., Mokrane, M., Navale, V., Petters, J., Sierman, B., Sokolova, D. V., Stockhause, M., & Westbrook, J. (2020). The TRUST Principles for digital repositories. *Scientific Data*, 7(1), 144. <https://doi.org/10.1038/s41597-020-0486-7>

Zertifizierung

Die Zahl digital archivierter Daten steigt stetig an, ebenso nimmt die Anzahl digitaler Archive zu. Da die Langzeitarchivierung von Forschungsdaten ein sehr komplexer und aufwändiger Prozess ist, wächst auch die Erwartung an die jeweiligen digitalen Archive, ihre Vertrauenswürdigkeit nachzuweisen. Ein Langzeitarchiv, das beispielsweise eine → [Bitstream Preservation](#) anbietet, ist vertrauenswürdig, wenn es das Versprechen halten kann, dass seine ausgelieferten Daten nach einem vereinbarten Zeitraum verfügbar, korrekt und vollständig sind.

Zum Nachweis der Vertrauenswürdigkeit wurden Kriterien und Anforderungskataloge erarbeitet. Digitale Langzeitarchive können sich einer Überprüfung unterziehen und sich somit die Erfüllung dieser Kriterien zertifizieren lassen. Ein solches Zertifikat hilft Datengebenden und -nehmenden, sich Klarheit über die Fähigkeiten, Hintergründe und Arbeitsweisen eines Archivs zu verschaffen. Es dient auch als Kennzeichen der besonderen Qualität eines Archivs.

Zu den bekanntesten Zertifikaten für digitale Langzeitarchive gehören das [nestor-Siegel](#)⁵⁷ und das [CoreTrustSeal](#)⁵⁸. Letzteres hat 2018 das bis dahin bestehende „Data Seal of Approval“ (DSA) abgelöst. Die Kriterienkataloge, die im Rahmen einer Zertifizierung jeweils erfüllt werden müssen, sind publiziert und online verfügbar.

Eine Zertifizierung erfolgt meist durch Selbstevaluierung mit anschließender Begutachtung durch Experten der Zertifizierungsstelle. Sie ist in der Regel kostenpflichtig. Erfüllt ein Archiv die Kriterien für ein Zertifikat, erhält es das Siegel und wird auf der Webseite der Organisation gelistet. Ein transparenter Zertifizierungsprozess sorgt dafür, dass eine Vergleichbarkeit zwischen Archiven möglich und für Außenstehende nachvollziehbar ist.

Die „ISO 16363“-Zertifizierung besitzt den umfangreichsten Kriterienkatalog. Auf dem ISO-Standard 16363 basieren viele Zertifizierungen, darunter auch das nestor-Siegel und das

⁵⁷ https://www.langzeitarchivierung.de/Webs/nestor/DE/Services/nestor_Siegel/nestor_siegel_node.html

⁵⁸ <https://www.coretrustseal.org/>

CoreTrustSeal. Bisher wurde das ISO 16363-Zertifikat weltweit nur einmal vergeben. Aufgrund des aufwändigen Zertifizierungsprozesses hat sich dieses im deutschsprachigen Raum noch nicht etabliert. Die folgende Tabelle stellt die genannten drei Zertifikate für digitale Langzeitarchivierung gegenüber:

	CoreTrustSeal	nestor-Siegel	ISO 16363
Logo			
Organisation	CoreTrustSeal Foundation ⁵⁹	nestor – Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen in Deutschland e. V. ⁶⁰	PTAB – Primary Trustworthy Digital Repository Authorisation Body Ltd ⁶¹
Kosten	3.000 EUR ⁶²	500 EUR ⁶³	18.300 GBP zzgl. Reise- und Aufenthaltskosten für Prüfer
Gültigkeit	3 Jahre	uneingeschränkt	1-3 Jahre (mit Zwischen-Audits)
Kriterien	16 Kriterien auf Basis von ISO 16363 ⁶⁴	34 Kriterien auf Basis von DIN 31644 ⁶⁵ und ISO 16363	alle 125 Kriterien der ISO 16363 ⁶⁶
Zertifizierte Archive	79, Übersicht ⁶⁷	4, Übersicht ⁶⁸	1, Übersicht ⁶⁹

Tabelle 1: Gegenüberstellung der Zertifizierungen von CoreTrustSeal, nestor-Siegel und ISO 16363. (Stand: 07/2023)

⁵⁹ <https://www.coretrustseal.org/>

⁶⁰ https://www.langzeitarchivierung.de/Webs/nestor/DE/Home/home_node.html

⁶¹ <http://www.iso16363.org/>

⁶² <https://www.coretrustseal.org/why-certification/frequently-asked-questions/>

⁶³ https://www.langzeitarchivierung.de/Webs/nestor/DE/Zertifizierung/nestor_Siegel/nestor_siegel_node.html

⁶⁴ <http://www.iso16363.org>

⁶⁵ <https://www.beuth.de/de/-/-/147058907>

⁶⁶ <http://www.iso16363.org>

⁶⁷ <https://amt.coretrustseal.org/certificates/>

⁶⁸ https://www.langzeitarchivierung.de/Webs/nestor/DE/Zertifizierung/nestor_Siegel/nestor_siegel_node.html

⁶⁹ <http://www.iso16363.org/iso-certification/certified-clients/>

Voraussetzung für die Durchführung einer digitalen Langzeitarchivierung sind Personalressourcen. Für das nestor-Siegel und das CoreTrustSeal nimmt eine adäquate qualifizierte Personaldecke einen höheren Stellenwert ein als die infrastrukturelle Ausstattung. Bei Personalabbau muss geprüft werden, ob das Archiv weiterhin die Kriterien eines vertrauenswürdigen Langzeitarchivs erfüllt. Daher prüfen aktuelle Zertifizierungen die Personaldecke und beschränken gleichzeitig die Gültigkeitsdauer der Zertifikate, sodass das Alter des Siegels ersichtlich ist.



Zertifizierte LZA-Angeboten innerhalb von NFDI4Culture:

- [SLUBArchiv.digital](https://slubarchiv.slub-dresden.de/)⁷⁰
- [TIB Preservation-as-a-Service](https://www.tib.eu/de/publizieren-archivieren/digitale-langzeitarchivierung)⁷¹



Vertrauenswürdigkeit digitaler Langzeitarchive:

- Neuroth, H., Oßwald, A., Scheffel, R., Strathmann, S., & Huth, K. (2010). *nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*. <http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:0008-2010071949>
- nestor-Arbeitsgruppe Zertifizierung. (2019). *Erläuterungen zum nestor-Siegel für vertrauenswürdige digitale Langzeitarchive, Version 2.1*. Deutsche Nationalbibliothek. <http://d-nb.info/1189191830/34>
- Schoger, A., Dobratz, S., Altenhöner, R., & nestor-Arbeitsgruppe Vertrauenswürdige Archive - Zertifizierung (Hrsg.). (2008). *Kriterienkatalog vertrauenswürdige digitale Langzeitarchive (Version 2)*. nestor - Kompetenznetzwerk Langzeitarchivierung c/o Deutsche Nationalbibliothek. <http://nbn-resolving.de/urn:nbn:de:0008-2008021802>

Zertifikate und Normen:

- nestor – Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen in Deutschland e. V. (2023, Januar 23). *Nestor-Siegel*. https://www.langzeitarchivierung.de/Webs/nestor/DE/Services/nestor_Siegel/nestor_siegel_node.html  [Wayback](https://www.waybackmachine.org/wayback/https://www.langzeitarchivierung.de/Webs/nestor/DE/Services/nestor_Siegel/nestor_siegel_node.html)⁷².
- Pohlkamp, S., & Zarnitz, M. (2022). Die lange Sicht ...: Nestor – Das Kompetenznetzwerk für Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen. *Bibliotheksdienst*, 56(5), 274–284. <https://doi.org/10.1515/bd-2022-0044>
- CoreTrustSeal. (2023). *CoreTrustSeal*. CoreTrustSeal. <https://www.coretrustseal.org/>
- PTAB – Primary Trustworthy Digital Repository Authorisation Body Ltd. (2023). *ISO16363*. PTAB - Primary Trustworthy Digital Repository Authorisation Body Ltd. <http://www.iso16363.org/>

⁷⁰ <https://slubarchiv.slub-dresden.de/>

⁷¹ <https://www.tib.eu/de/publizieren-archivieren/digitale-langzeitarchivierung>

⁷² https://web.archive.org/web/20230301155634/https://www.langzeitarchivierung.de/Webs/nestor/DE/Services/nestor_Siegel/nestor_siegel_node.html

- Lindlar, M., & Schwab, F. (2022). *203.5 All that work ... for what? Return on investment for trustworthy archive certification processes – a case study*. <https://doi.org/10.17605/OSF.IO/8A3SC>
- Lindlar, M., & Rudnik, P. (2021). *Deutsche Übersetzung der 2019 NDSA Levels of Digital Preservation Matrix*. <https://doi.org/10.5281/zenodo.4718662>

2.3 LZA-Communities

Um bewährte Verfahren und Standards für die dLZA zu entwickeln und zu fördern, ist der Austausch mit Gruppen wichtig. Diese bestehen aus Institutionen, Experten sowie Organisationen und teilen ihr Wissen sowie ihre Ressourcen.

Im deutschsprachigen Raum ist das [Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen in Deutschland e. V. \(nestor\)](#)⁷³ am bekanntesten. nestor bietet auf deren Webseite eine Reihe an Web-Seminaren namens [nestor virtuell](#)⁷⁴  [Wayback](#)⁷⁵ sowie viele publizierte Ressourcen. Die bekannteste Ressource ist dabei das [nestor-Handbuch](#)⁷⁶. Interessant sind auch die [nestor Community Profiles](#)⁷⁷  [Wayback](#)⁷⁸, die einen Überblick über Gruppen in der dLZA geben. Schließlich organisiert nestor auch Arbeitsgruppen zu diversen LZA-Themen sowie eine jährliche Veranstaltung zur LZA-Praxis – den [nestor-Praktikertag](#)⁷⁹  [Wayback](#).

Die [Open Preservation Foundation \(OPF\)](#)⁸⁰ ist eine international ausgelegte LZA-Community. Sie ist verantwortlich für die Entwicklung renommierter LZA-Tools wie etwa von [JHOVE](#)⁸¹, [veraPDF](#)⁸² und [fido](#)⁸³, veranstaltet [Web-Seminare](#)⁸⁴ (registrierungspflichtig) und betreibt ebenso einen sehr aktiven [Blog](#)⁸⁵.

⁷³ https://www.langzeitarchivierung.de/Webs/nestor/DE/Home/home_node.html

⁷⁴ https://www.langzeitarchivierung.de/Webs/nestor/DE/Services/nestor_virtuell/nestor_virtuell_node.html

⁷⁵ https://web.archive.org/web/20230607110902/https://www.langzeitarchivierung.de/Webs/nestor/DE/Services/nestor_virtuell/nestor_virtuell_node.html

⁷⁶ <http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:0008-2010071949>

⁷⁷ https://www.langzeitarchivierung.de/Webs/nestor/EN/Services/nestor_Community_Profiles/nestor_Community_Profiles_node.html

⁷⁸ https://web.archive.org/web/20230531025139/https://www.langzeitarchivierung.de/Webs/nestor/EN/Services/nestor_Community_Profiles/nestor_Community_Profiles_node.html

⁷⁹ https://www.langzeitarchivierung.de/Webs/nestor/DE/Services/nestor_Praktikertag/praktikertag_node.html

⁸⁰ <https://openpreservation.org/>

⁸¹ <http://jhove.openpreservation.org/>

⁸² <https://openpreservation.org/tools/veraPDF>

⁸³ <https://openpreservation.org/tools/fido>

⁸⁴ <https://openpreservation.org/webinars/>

⁸⁵ <https://openpreservation.org/blogs/>

Die [Digital Preservation Coalition \(dpc\)](#)⁸⁶ fokussiert sich zwar auf Großbritannien, besitzt jedoch auch ein gutes Handbuch zum Thema dLZA, das [Digital Preservation Handbook](#)⁸⁷  [Wayback](#)⁸⁸, sowie zahlreiche [Technology Watch Publications](#)⁸⁹  [Wayback](#)⁹⁰, die sich u. a. thematisch auf bestimmte Formatgruppen fokussieren. Ebenso bietet die dpc auch [Webseminare](#)⁹¹  [Wayback](#)⁹² und einen [Blog](#)⁹³  [Wayback](#)⁹⁴ an. Für Anfänger eignet sich hier auch das Online-Training namens [Novice to Know-How](#)⁹⁵  [Wayback](#)⁹⁶.

Darüber hinaus gibt es noch Konferenzen. Die bekannteste ist dabei die [International Conference on Digital Preservation \(iPRES\)](#)⁹⁷, die einmal jährlich stattfindet.

Schließlich kann man sich auch über Mailing-Listen wie der [Digital-Preservation](#)⁹⁸  [Wayback](#)⁹⁹ eintragen, um sich auf den neuesten Stand zu halten.

⁸⁶ <https://www.dpconline.org/>

⁸⁷ <https://www.dpconline.org/handbook/contents>

⁸⁸ <https://web.archive.org/web/20231123175017/https://www.dpconline.org/handbook/contents>

⁸⁹ <https://www.dpconline.org/digipres/discover-good-practice/tech-watch-reports>

⁹⁰ <https://web.archive.org/web/20231031115349/https://www.dpconline.org/digipres/discover-good-practice/tech-watch-reports>

⁹¹ <https://www.dpconline.org/events/all-webinars>

⁹² <https://web.archive.org/web/20230519194301/https://www.dpconline.org/events/all-webinars>

⁹³ <https://www.dpconline.org/blog>

⁹⁴ <https://web.archive.org/web/20231203121904/https://www.dpconline.org/blog>

⁹⁵ <https://www.dpconline.org/digipres/collaborative-projects/n2kh-project>

⁹⁶ <https://web.archive.org/web/20231205142230/https://www.dpconline.org/digipres/collaborative-projects/n2kh-project>

⁹⁷ <https://ipres-conference.org/>

⁹⁸ <https://www.jiscmail.ac.uk/cgi-bin/webadmin?A0=digital-preservation>

⁹⁹ <https://web.archive.org/web/20230604185415/https://www.jiscmail.ac.uk/cgi-bin/webadmin?A0=digital-preservation>

2.4 Angebote von NFDI4Culture

NFDI4Culture-Langzeitarchive

Im Folgenden werden Archive der [NFDI4Culture Co-Applicants](#)¹⁰⁰ vorgestellt, die Langzeitarchivierung ermöglichen.

Grundsätzlich sollten Forschende zunächst immer nach fachspezifischen Archiven schauen. Im Kontext der NFDI4Culture gehören dazu SLUBArchiv.digital, TIB Preservation-as-a-Service und heiARCHIVE. Wenn fachspezifische Archive nicht infrage kommen, kann auch auf ein generisches Archiv wie RADAR4Culture zurückgegriffen werden.

Auf alle vier genannten Archive wird in dieser Handreichung häufig Bezug genommen. Die hier folgenden Steckbriefe bieten sowohl einen Überblick über Anbieter und Zielgruppen als auch die angewandten Erhaltungsstrategien sowie eventuelle Kosten und Einschränkungen.

¹⁰⁰ <https://nfdi4culture.de/about-us/partners.html#c310>

	SLUBArchiv.digital
URL	https://slubarchiv.slub-dresden.de/
NFDI4Culture ID	https://nfdi4culture.de/id/E3090
Anbieter	SLUB Dresden
Produktiv seit	2015
Zweck	Langzeitarchivierung
Schwerpunkt	Architekturgeschichte, Kunstgeschichte und Musikwissenschaft; Voraussetzung ist eine individuelle fachliche Bewertung durch die SLUB in Bezug auf die Passfähigkeit zu deren Sammlungsschwerpunkten
Zielgruppe	Öffentliche Kultur- und Wissenschaftseinrichtungen
Kosten	kostenpflichtig, volumenbasiert (Preise auf Anfrage ¹⁰¹)
Haltefrist	unbegrenzt
Dateiformate	nur nach Archivvorgaben ¹⁰²
Erhaltungsstrategien	Content Preservation (Formatmigration)
Zugriff	geschlossen (Dark Archive)
Zertifizierung	Data Seal of Approval (2014); CoreTrustSeal (2022)
Weiterführende Literatur	SLUB Dresden. (2023). <i>Technische Standards für die Ablieferung von digitalen Dokumenten</i> . https://slubarchiv.slub-dresden.de/technische-standards-fuer-die-ablieferung-von-digitalen-dokumenten

¹⁰¹ [Gerald Huebsch \[at\] slub-dresden.de](mailto:Gerald.Huebsch@slub-dresden.de)

¹⁰² <https://slubarchiv.slub-dresden.de/technische-standards-fuer-die-ablieferung-von-digitalen-dokumenten>

TIB Preservation-as-a-Service

	
URL	https://www.tib.eu/de/publizieren-archivieren/digitale-langzeitarchivierung
NFDI4Culture ID	https://nfdi4culture.de/id/E2939
Anbieter	TIB Hannover
Produktiv seit	2013
Zweck	Langzeitarchivierung
Schwerpunkt	Dokumente, Retro-Digitalisate, A/V-Daten, 3D-Daten und Architektur
Zielgruppe	Wissenschaftler:innen, Lehrende, Studierende an Universitäten, Hochschulen und außeruniversitären Forschungseinrichtungen, Mitarbeitende in der außerakademischen und kommerziellen Forschung, Bibliotheken, Archive, Museen und Infrastruktureinrichtungen, Projektträger wie die Deutsche Forschungsgemeinschaft (DFG) oder das Bundesministerium für Bildung und Forschung (BMBF)
Kosten	kostenpflichtig (Preise auf Anfrage ¹⁰³)
Haltefrist	unbegrenzt
Dateiformate	keine Beschränkung
Erhaltungsstrategien	Content Preservation (Formatmigration) und Bitstream Preservation
Zugriff	geschlossen (Dark Archive)
Zertifizierung	Data Seal of Approval (2015); CoreTrustSeal (2020); nestor-Siegel (2017, 2022)
Weiterführende Literatur	TIB Hannover. (2023, Februar 24). <i>Digitale Langzeitarchivierung an der TIB</i> . TIB Wiki. https://wiki.tib.eu/confluence/display/lza/Digitale+Langzeitarchivierung+an+der+TIB

¹⁰³ [Thomas Baehr \[at\] tib.eu](mailto:Thomas.Baehr@tib.eu)

	
URL	https://heiarchive.uni-heidelberg.de/
NFDI4Culture ID	https://nfdi4culture.de/id/E3788
Anbieter	Universität Heidelberg, Kompetenzzentrum Forschungsdaten (KFD) ¹⁰⁴ , eine gemeinsame Serviceeinrichtung des Universitätsrechenzentrums ¹⁰⁵ und der Universitätsbibliothek Heidelberg ¹⁰⁶
Produktiv seit	28.04.2023 (Pilotbetrieb)
Zweck	Langzeitarchivierung
Schwerpunkt	fachunabhängiges Archiv für Forschungsdaten von Mitgliedern der Universität Heidelberg; Kunstgeschichte
Zielgruppe	Mitglieder der Universität Heidelberg; Kunsthistoriker:innen weltweit in Verbindung mit einer Open-Access-Publikation in arthistoricum.net
Kosten	Langzeitarchivierung in Verbindung mit Online-Publikationen auf arthistoricum.net im Rahmen des Fachinformationsdienst Kunst kostenfrei, ggf. anfallende Publikationsgebühren
Haltefrist	Archivierung von Publikationen und kulturellem Erbe: unbegrenzt, sonst: in Stufen, mind. 10 Jahre
Dateiformate	Bitstream Preservation: keine Beschränkung, Content Preservation: TXT, PDF, XML und TIFF (Stand: 08/2023)
Erhaltungsstrategien	Bitstream Preservation und Content Preservation (Formatmigration)
Zugriff	geschlossen (Dark Archive)
Zertifizierung	in Planung
Weiterführende Literatur	Universitätsbibliothek Heidelberg. (2023). <i>heiARCHIVE - der zentrale Datenarchivierungsdienst der Universität Heidelberg</i> . https://heiarchive.uni-heidelberg.de/

¹⁰⁴ <https://data.uni-heidelberg.de>

¹⁰⁵ <https://www.urz.uni-heidelberg.de/>

¹⁰⁶ <https://www.ub.uni-heidelberg.de/>

	
URL	https://radar4culture.radar-service.eu/
NFDI4Culture ID	https://nfdi4culture.de/id/E2853
Anbieter	FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur
Produktiv seit	2022
Zweck	Publizieren mit Langzeitarchivierung
Schwerpunkt	generisch (NFDI4Culture-Disziplinen)
Zielgruppe	Forschende an öffentlich geförderten Forschungseinrichtungen und (Kunst-)Hochschulen sowie nicht-kommerziellen Akademien, Galerien, Bibliotheken, Archiven und Museen in Deutschland
Kosten	kostenfreie Publikationen bis 10 GB (Preise für größere Volumina ¹⁰⁷)
Haltefrist	unbegrenzt (mind. 25 Jahre)
Dateiformate	keine Beschränkung
Erhaltungsstrategien	dreifach redundante Bitstream Preservation mit kontextuellen Metadaten
Zugriff	öffentlich (Datensätze sind jeweils per DOI identifizierbar)
Zertifizierung	CoreTrustSeal-Zertifizierungen sind bisher für generische Repositorien ohne fachliche Kuratierung noch nicht möglich
Weiterführende Literatur	Soltau, K., & Goeller, S. (2023). <i>RADAR4Culture: Quickstart-Guide für Datengeberinnen und Datengeber</i> . https://doi.org/10.5281/ZENODO.8221341 Soltau, K. (2023). <i>RADAR4Culture: Handreichung zu personenbezogenen Daten</i> . https://doi.org/10.5281/ZENODO.8221496

¹⁰⁷ <https://radar.products.fiz-karlsruhe.de/de/radaragreementsprices/vertraege-preise>

NFDI4Culture Helpdesk

URL	https://nfdi4culture.de/services/details/nfdi4culture-helpdesk.html
NFDI4Culture ID	https://nfdi4culture.de/id/E2409

Der [NFDI4Culture Helpdesk](https://nfdi4culture.de/services/details/nfdi4culture-helpdesk.html)¹⁰⁸ bietet umfassende Beratungsdienste für Forschende und Institutionen aus dem Kulturbereich. Er unterstützt dabei in verschiedenen Aspekten der Datenverarbeitung und -verwaltung. Dazu gehören:

- Beratung zur Speicherung, Veröffentlichung und **Langzeitarchivierung** von Forschungsergebnissen im Kulturbereich.
- Hilfe bei der **Anwendung von Qualitätsstandards** wie den FAIR-Kriterien für Forschungsdaten.
- Unterstützung bei Fragen zur **Datenqualität und -standards** in allen Phasen des Datenlebenszyklus.
- Beratung in der **Planungsphase von Projekten** und bei der Umsetzung von qualitätsorientierten Datenmanagement Strategien.
- Hilfe bei der Erstellung von **Datenmanagementplänen** und der Auswahl von **Metadatenstandards**.
- Unterstützung und Informationen zu **datenrechtlichen Fragen** wie Urheberrecht und Datenschutz.
- Klärung rechtlicher Fragen, um Forschungsdaten konform mit den FAIR-Prinzipien zu halten.
- Unterstützung und Informationen zu **ethischen Fragen**, insbesondere im Umgang mit kulturell sensiblen Objekten.
- Sicherstellung, dass Forschungsdaten **ethischen Richtlinien** entsprechen.
- Unterstützung bei der Identifizierung geeigneter Verfahren im **Digitalisierungsprozess** unter Berücksichtigung von Anforderungen wie Interoperabilität und Wiederverwendbarkeit.

¹⁰⁸ <https://nfdi4culture.de/id/E2409>



Identifizieren Sie ein geeignetes Langzeitarchiv:

- Kennen Sie die **für Ihren Fachbereich geeigneten Langzeitarchive**?
Machen Sie sich mit deren Anforderungen an Forschungsdaten vertraut.
- Sind ihre **Forschungsdaten in einem Repository** veröffentlicht? Planen Sie das zu tun? Haben Sie bereits Kontakt zu diesem Repository aufgenommen?
- Ist das von Ihnen gewählte **Repository an ein Langzeitarchiv angeschlossen**?
- Nehmen Sie **Kontakt zum Archiv** auf und besprechen Sie Ihre Daten und Ihre Anforderungen.

3 Wie kommen Materialien in ein digitales Langzeitarchiv?

In diesem Abschnitt wird der Weg von Forschungsdaten in ein digitales Langzeitarchiv modellhaft beschrieben (siehe folgende Abbildung). Je nach Ausgangslage variieren die Schritte in ein digitales Langzeitarchiv. So fällt zum Beispiel bei einem born-digital Objekt der Digitalisierungsprozess komplett weg. Im folgenden Abschnitt wird daher von noch nicht digitalisierten analogen Kulturgütern ausgegangen, wie etwa historischen Handschriften, Tonbändern, Printmedien oder Filmen als Ausgangsmaterial. Anhand der schematischen Darstellung werden die wesentlichen Schritte, die Materialien auf dem Weg in ein digitales Langzeitarchiv durchlaufen, skizziert.

Der Weg beginnt bei den Datengebenden: Basierend auf einem Grundverständnis für die digitale Langzeitarchivierung gilt es zunächst, die →[Archivwürdigkeit](#) des einzuliefernden Materials zu bewerten und in diesem Zusammenhang →[rechtliche Aspekte](#) sowie →[ethische Aspekte](#) zu klären. Falls die Nutzbarkeit der gespeicherten Informationen erhalten werden soll, müssen dann die →[erhaltenswerten Eigenschaften](#) identifiziert werden. Danach folgen die →[Prüfung und Auswahl archivfähiger Formate](#). Unabhängig von der Erhaltungsstrategie verläuft die →[Digitalisierung](#), insofern diese noch nicht durchgeführt wurde. Abschließend werden Dateien des einzuliefernden Materials mit →[erfassten Metadaten](#) zu →[Informationspaketen gepackt](#) und ins →[digitale Langzeitarchiv eingeliefert](#).

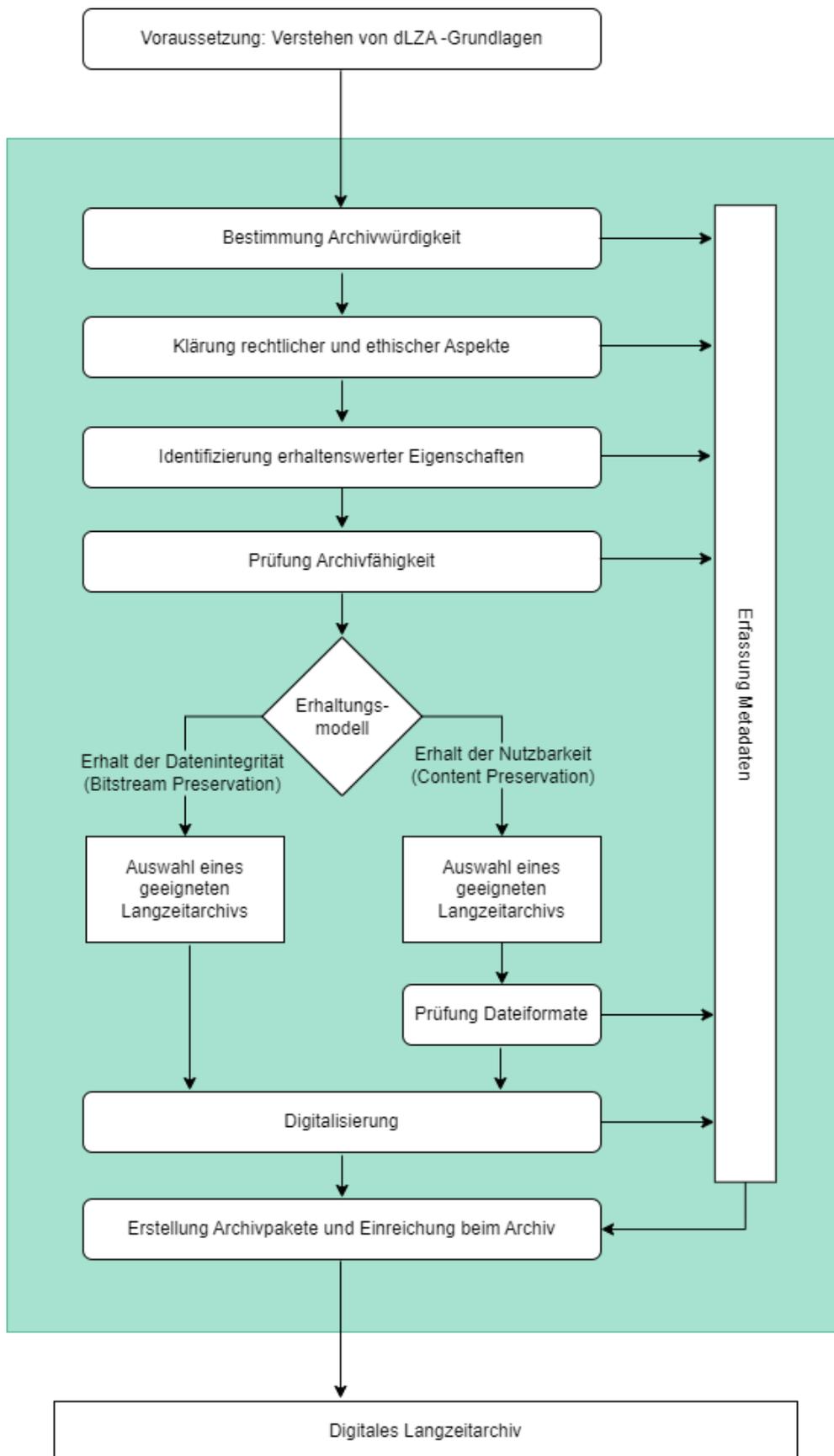


Abbildung 1: Schematische Darstellung des Wegs von Forschungsdaten in ein digitales Langzeitarchiv.

Bei dieser Abbildung handelt es sich um eine vereinfachte schematische Darstellung, die von Archiv zu Archiv immer etwas anders aussehen wird. Auch ist die digitale Langzeitarchivierung ein Prozess, der dynamischen Veränderungen unterliegt. Archive müssen kontinuierlich und flexibel auf Veränderungen im Kontext der Community (z. B. sich ändernde Dateiformate von Forschungsdaten) und auf äußere technische Veränderungen (z. B. neue Softwareentwicklungen) reagieren.

Es gibt daher nicht den EINEN Weg von Daten in ein Langzeitarchiv. Alles ist abhängig von den spezifischen Forschungsdaten und den Bestimmungen des jeweiligen Archivs.



Weitere Modelle und Wege in ein digitales Langzeitarchiv:

- nestor-Arbeitsgruppe Standards für Metadaten, Transfer von Objekten in digitale Langzeitarchive und Objektzugriff (Hrsg.). (2009). *Wege ins Archiv: Ein Leitfaden für die Informationsübernahme in das digitale Langzeitarchiv*. (Version 1-zur öffentlichen Kommentierung). nestor - Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit Digitaler Ressourcen für Deutschland. <https://doi.org/10.18452/1527>

3.1 Datenmanagementplan

Förderprogramme für Forschungsprojekte setzen im Regelfall einen [Datenmanagementplan \(DMP\)](#)¹⁰⁹ voraus, der zusammen mit dem Antrag auf Fördermittel eingereicht wird. Ein DMP deckt alle Bereiche des Forschungsdatenlebenszyklus (Datenerhebung, -verarbeitung, -analyse, -speicherung und -freigabe) ab und beinhaltet auch die Planung einer Strategie für die Langzeitarchivierung. Um die entsprechenden Kosten für die Langzeitarchivierung zu ermitteln, muss jedoch zum einen ein Überblick über das archivwürdige und -fähige Datenvolumen bestehen und zum anderen bereits vor Antragsstellung mit einem geeigneten Langzeitarchiv Kontakt aufgenommen werden. Deshalb wird ausdrücklich empfohlen, bereits in der Planungsphase bzw. während der Antragsstellung festzulegen, welche Projektergebnisse und Daten dauerhaft archiviert und bewahrt werden sollen. Genau diese Entscheidungen müssen für die digitale Langzeitarchivierung ebenfalls getroffen werden. Man spricht in diesem Kontext von der →[Bestimmung der Archivwürdigkeit](#) bzw. von →[erhaltenswerten Eigenschaften](#) der Forschungsdaten.

¹⁰⁹ <https://forschungsdaten.info/themen/informieren-und-planen/datenmanagementplan/>



Weiterführende Informationen zu Datenmanagementplänen:

- forschungsdaten.info. (2023, Mai 5). *Datenmanagementplan | Informieren und Planen*. <https://forschungsdaten.info/themen/informieren-und-planen/datenmanagementplan/>
- NFDI4Culture-Handreichung: Albrecht-Hohmaier, Martin, Bergmann, Katharina, Polywka, Andrea, Eggensgluß, Christoph, & Stark, Alexander. (2022). *Kulturwissenschaftliche Forschungsdaten – eine Einführung* [dataset]. Zenodo. <https://doi.org/10.5281/ZENODO.7327498> (siehe insbesondere Kapitel 3)
- NFDI4Culture-Handreichung: Kailus, A. (2023). *Handreichung für ein FAIRes Management kulturwissenschaftlicher Forschungsdaten*. <https://doi.org/10.5281/zenodo.7716941>
- forschungsdaten.info. (2023, Mai 5). *Förderrichtlinien | Informieren und Planen*. <https://forschungsdaten.info/themen/informieren-und-planen/foerderrichtlinien/>



Verschaffen Sie sich einen Überblick über Ihre Forschungsdaten:

- Welche **Medientypen** benutzen Sie? Klassifizieren Sie Ihre Daten nach Texten, Bildern, Video, Audio, Annotationen, 3D-Daten, Tabellen, etc.
- Welche **Datenmenge** umfassen Ihre Daten? Klassifizieren Sie nach Dimensionen wie < 10 GB, < 100 GB, < 1 TB, < 5 TB und ≥ 5 TB. Teilen Sie Ihre Daten in Größenklassen ein.
- Sind die **Daten unabhängig** oder Teil eines Datenbanksystems? Können die Daten extrahiert werden und in welcher Form (z. B. Export als Rohdaten in Tabellenform, verlustfreie und offene Formate)?
- Sind Ihre **Daten strukturiert**? Gibt es für diese Datenart in Ihrem Fachbereich Datenstandards und benutzen Sie diese (z. B. TEI, MEI)?
- Dokumentieren Sie alles**, idealerweise in Ihrem Datenmanagementplan.

3.2 Bestimmung der Archivwürdigkeit

Die langfristige Archivierung digitaler Daten ist mit erheblichen Kosten verbunden, weshalb grundsätzlich nicht alle Informationen archiviert werden können. Die Entscheidung darüber, welche Daten aufbewahrt werden sollen, bezeichnet man als Bestimmung der Archivwürdigkeit. Diese erfolgt sowohl vor Beginn der digitalen Langzeitarchivierung als auch nach Ablauf der Aufbewahrungsfrist.

Die Bestimmung der Archivwürdigkeit stellt einen entscheidenden Schritt in der digitalen Langzeitarchivierung dar und liegt in der Regel in der Verantwortung von Archivaren. Dabei erfolgt die Beurteilung anhand nachvollziehbarer Kriterien, die von Archiv zu Archiv variieren können. Archivwürdigkeit wird sowohl durch Gesetze als auch durch vertragliche Vereinbarungen geregelt, wie beispielsweise die Pflichtablieferung von Medienwerken an Landesbibliotheken in Deutschland.

Generell gilt der Grundsatz, dass **Materialien archivwürdig sind, wenn sie einen zeitlosen, bleibenden Wert für die Gesellschaft besitzen**. Dieser „bleibende Wert“ wird durch Merkmale wie historischen Wert, Einzigartigkeit, Authentizität (d. h. die Unverändertheit von Informationen), Forschungspotenzial und Relevanz für die Institution bestimmt (Kommission Archäologie und Informationssysteme im Verband der Landesarchäologen Deutschlands, 2017). Die Bewertung berücksichtigt auch die wissenschaftliche Qualität gemäß [DFG-Empfehlungen zur Sicherung guter wissenschaftlicher Praxis](#)¹¹⁰ (d. h. Originalität, Innovationshöhe und Beitrag zum Erkenntnisfortschritt, Einzigartigkeit, wissenschaftliche Relevanz sowie Verständlichkeit) sowie ökologische Aspekte der Langzeitarchivierung (z. B. der Energie- und Ressourcenverbrauch durch Austausch obsoleter Speichermedien, den Rechenzentren bei der Speicherung und Verarbeitung archivierter Daten benötigen).

Ausschlusskriterien für die Archivierung sind in der Regel nicht abgeschlossene Projektdaten, Duplikate, unbeschriebene Daten und hochspezialisierte Formate mit technisch eng begrenzter Nutzbarkeit. (Kommission Archäologie und Informationssysteme im Verband der Landesarchäologen Deutschlands, 2017)

Die Herausforderung besteht darin, eine eindeutige Bewertung der Archivwürdigkeit vorzunehmen, da sogar scheinbar irrelevante Daten möglicherweise erst nach vielen Jahren ihre Bedeutung zeigen können, wie folgendes Beispiel zeigt:

¹¹⁰ <https://doi.org/10.5281/zenodo.6472827>

Fallbeispiel: Archivwürdigkeit alter Seefahrtslogbücher

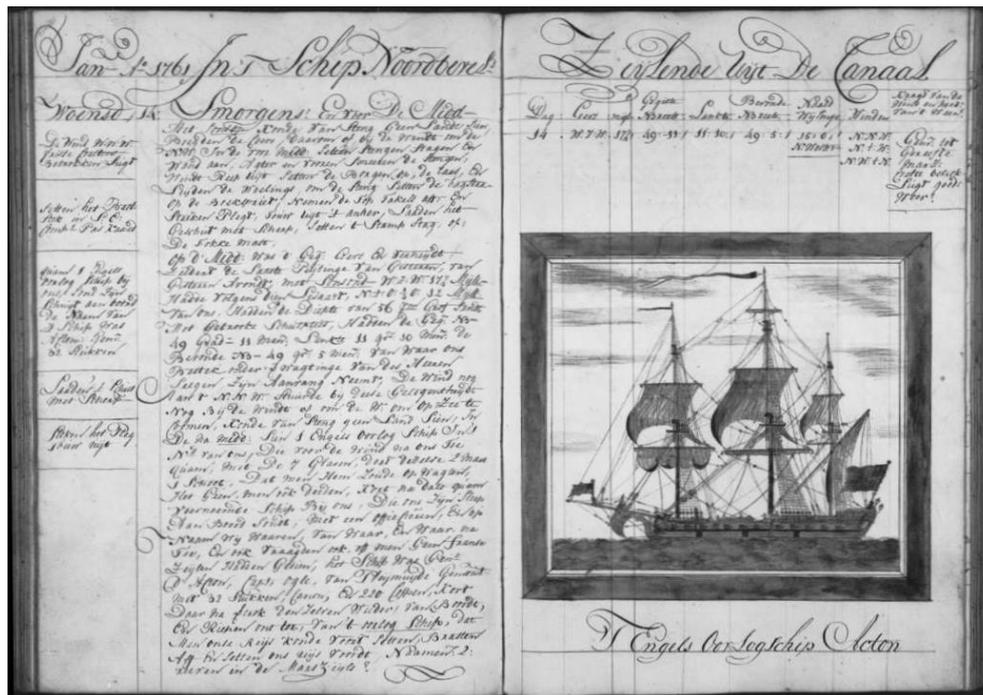


Abbildung 2: Schiffstagebuch eines niederländischen Schiffes aus dem Jahre 1761.
(Foto: © [Royal Netherlands Meteorological Institute \(KNMI\)](https://www.knmi.nl)¹¹¹)

Im [Reclaim-Projekt](https://reclaim.noaa.gov/)¹¹² [Wayback](https://web.archive.org/web/20230718124021/https://icoads.noaa.gov/reclaim/)¹¹³ der U.S. National Oceanic and Atmospheric Administration werden Logbücher von Schiffen digitalisiert (ähnlich wie in der Abbildung oben). Solche Logbücher enthalten u. a. auch diverse Wetterbeobachtungen. Die Sammlung und Digitalisierung historischer Logbücher ist heute interessant für die Klimaforschung. Aus ihnen können viele historische Daten entnommen werden, mit denen man Klimamodelle nachrechnen, validieren und genauer in die Zukunft extrapolieren kann. Logbücher müssen auch heute noch geführt werden. Sie enthalten eine Dokumentation der täglichen Ereignisse während einer Reise und können z. B. bei Rechtsstreitigkeiten wichtige Beweismittel sein. Heutzutage gilt, dass Seetagebücher nach der letzten Eintragung mindestens drei Jahre lang aufbewahrt¹¹⁴ [Wayback](https://web.archive.org/web/20220701070938/https://www.segeln360.de/yachtcharter/unterwegs-waehrend-des-toerns/logbuch.html)¹¹⁵ werden müssen.

Referenz:

Wilkinson, C., Woodruff, S. D., Brohan, P., Claesson, S., Freeman, E., Koek, F., Lubker, S. J., Marzin, C., & Wheeler, D. (2011). Recovery of logbooks and international marine data: The RECLAIM project. *International Journal of Climatology*, 31(7), 968–979.
<https://doi.org/10.1002/joc.2102>

¹¹¹ <https://cdn.knmi.nl/system/updates/image1s/000/002/103/xlarge/101170.jpeg>

¹¹² <https://icoads.noaa.gov/reclaim/>

¹¹³ <https://web.archive.org/web/20230718124021/https://icoads.noaa.gov/reclaim/>

¹¹⁴ <https://www.segeln360.de/yachtcharter/unterwegs-waehrend-des-toerns/logbuch.html>

¹¹⁵ <https://web.archive.org/web/20220701070938/https://www.segeln360.de/yachtcharter/unterwegs-waehrend-des-toerns/logbuch.html>



Archivwürdigkeit:

- Kommission Archäologie und Informationssysteme im Verband der Landesarchäologen Deutschlands. (2017). *Themenblätter zur Archivierung digitaler Daten – Archivwürdigkeit*. https://www.landesarchaeologien.de/fileadmin/mediamanager/004-Kommissionen/Archaeologie-und-Informationssysteme/Archivierung/TB_Archivwuerdigkeit.pdf [Wayback](#)¹¹⁶
- Schweizerisches Bundesarchiv BAR. (2022, Oktober 31). *Archivwürdigkeit*. <https://www.bar.admin.ch/bar/de/home/informationsmanagement/archivwuerdigkeit.html>
- Chiquet, S. (2001). Was heisst eigentlich archivwürdig? *Schweizerische Zeitschrift für Geschichte = Revue suisse d'histoire = Rivista storica svizzera*, 51(2001), 470–486. <https://doi.org/10.5169/seals-107875>
- ISO 15489. (2023). In *Wikipedia*. https://de.wikipedia.orghttps://de.wikipedia.org/w/index.php?title=ISO_15489&oldid=239913830

Weiterführende Informationen zum ökologischen Fußabdruck der Digital Humanities-Forschungsaktivität:

- Verband Digital Humanities im deutschsprachigen Raum e. V. (2022). *AG Greening DH*. <https://dig-hum.de/ag-greening-dh> [Wayback](#)¹¹⁷



Archivwürdigkeit der Forschungsdaten:

- Können Teile Ihrer Daten potentiell **von anderen Forscher:innen nachgenutzt** werden?
- Ist das **Projekt**, in dem die Daten erstellt wurden, bereits **abgeschlossen**?
- Sind Teile Ihrer Daten bereits **anderswo archiviert** worden?
- Erfüllen Ihre Daten die **Qualitätskriterien** gemäß den [Empfehlungen der DFG](#)¹¹⁸?

¹¹⁶ https://web.archive.org/web/20231024161326/https://www.landesarchaeologien.de/fileadmin/mediamanager/004-Kommissionen/Archaeologie-und-Informationssysteme/Archivierung/TB_Archivwuerdigkeit.pdf

¹¹⁷ <https://web.archive.org/web/20230708202950/https://dig-hum.de/ag-greening-dh>

¹¹⁸ <https://doi.org/10.5281/zenodo.6472827>

3.3 Klärung rechtlicher und ethischer Aspekte

Rechtliche Aspekte

Wurde festgestellt, dass Daten erhaltenswert sind, müssen rechtliche Aspekte geprüft werden. Repositorien möchten in der Regel die rechtlichen Risiken vermeiden, nehmen dafür die Forschenden in die Pflicht und halten das Abkommen in einer schriftlichen Vereinbarung fest. Diese ist oft Bestandteil des Dienstleistungsvertrages.

Bei der digitalen Langzeitarchivierung ist zu beachten, dass es in der Verantwortung des Forschenden bzw. Datengebenden liegt, alle Rechte Dritter vor Veröffentlichung und Archivierung zu klären. Dies ist im → [Datenmanagementplan](#) bereits vor Projektbeginn zu berücksichtigen.

Hinsichtlich der rechtlichen Aspekte spielen Datenschutz und Urheberrechte eine tragende Rolle. In der Regel verlangen Archive von ihren Datengebenden, dass eingelieferte digitale Dokumente frei von Rechten Dritter (wie Urheberrechten, personenbezogenen Daten sowie Betriebs- und Geschäftsgeheimnissen) sind. Deshalb ist es notwendig und auch in der Verantwortung der Autor:innen, bereits vor der Publikation und der Langzeitarchivierung diese von Dritten einzuholen. Im Idealfall wird dies bereits im Datenmanagementplan berücksichtigt und zeitlich mit eingeplant. Hierfür werden die Rechteinhaber:innen oder deren Vertreter:innen direkt kontaktiert, um sich die Nutzungsrechte einräumen zu lassen. Diese und die Konditionen dafür sollten immer schriftlich mit den Rechteinhaber:innen festgehalten werden, da diese ggf. dem Langzeitarchiv oder der Publikationsplattform vorgelegt werden müssen. Ein solches Abkommen ist dann Grundlage der Vereinbarung zwischen Archiv und Forschenden. Einige der sich daraus ergebenden Probleme lassen sich durchaus lösen, allerdings nicht ohne einen gewissen Aufwand. Bestehen Urheberrechte, kann die notwendige Werknutzung möglicherweise lizenziert werden, wodurch jedoch Kosten entstehen können.

Ein wesentlicher Faktor ist der zukünftige Zugriff zu den archivierten Daten: Soll er offen sein oder durch das Langzeitarchiv verwaltet werden? Dies ist auch relevant, wenn im Datenarchiv datenschutzrechtliche Positionen geschützt werden müssen. Das Archivrecht bietet hier ein Modell des gestaffelten und kontrollierten Zugangs, das grundsätzlich auf digitale Daten übertragbar ist, jedoch zukünftigen Aufwand bedeutet. Das bedeutet hier, dass Zugriffsrechte an vordefinierte Account-Typen vergeben werden können. In Einzelfällen müssen Anfragen zur Berechtigung geprüft werden, was mit der Klärung datenschutzrechtlicher Aspekte einhergeht: Habe ich die betroffenen Personen darüber informiert, ob und dass ich ihre Daten in ein Langzeitarchiv geben möchte? Muss ich bestimmte Betroffenenrechte beachten, Teile der Daten anonymisieren oder gar vor der Speicherung löschen? Und schließt die Archivierung die Weitergabe an Dritte zugunsten einer Nachnutzung für die Forschung ein? Diese

und weitere Fragen doppeln sich oft in beiden Rechtsgebieten, führen aber gelegentlich zu unterschiedlichen Lösungswegen. Diese in Einklang zu bringen, erfordert oft juristischen Rat. Wie die Forschungsdaten nachgenutzt werden können, wird durch die Vergabe von Lizenzen geregelt. Diese werden von Forschenden bzw. Datengebenden nach Klärung der Rechte ausgewählt und meist in einer schriftlichen Einverständniserklärung mit dem Langzeitarchiv festgehalten. Hier ist zu beachten, dass Lizenzen für die enthaltenen Komponenten (z. B. Bilder, Software etc.) differenziert anzugeben sind.



Rechtliche Aspekte:

- Klimpel, Paul. „Urheberrechtsreform 2021: neue Chancen für das kulturelle Erbe.“ Digitales Deutsches Frauenarchiv. November 2021. <https://nbn-resolving.org/urn:nbn:de:0297-zib-84315>
- “Handling Sensitive Data.” *The Qualitative Data Repository*, March 5, 2018. <https://qdr.syr.edu/policies/sensitive-data>  [Wayback](https://web.archive.org/web/20221104024025/https://qdr.syr.edu/policies/sensitive-data)¹¹⁹.
- forschungsdaten.info. (2023, Mai 12). *Forschungsdaten veröffentlichen | Rechte und Pflichten*. <https://forschungsdaten.info/themen/rechte-und-pflichten/forschungsdaten-veroeffentlichen/>
- Beinert, T. (2009). *Wege ins Archiv: Ein Leitfaden für die Informationsübernahme in das digitale Langzeitarchiv*. <https://doi.org/10.18452/1527>

Umgang mit personenbezogenen Daten:

- Soltau, K. (2023). *RADAR4Culture: Handreichung zu personenbezogenen Daten [deutsch]*. <https://doi.org/10.5281/ZENODO.8221496>
- Herklotz, M., Nguyen, L., & Oberländer, L. (2022, August 9). *Interactive Virtual Assistant iVA1*. Anwendungsbereich DSGVO. https://wiki.bib.uni-mannheim.de/xerte/play.php?template_id=224#page1
- Gerlach, R., Rex, J., Lang, K., Neute, N., & Schröter, A. (2020). *Fact Sheet: Personal Data*. <https://doi.org/10.5281/zenodo.4035992>

Bildrechte in der Kunstgeschichte:

- Verband Deutscher Kunsthistoriker. (2022). *Bildrechte in der kunsthistorischen Praxis – ein Leitfaden*. - Zweite, überarbeitete und erweiterte Auflage. Heidelberg University Library. <https://doi.org/10.11588/ART-DOK.00007769>

Lizenzen:

- für Texte, Abbildungen und Daten:
forschungsdaten.info. (2023, Mai 12). *Creative-Commons-Lizenzen | Forschungsdaten veröffentlichen | Rechte und Pflichten*. <https://forschungsdaten.info/themen/rechte-und-pflichten/forschungsdaten-veroeffentlichen/creative-commons-lizenzen/>
- für Software:
Free Software Foundation, Inc. (2023, Juni 11). *GNU General Public License, Version 3—GNU-Projekt—Free Software Foundation*. <https://www.gnu.org/licenses/gpl-3.0.html>

¹¹⁹ <https://web.archive.org/web/20221104024025/https://qdr.syr.edu/policies/sensitive-data>

- für Datenbanken:
Open Data Commons: Legal tools for open data. (o. J.). Abgerufen 4. August 2023, von <https://opendatacommons.org/>
- für Daten:
The Linux Foundation®. (2021, Juni 22). *Community Data Licence Agreement*. CDLA. <https://cdla.dev/>

Ethische Aspekte

Neben rechtlichen Aspekten müssen beim verantwortungsvollen Umgang mit Forschungsdaten auch ethische Aspekte geprüft werden. Der Provenienznachweis spielt dabei eine tragende Rolle, aber auch Fragen nach der verantwortungsvollen Übernahme von Objekten im Akquise- bzw. Aufnahmeprozess oder der Sicherung bzw. Dokumentation von Inhalten, die unter ethischen Gesichtspunkten problematisch sein können. Ebenso gehören dazu Fragen wie Zugriffsregelungen oder Anonymisierung der Daten. Generell werden bei der Klärung ethischer Aspekte in Zukunft auch die bereits erwähnten →[CARE-Prinzipien](#) oder die insbesondere auf Big Data-Analysen abzielenden [FACT-Prinzipien](#) eine zunehmende Rolle spielen. Neben den modernen ethischen Prinzipien-Bündeln ist aber auch auf etablierte wissenschaftsethische Standards der jeweiligen Disziplin zurückzugreifen.



Weiterführende Informationen:

- GIDA - Global Indigenous Data Alliance: CARE Principles for Indigenous Data Governance. 2018.
<https://www.gida-global.org/care>
- Deppe, A. (2020). FAIR, CARE und mehr. Prinzipien für einen verantwortungsvollen Umgang mit Forschungsdaten. In Matthias Schulze (Hrsg.): *Historisches Erbe und zeitgemäße Informationsinfrastrukturen: Bibliotheken am Anfang des 21. Jahrhunderts* (S. 299–312). Kassel university press.
<http://dx.doi.org/doi:10.17170/kobra-202010131934-299>



Rechtliche und ethische Fragen:

- Sind Sie in allen Teilen Ihrer Forschungsdaten auch Rechteinhaber oder **bestehen Rechte Dritter**?
- Bestehen Rechte Dritter, ist es notwendig, für jedes einzelne Objekt eine entsprechende **Erlaubnis** einzuholen.
- Sollen Ihre Daten in der Zukunft **frei zugänglich** gemacht werden? Ist das bei Fremddaten beispielsweise durch **Lizenzen** abgedeckt?
- Dürfen Ihre Forschungsdaten nachgenutzt werden, vergeben Sie **geeignete Lizenzen** (z. B. [Creative Commons](#)).
- Prüfen Sie den **Datenschutz**, insbesondere bei personenbezogenen Daten.
- Prüfen Sie, inwiefern Ihre Forschungsdaten **ethische Aspekte** berühren.
- Sind Sie sich einzelnen Punkten nicht sicher, holen Sie sich den **Rat von Expert:innen**, z. B. des Langzeitarchivs Ihrer Wahl oder über den →[NFDI4Culture Helpdesk](#).

3.4 Identifizierung erhaltenswerter Eigenschaften

Erhaltenswerte Eigenschaften (auch „signifikante Eigenschaften“ genannt) bilden die Grundlage für das archivarische Handeln, wenn es darum geht, die Nutzbarkeit der Forschungsdaten zu erhalten (→siehe Abschnitt [Content Preservation](#)). In diesem Kontext bilden sie die Basis für den Erhalt der Nutzbarkeit des Archivmaterials. Selbst wenn es zunächst nur darum geht, die Datenintegrität des eingelieferten Archivmaterials zu erhalten, ist die Erfassung erhaltenswerter Eigenschaften zu empfehlen, um auch später noch Maßnahmen zur Erhaltung der Nutzbarkeit durchführen zu können.

Eine erhaltenswerte Eigenschaft wäre zum Beispiel die Lesbarkeit von Text innerhalb eines Dokumentes bei üblichem Betrachtungsabstand.

Das Fehlen oder der Verlust von nur einer erhaltenswerten Eigenschaft kann eventuell ein Objekt unbrauchbar machen. Auch Entscheidungen zur →[Formatmigration](#) und zur →[Auswahl von Dateiformaten](#) hängen wesentlich von erhaltenswerten Eigenschaften ab.

Fallbeispiel: signifikante Eigenschaften bei Tabelleninhalten

Kalkulierte Kosten:		(ab dem vierten Jahr muß kein Vertragsverhältnis bestehen)					
Jahr	Jahr	Jahr	Jahr	Jahr	Jahr	Jahr	
3	4	5	6	7	8	9	
1.006,35 €	1.506,35 €	-1.506,35 €	-1.506,35 €	-1.506,35 €	-1.506,35 €	-1.506,35 €	
1.063,50 €	1.563,50 €	-1.563,50 €	-1.563,50 €	-1.563,50 €	-1.563,50 €	-1.563,50 €	
1.635,00 €	2.135,00 €	-2.135,00 €	-2.135,00 €	-2.135,00 €	-2.135,00 €	-2.135,00 €	
7.350,00 €	7.850,00 €	7.850,00 €	7.850,00 €	7.850,00 €	7.850,00 €	7.850,00 €	
64.500,00 €	65.000,00 €	65.000,00 €	65.000,00 €	65.000,00 €	65.000,00 €	65.000,00 €	
534.000,00 €	544.000,00 €	544.000,00 €	544.000,00 €	544.000,00 €	544.000,00 €	544.000,00 €	

Kosten: (bei dieser Kalkulation bleibt das Vertragsverhältnis bestehen, falls...)

Jahr	Jahr	Jahr
5	6	7

Abbildung 3: Beispiel einer Tabellenkalkulationsdatei mit farbig codierten Informationen. (Foto: © Matthias Razum, CC BY 4.0 International)

Tabellarische Daten sollten maschinenlesbar sein und nicht in proprietären Formaten vorliegen. Dafür eignet sich z. B. das Tabellenformat CSV, in dem Werte durch Kommata getrennt werden. In diesem Beispiel sind Teile der Ursprungsdatei farbig codiert. Das könnte im CSV-Format problematisch sein, denn dieses deckt derartige Zusatzinformationen nicht ab. Es ist also zu prüfen, ob die Farbwerte eine erhaltenswerte Eigenschaft sind. Da es sich im Beispiel bei der roten Markierung um die Hervorhebung von negativen Zahlen handelt, diese aber zusätzlich das Negativzeichen aufweisen, bedeutet der Wegfall der Farbe keinen Informationsverlust. Die Farbe ist damit keine erhaltenswerte Eigenschaft, und das CSV-Format kann hier verwendet werden.

Falls nun das Archiv Tabellen nur als CSV archiviert, müsste der Rat an die Datenproduzierenden lauten, über Farbauszeichnung getragene Informationen in zusätzliche Tabellenspalten zu extrapolieren oder Minuszeichen einzufügen. Andernfalls wäre eine Überführung in das CSV-Format nicht möglich.

Fallbeispiel: signifikante Eigenschaften bei 3D-Modellen



Abbildung 4: [3D-Tastmodell des Voelkerschlachtdenkmals](#)¹²⁰ des Stadtgeschichtlichen Museums in Leipzig (links, Maßstab: 1:100, Foto: © Sandrino Donnhauser) sowie seine [virtuelle Präsentation](#)¹²¹ (rechts, 3D-Modell: © [Eric Eschrich](#)¹²²).

3D-Modelle sollten in nicht-proprietären Formaten vorliegen. Dafür eignet sich das STL-Format, das Geometrieinformationen in Form von Dreiecksfacetten speichern kann. Möchte man 3D-gescannte Gegenstände oder Gebäude mit einfarbigem Filament 3D-drucken, um diese z. B. als Kunststoffastmodelle für sehbeeinträchtigte und blinde Menschen in Museen wie im linken Bild der oberen Abbildung zu nutzen, so stellt der Erhalt von Farbinformationen keine signifikante Eigenschaft dar. Soll das digitalisierte Denkmal hingegen für eine wirklichkeitstreue virtuelle Präsentation gesichert werden, sind Farbinformationen eine erhaltenswerte Eigenschaft und ein anderes archivfähiges 3D-Format wie etwa glTF oder X3D muss verwendet werden. Falls das Archiv keine 3D-Daten archiviert, müsste der Rat an die Datenproduzierenden lauten, alternative Darstellungsweisen wie 360°-Videos oder Bilder aus unterschiedlichen Perspektiven einzuliefern.

Erhaltenswerte Eigenschaften werden meist von den Datengebenden mit Unterstützung der Archivar:innen bestimmt. Wichtig ist, dass bei der Einlieferung von Materialien ins Archiv die Bestimmung erhaltenswerter Eigenschaften generell ganz am Anfang stehen sollte. Auch sollten sich Forschende bereits zu Projektbeginn mit dem digitalen Archiv beraten, um die Weichen rechtzeitig in ihrer Datenhaltung zu stellen.

Bei der Ermittlung dieser Eigenschaften muss in Betracht gezogen werden, um welches Material es sich handelt, zu welchem Zweck es archiviert werden soll und für welche Zielgruppen es geschaffen wurde. Dabei definiert das Archiv auf Basis von Checklisten sogenannte

¹²⁰ <https://de.wikipedia.org/wiki/Tastmodell>

¹²¹ <https://sketchfab.com/3d-models/voelkerschlachtdenkmal-0fca79c19c564a36a2ad479406a26bf4>

¹²² https://hayle.artstation.com/projects/bKd1Xk?album_id=3483

„Objekttypgruppen“ (Material mit gleichen Eigenschaften), von denen angenommen werden kann, dass sie auf ähnliche Art und Weise genutzt werden. Das Archiv überträgt die erfassten Eigenschaften dann in ein geeignetes Format. Das SLUBArchiv.digital stellt dafür ein eigenes XML-Schema „[significant properties](#)“¹²³ ([Anleitung](#)¹²⁴) bereit, mit dem erhaltenswerte Eigenschaften in einer XML-Datei erfasst werden können. Bei der Einlieferung von Datenpaketen prüft das Archiv das Vorhandensein erhaltenswerter Eigenschaften und überführt diese dann in **technische Eigenschaften**: So ist die Anforderung, dass auf einem Bild enthaltene Textinformationen bei üblichem Betrachtungsabstand auf einem Bildschirm lesbar sein sollen, eine abstrakte erhaltenswerte Eigenschaft; die Anforderung, dass bei Rastergrafiken mit einer Auflösung von mindestens 300dpi gearbeitet werden muss, wäre eine abgeleitete technische Eigenschaft.

Bei der Bestimmung erhaltenswerter Eigenschaften gibt es unterschiedliche Ansätze, wie es am Fallbeispiel „Stein von Rosetta“ (siehe Abbildung 5) veranschaulicht werden kann.

Der **provenienzbasierte Ansatz** fragt danach, welche überlieferte oder vermutete Absicht in der Vergangenheit bei der Erstellung eines analogen oder digitalen Objekts verfolgt wurde. Im Falle des Steins von Rosetta war etwa die Vermittlung des eingemeißelten Textes so wichtig, dass dieser gleich in drei Sprachen verfasst wurde. Nach diesem Ansatz hätte bei der digitalen Langzeitarchivierung daher der semantische Erhalt des gesamten Textes mit umfangreichen Kontextinformationen Priorität.

Der **nutzergruppenbasierte Ansatz** hingegen fragt danach, welche Bedürfnisse eine Community zukünftig haben wird und was daher mit den archivierten Objekten geschehen soll. Außerdem stellt sich hier die Frage, welcher Anspruch daraus an die Qualität der Daten im Archiv erwächst. Für die Langzeitarchivierung des Steins von Rosetta hätten dann erhaltenswerte Eigenschaften vor allem zum äußeren Erscheinungsbild Priorität. Dabei ist der Zweck entscheidend, je nachdem, ob die Objektdaten nach einer 3D-Digitalisierung online zur virtuellen Erkundung oder originalgetreu zur Reproduktion im Falle einer Zerstörung aufbewahrt werden sollen.

Die durch beide Ansätze ermittelten erhaltenswerten Eigenschaften ergänzen sich oft. Zusammen werden sie auch als „Preservation Intent“ bezeichnet. Darin legt das Archiv – basierend auf den Informationen der Datengebenden – fest, was wie erhalten werden soll.

¹²³ <https://slubarchiv.slub-dresden.de/fileadmin/groups/slubsite/slubarchiv/standards/sigprops/sigprops.xsd>

¹²⁴ https://slubarchiv.slub-dresden.de/fileadmin/groups/slubsite/slubarchiv/SLUBArchiv_Diskussionspapier_Signifikante_Eigenschaften_v1.2.pdf

Fallbeispiel: erhaltenswerte Eigenschaften aus unterschiedlichen Perspektiven



Abbildung 5: Der [Stein von Rosetta](#)¹²⁵. Links: Bruchstück; ausgestellt im [British Museum](#)¹²⁶ [Wayback](#)¹²⁷; enthält eine dreisprachige Inschrift aus dem Jahr 196 v. Chr., in dem der ägyptische König Ptolemaios V. geehrt wird (Foto: © Olaf Herrmann, free use). Rechts: digitale Rekonstruktion des kompletten Steins (Grafik: © [Captmondo](#)¹²⁸, CC BY-SA 3.0).

Beim Stein von Rosetta wurden sowohl einfache Fotografien aus unterschiedlichen Perspektiven als auch 3D-Modelle mittels Photogrammetrie-Methoden für eine breite Zielgruppe – „vom Wissenschaftler bis zum Bürger“ – erstellt: Einerseits für museale Zwecke, vom Ansehen der Steinoberfläche bis zum Erschließen von Kontextinformationen, andererseits zum Zwecke der [Rekonstruktion](#)¹²⁹ [Wayback](#)¹³⁰. 3D-Modelle werden dabei online auf der [Sketchfab-Seite](#)¹³¹ [Wayback](#)¹³² des British Museum zur Verfügung gestellt. Fotografien mit weiterführenden Informationen z. B. über das Objekt und zur Handhabung sowie Aufzeichnungen zur Konservierung werden über das [Webpräsentationssystem COL](#)¹³³ [Wayback](#)¹³⁴ des British Museum kostenlos bereitgestellt. Einige Felder sind aus Gründen der Sicherheit und Vertraulichkeit nicht öffentlich einsehbar. Alle Aufzeichnungen werden dabei regelmäßig mit neuen Forschungsergebnissen und Bibliografien aktualisiert und sind einige Tage später online sichtbar.

Fotografien und 3D-Modelle werden zur Webpräsentation in niedriger Auflösung der Öffentlichkeit zur Verfügung gestellt. Fotografien und 3D-Modelle in hoher Auflösung samt allen Metadaten werden hingegen im digitalen Archiv des British Museum aufbewahrt und können auf Anfrage bereitgestellt werden.

¹²⁵ https://de.wikipedia.org/wiki/Stein_von_Rosette

¹²⁶ <https://www.britishmuseum.org/blog/everything-you-ever-wanted-know-about-rosetta-stone>

¹²⁷ <https://web.archive.org/web/20231115055246/https://www.britishmuseum.org/blog/everything-you-ever-wanted-know-about-rosetta-stone>

¹²⁸ <https://commons.wikimedia.org/wiki/User:Captmondo>

¹²⁹ <https://www.britishmuseum.org/exhibitions/hieroglyphs-unlocking-ancient-egypt>

Es ist empfehlenswert, zuerst erhaltenswerte Eigenschaften zu identifizieren, dann darauf aufbauend technische Parameter festzulegen und anschließend erst geeignete Dateiformate auszuwählen. Erst danach sollte die Digitalisierung durchgeführt werden.

Vor der Digitalisierung großer homogener Datenmengen eines Projektes (wie z. B. bei der Digitalisierung historischer Kunstzeitschriften oder mittelalterlicher Handschriften) müssen erhaltenswerte Eigenschaften oft nur einmal für den jeweiligen Objekttyp erfasst werden. Bei großen, inhomogenen Datenmengen verschiedener Projekte kann dies aufwändiger werden. Da die digitale Langzeitarchivierung in den meisten Infrastruktureinrichtungen noch nicht fest etabliert ist, wird dieser Schritt im Prozess der Digitalisierung oft ausgelassen. Dies kann jedoch die Nutzbarkeit der Objekte langfristig gefährden, wenn diese Jahrzehnte später wieder aus dem Archiv ausgespielt werden.



Weitere Informationen zu erhaltenswerten Eigenschaften:

- TIB Hannover. (2023, Mai 16). *Signifikante Eigenschaften. Digitale Langzeitarchivierung an der TIB - TIB Wiki*. <https://wiki.tib.eu/confluence/display/lza/Signifikante+Eigenschaften>  ¹³⁵
- SLUB Dresden. (2023). *Technische Standards für die Ablieferung von digitalen Dokumenten*. <https://slubarchiv.slub-dresden.de/technische-standards-fuer-die-ablieferung-von-digitalen-dokumenten>  ¹³⁶
 - SLUB Dresden. (2022). *Signifikante Eigenschaften im SLUB Langzeitarchiv – Diskussionspapier*. SLUB Dresden. https://slubarchiv.slub-dresden.de/fileadmin/groups/slubsite/slubarchiv/SLUBArchiv_Diskussionspapier_Signifikante_Eigenschaften_v1.2.pdf
 - SLUB Dresden. (2022). *XML-Schema von signifikanten Eigenschaften im SLUBArchiv (1.1)*. SLUB Dresden. <https://slubarchiv.slub-dresden.de/fileadmin/groups/slubsite/slubarchiv/standards/sigprops/sigprops.xsd>
 - SLUB Dresden. (2023). *Maschinenlesbare Definitionen von signifikanten Eigenschaften der im SLUBArchiv bewahrten Objektgruppen*. SLUB Dresden. <https://git.slub-dresden.de/digital-preservation/significantproperties>

¹³⁰ <https://web.archive.org/web/20231031132944/https://www.britishmuseum.org/exhibitions/hieroglyphs-unlocking-ancient-egypt>

¹³¹ <https://sketchfab.com/models/1e03509704a3490e99a173e53b93e282>

¹³² <https://web.archive.org/web/20230317101059/https://sketchfab.com/3d-models/the-rosetta-stone-1e03509704a3490e99a173e53b93e282>

¹³³ https://www.britishmuseum.org/collection/object/Y_EA24

¹³⁴ https://web.archive.org/web/20230911225446/https://www.britishmuseum.org/collection/object/Y_EA24

¹³⁵ <https://web.archive.org/web/20221124215924/https://wiki.tib.eu/confluence/display/lza/Signifikante+Eigenschaften>

¹³⁶ <https://web.archive.org/web/20230605042133/https://slubarchiv.slub-dresden.de/technische-standards-fuer-die-ablieferung-von-digitalen-dokumenten>

- Web, C., Pearson, D., & Koerben, P. (2013). „Oh, you wanted us to preserve that?!“ Statements of Preservation Intent for the National Library of Australia’s Digital Collections. *D-Lib Magazine*, 19(1/2). <https://doi.org/10.1045/january2013-webb>
- Bussmann, B. (2014). *Die Bestandserhaltung digitaler Informationen mittels der Definition von signifikanten Eigenschaften*. <https://nbn-resolving.org/urn:nbn:de:kobv:525-8831>



Identifizierung erhaltenswerter Eigenschaften:

- Welchen **Objekttypgruppen** lässt sich das einzuliefernde Material zuordnen? (z. B. elektronische Publikationen, retrodigitalisierte Bücher, 16mm-Film, Fotografien, Tonkassetten, Münzen, ...)
- Von wem** sollen die digitalisierten Objekte zukünftig genutzt werden? (z. B. breite Öffentlichkeit, Museen, Aussteller:innen, Wissenschaftler:innen, Privatpersonen, ...)
- Wozu** sollen die digitalisierten Objekte **zukünftig** genutzt werden? Welche Absicht stand **ursprünglich** hinter der Herstellung der Objekte? Entwickeln Sie **Nutzungsszenarien** (z. B. ansehen, anhören, Material analysieren, vergleichen, reproduzieren, ausdrucken, annotieren, auffinden, zitieren, ...).
- Welche **signifikanten Eigenschaften** müssen erhalten werden, damit diese Intention weiterhin erfüllt wird?
 - **Inhalt** (z. B. Ist gesprochene Sprache verständlich?)
 - **Kontext** (z. B. Ist die Authentizität durch Prüfsummen nachweisbar?)
 - **Struktur** (z. B. Sind Seitenzahlen, Lesezeichen oder Timecodes erhalten geblieben?)
 - **Aussehen** (z. B. Gibt es keine visuell wahrnehmbaren Änderungen zum Original?)
 - **Verhalten** (z. B. Sind Ton und Bild eines Videos synchron?)
- Wie können diese Eigenschaften in **technische Parameter** übersetzt werden? Welche **Dateiformate** sind hierfür geeignet?

Dateiformate für die Langzeitarchivierung

Die meisten Langzeitarchive haben bei der Einlieferung keine Formatbeschränkung. Die Prüfung und Auswahl von Formaten einzuliefernder Materialien ist vorrangig für die → [Formatmigration](#) entscheidend. Hier muss auf die Verwendung archivfähiger Formate geachtet werden, um die Erhaltung der Nutzbarkeit noch Jahrzehnte später zu gewährleisten. Für andere Erhaltungsmaßnahmen sind die Datenformate weniger wichtig. Fällt die Entscheidung bspw. im Kontext von komplexen Objekten auf eine → [Emulation](#) (den Prozess des Nachahmens eines Computersystems durch ein anderes), existieren bisher kaum archivfähige Formate und Standards.



Die Festlegung auf Dateiformate für die Langzeitarchivierung sollte bereits vor Projektanfang bedacht werden. Die Auswahl der Formate sollte auch im → [Datenmanagementplan](#) berücksichtigt werden.

In Forschungsprojekten ist es nicht unüblich, Forschungsdaten in verschiedenen Formaten zu sichern, beispielsweise einer archivfähigen Masterkopie und einem Derivat für die Online-Präsentation. Bei der Archivierung werden beide Dateiformate eingeliefert.

Empfohlene Dateiformate für die digitale Langzeitarchivierung

Es folgt eine kurze Übersicht an Dateiformaten, die allgemein für die digitale Langzeitarchivierung mit Formatmigration empfohlen werden. Alle **fett** dargestellten Formate werden dabei für die Langzeitarchivierung **präferiert**, alle nicht fett dargestellten Formate hingegen akzeptiert. Da alle Formate dem Alterungsprozess unterliegen, können sich in der Zukunft Änderungen ergeben.



Die Prüfung und Auswahl von Formaten für die digitale Langzeitarchivierung sollte immer eine Schlussfolgerung der → [Identifizierung erhaltenswerter Eigenschaften](#) sein.

Anwendungsbereich	Empfohlene Archivformate
Rasterbilder	Unkomprimiertes Baseline- TIFF (ab Version 6.0), GeoTIFF
Vektorbilder	Unkomprimiertes SVG
CAD-Modelle	IFC, STP, DXF . Wenn funktional nicht ausreichend, dann DWG mit älterer Version 2010 (AC1024)
3D-Modelle	Für Punktwolken: E57 . Sonst: gITF, DAE, OBJ . Wenn funktional nicht ausreichend, dann X3D

Text	PDF/A-2, PDF/A-1 (aber keine anderen PDF-Varianten), ODT, DOCX (aber nicht DOC), TXT mit UTF-8 ohne BOM-Kodierung
Strukturierte Texte	XML mit DTD-Datei oder XML-Schema, JSON, HTML
Tabellen	CSV mit UTF-8 ohne BOM-Kodierung, ODS, XLSX (aber nicht XLS), XML mit DTD-Datei oder XSD-Schema, HTML
Datenbanken	SIARD2 (ab Version 2.0), SQL (ab Version 2008), XML mit XSD-Schema
Dokumente	PDF/A-2, PDF/A-1 (aber keine anderen PDF-Varianten), ODT, DOCX (aber nicht DOC)
Video	Matroska MKV mit Video-Codec FFV1 sowie FLAC als Audio-Codec
Audio	FLAC, BWF mit linearer Puls-Code-Modulation (LPCM), WAVE mit LPCM, Matroska MKA mit LPCM oder FLAC-Kodierung. Alle mit mindestens 48 kHz und 24 bit pro Kanal
Webseiten	PDF/A-2, PDF/A-1 (aber keine anderen PDF-Varianten), WARC, HTML (z. B. ohne JavaScript)

Tabelle 2: Empfohlene Dateiformate für die digitale Langzeitarchivierung im Überblick. (Stand: 07/2023)



Weitere Informationen zu empfohlenen Dateiformaten:

- forschungsdaten.info. (2023, Mai 12). *Forschungsdaten veröffentlichen | Rechte und Pflichten*. <https://forschungsdaten.info/themen/rechte-und-pflichten/forschungsdaten-veroeffentlichen/>
- Deutsches Archäologisches Institut. (2023, Juni 1). *IANUS - Forschungsdatenzentrum Archäologie & Altertumswissenschaften*. <https://ianus-fdz.de/langzeitarchivierung/>
- SLUB Dresden. (2021). *Langzeitarchivfähige Dateiformate*. SLUB Dresden. https://slubarchiv.slub-dresden.de/fileadmin/groups/slubsite/slubar-chiv/SLUBArchiv_langzeitarchivfaehige_Dateiformate_v2.0.pdf
- Open Preservation Foundation. (2022, Dezember). *International Comparison of Recommended File Formats*. Google Docs. https://docs.google.com/spreadsheets/d/1XjE-jFBCGF3N1spNZc1y0DG8_Uyw18uG2j8V2bsQdYjk/edit
- Laagland, H., Alloing, S., Gilissen, V., Steeman, M., & Swagemakers, W. (2020). *White paper on preferred formats*. <https://doi.org/10.5281/zenodo.4518486>
- Rudnik, P. (Regisseur). (2020, August 14). *Video: Crashkurs Digitale Langzeitarchivierung - Dateiformate*. Zenodo. <https://doi.org/10.5281/zenodo.3985075>
- Quast, A., & Ogan, K. (2023). *Interaktive Tafel gängiger Dateiformate*. <https://www.lzv.nrw/dateiformate/>

Auswahlkriterien für Dateiformate

Das wichtigste Kriterium für die Auswahl eines Dateiformats ist der **Grad an technischen Eigenschaften**, die zum Bewahren signifikanter Eigenschaften benötigt werden. So kann etwa das 3D-Format „STL“, welches die Oberfläche von dreidimensionalen Objekten mittels Dreiecksfacetten beschreibt, standardmäßig nur farblose Drahtgittermodelle speichern. Werden Gegenstände 3D-gescannt, um sie mit einfarbigem Filament 3D-zu-drucken, ist das STL-Format zur Speicherung dieser Daten ausreichend. Sollen hingegen auch Farbinformationen abgelegt werden, muss ein anderes Datenformat wie z. B. „glTF“ verwendet werden.

Neben technischen Eigenschaften müssen bei Datenformaten noch weitere **Bedingungen** erfüllt sein, damit Daten **archivfähig** werden. Dazu gehören:

- ein hoher Offenlegungsgrad,
- eine hohe Verbreitung in der Community,
- eine geringe Komplexität der Formatspezifikation,
- eine hohe Selbstdokumentation,
- eine hohe Robustheit gegenüber Datenfehlern und
- eine geringe Abhängigkeit von anderen Ressourcen.

Ein **hoher Offenlegungsgrad** besteht, wenn die Spezifikation sowie Dokumentation eines Formates frei zugänglich sind und das Format mit einer offenen Lizenz versehen ist. Eine Formatspezifikation beschreibt die genaue Anordnung der Daten innerhalb eines Dateiformats und die Art der Codierung. Anhand dieser Spezifikation kann ein Archiv später im Rahmen einer Formatmigration Dateien lesen, interpretieren und erhalten. Ob ein bestimmtes Dateiformat gemäß der Formatspezifikation vorliegt, kann mit verschiedenen Validierungswerkzeugen geprüft werden (→siehe Abschnitt [Formatvalidierung](#)). Um einen hohen Offenlegungsgrad zu erreichen, sollte das Format zudem standardisiert sein, idealerweise auf internationaler Ebene mit einem ISO-Standard.

Für das Format PDF/A-1 ist beispielsweise die detaillierte Formatspezifikation im [ISO-Store](#)¹³⁷ online käuflich verfügbar. PDF/A-1 hat daher einen hohen Offenlegungsgrad. Die Formatspezifikation von Baseline-TIFF 6.0 steht [kostenlos bereit](#)¹³⁸. Das Format hat daher einen noch höheren Offenlegungsgrad.

Eine **hohe Verbreitung** bedeutet, dass viele verschiedene existierende Tools dieses Format bereits unterstützen und von der Community aktiv auch in der Zukunft genutzt werden. Mittlerweile können nicht nur aktuelle Computer-, sondern auch Smartphone-Betriebssysteme PDF-Dateien out-of-the-box rendern. PDF/A-1-Dateien können dabei mit jedem PDF-Viewer

¹³⁷ <https://www.iso.org/standard/38920.html>

¹³⁸ <https://www.itu.int/itudoc/itu-t/com16/tiff-fx/docs/tiff6.pdf>

präsentiert werden. Auch das Baseline-TIFF-Format ohne Komprimierung wird nahezu von jedem Bildbearbeitungsprogramm unterstützt. Beide haben daher eine hohe Verbreitung.

Eine **geringe Komplexität** ist erstrebenswert, um das Finden von Fehlern in Dateien zu erleichtern. Diese können z. B. bei der Digitalisierung entstehen. Wird bei der Aufnahme ins Archiv ein Fehler erkannt, muss dieser analysiert und behoben werden. Die Komplexität der [Formatspezifikation von PDF/A-1](#)¹³⁹ ist mit etwa 29 Seiten gering. Zählt man jedoch die Seiten aller PDF-Varianten zusammen, so kommt man auf mindestens 9.000 Seiten, wobei hier genutzte Spezifikationen wie bspw. die JPEG2000-Bildkompression noch nicht inbegriffen sind. Das Format PDF mit allen Varianten gilt daher als komplex. Die [Formatspezifikation von TIFF 6.0](#)¹⁴⁰ hingegen umfasst etwa 121 Seiten und weist damit eine geringe bis mäßige Komplexität auf.

Ein **hoher Grad an Selbstdokumentation** liegt vor, wenn das Format wie PDF und die genaue Version wie PDF/A-1a aus dem Dateinhalt abgelesen werden können. Das wird in der Regel durch Signaturen, sogenannte „MagicBytes“, gelöst, die im Bitstream einer Datei enthalten sind. Bei allen Varianten des PDF- und TIFF-Formats können Format und genaue Version aus dem Bitstream abgelesen werden. Diese sind daher selbstdokumentierend.

Eine **hohe Robustheit** wird erreicht, wenn die ungewollte Veränderung eines Bits bzw. Bytes innerhalb einer Datei kaum Auswirkungen auf die Interpretierbarkeit bzw. die Darstellung des Inhaltes hat. Dabei sind mit Kompression arbeitende Formate anfälliger gegenüber Bitfehlern – unabhängig davon, ob sie verlustfrei oder verlustbehaftet arbeiten. Durch Fehler in komprimierten Daten können größere Teile der Datei unbrauchbar werden. Daher sollte Kompression in der digitalen Langzeitarchivierung vermieden werden. Kompression kann aber zulässig sein, wenn sie zum Beispiel durch Redundanz abgesichert ist, verlustfrei erfolgt, oder die zu erwartende Einsparung an Speicherplatz das Risiko von Datenverlusten aufwiegt. Andererseits benötigen unkomprimierte Daten meist mehr Speicherplatz, daher müssen – besonders bei speicherintensiven Medien wie Video oder Audio – Kosten und Nutzen besonders abgewogen werden. Zwar ist eine PDF/A-1-Datei binär codiert, aber deren Dokumentendatei an sich ist nicht komprimiert. Das Baseline-TIFF-Format ist unkomprimiert, bietet genügend Redundanz, um die meisten Fehler zu beheben, und nur ein kleiner Bereich (genauer: Teile des „Image File Directory“) reagiert stark auf Bitfehler. Daher ist die Robustheit bei beiden Beispielen mäßig bis hoch.

Schließlich bedarf es noch eines möglichst **geringen Grades an Abhängigkeiten** von anderen Ressourcen, wie spezieller Hardware, Software oder Internet-Ressourcen, da diese in Zukunft möglicherweise nicht mehr zur Verfügung stehen. Alle Medien einer PDF/A-Datei

¹³⁹ <https://www.iso.org/standard/38920.html>

¹⁴⁰ <https://www.itu.int/itu-t/com16/tiff-fx/docs/tiff6.pdf>

müssen in den gängigsten Bild- und Vektorformaten eingebettet sein. Links auf externe Ressourcen dürfen darin nicht enthalten sein. Für Baseline-TIFF gibt es keine externen Abhängigkeiten. Daher ist bei beiden Formaten in den jeweiligen Versionen der Grad an Abhängigkeiten gering.

Die Prüfung und Auswahl der Datenformate sind entscheidend für die Datenübernahme in das Archiv und für die spätere Bestandserhaltung, die weitgehend automatisiert abläuft (→siehe Abschnitt [Einreichung beim Archiv](#)).



Robustheit:

- Romeyke, A. (2013, August 27). Art1Pirat Blog: Baseline TIFF – Versuch einer Rekonstruktion, Teil1. *Art1Pirat Blog*. <https://art1pirat.blogspot.com/2013/08/baseline-tiff-versuch-einer.html>  [Wayback](#)¹⁴¹



Geeignete Dateiformate:

- Bestehen **Formatvorgaben** seitens des von Ihnen präferierten Archivs?
- Werden **Erhaltungsmaßnahmen** wie Formatmigration oder Emulation Teil des Langzeitarchivierungsprozesses sein?
- Erfüllt das Format des einzuliefernden Materials alle **technischen Eigenschaften**, die aus den zuvor ermittelten erhaltenswerten Eigenschaften hervorgehen?
- Ist das Format des einzuliefernden Materials **archivfähig**?
 - Ist es standardisiert?
 - Ist die Spezifikation frei zugänglich? Ist es mit einer offenen Lizenz versehen?
 - Ist das Format aktuell und verbreitet?
 - Können Metadaten in die Datei eingebettet werden?
 - Wurde eine menschenlesbare bzw. ASCII-basierte Version gewählt?
- Darf eine **Datenkompression** vorgenommen werden?
- Bestehen **Abhängigkeiten** von anderen Ressourcen, wie spezieller Hardware, Software oder Internet-Ressourcen?

¹⁴¹ <https://web.archive.org/web/20220125221158/http://art1pirat.blogspot.com/2013/08/baseline-tiff-versuch-einer.html>

3.5 Erfassen von Metadaten

In der digitalen Langzeitarchivierung werden nicht nur Forschungsdaten archiviert, sondern immer auch Informationen über diese Dateien – die Metadaten. Erst durch sie wird das Archivpaket nachhaltig nutzbar. Aus Metadaten muss ablesbar sein, was das digitale Objekt repräsentiert, wenn dieses Jahre später wieder aus einem Archiv entnommen wurde.

Metadaten sollten spätestens bei der Einreichung ins Archiv bzw. der Erstellung von Informationspaketen vorliegen. Im Idealfall werden sie aber bereits bei der Erfassung der Forschungsdaten mitbedacht (→siehe Abschnitte [Datenmanagementplan](#) und →[FAIRe Forschungsdaten](#), bzw. den Abschnitt [Empfehlungen für Metadatenstrukturen](#)¹⁴² der NFDI4Culture-Handreichung zum Thema FAIR). Sie können deskriptive, strukturelle, administrative sowie technische Informationen enthalten. Bevor Datengebende bzw. Forschende mit der Erhebung von Metadaten beginnen, sollten sie sich beim Archiv erkundigen, welche Daten erfasst und wie diese dokumentiert werden.

Deskriptive Metadaten beschreiben den Projektkontext und erschließen die Inhalte der Projektdaten. Sie werden in der Regel manuell erfasst. Hierzu gehören u. a. Titel, Namen beteiligter Personen und Institutionen sowie Themenschlagwörter. Deskriptive Metadaten sind besonders für das Auffinden und auch die Wiederverwendung eines digitalen Objektes im Kontext von Linked-Data-Anwendungen wichtig.

Dabei sollte möglichst auf **Normdaten und kontrollierte Vokabulare** zurückgegriffen werden. Hierzu gehören:

- [Gemeinsame Normdatei \(GND\)](#)¹⁴³ für Personen, Körperschaften, Kongresse, Geografika (zu denen die GND auch Bauwerke zählt), Sachschlagwörter und Werke
- [Open Researcher and Contributor ID \(ORCID\)](#)¹⁴⁴ für Autoren und an Forschungsprojekten mitwirkende Personen
- [Research Organization Registry \(ROR\)](#)¹⁴⁵ für Forschungseinrichtungen
- [Crossref Funder Registry](#)¹⁴⁶ für Forschungsförderer
- [Wikidata](#)¹⁴⁷ für Personen, Körperschaften, Veranstaltungen, Geografika, Bauwerke und Werke. Sachschlagwörter werden hier jedoch nicht empfohlen, da die Definitionen oft qualitativ unzureichend sind.

¹⁴² <https://docs.nfdi4culture.de/ta2-fair-handreichung/empfehlungen-zu-metadatenstandards/empfehlungen-fuer-metadaten-strukturen>

¹⁴³ <https://lobid.org/gnd>

¹⁴⁴ <https://orcid.org/>

¹⁴⁵ <https://ror.org/>

¹⁴⁶ <https://www.crossref.org/documentation/funder-registry/>

¹⁴⁷ <https://www.wikidata.org/>

- [Art & Architecture Thesaurus \(AAT\)](#)¹⁴⁸ zur Beschreibung von Objekten der Kunst, Kulturgeschichte und der Architektur sowie weiteren kulturellen Konzepten mit Sachschlagwörtern
- [ICONCLASS](#)¹⁴⁹ zur Kategorisierung vorwiegend abendländischer Bildinhalte

Bei den meisten Repositorien sind bestimmte Felder für Normdaten-IDs zur Verschlagwortung und die Vergabe weiterer Metadaten vorab festgelegt.

In manchen Fällen besteht die Möglichkeit, in enger Abstimmung mit einem Datenkurator zusätzliche Schlagwörter zu vergeben. Das kann erforderlich sein, wenn angebundene Vokabularien bestimmte Forschungsfelder nicht oder nur unzureichend abdecken, z. B. im Kontext außereuropäischer Kulturen.

Strukturelle Metadaten erläutern die Dokumentstruktur. So kann z. B. die Struktur eines retdigitalisierten Buches durch Seitenzählung beschrieben werden, wobei jede Seite auf ein gescanntes Bild sowie einen mittels OCR-Software erstellten Volltext verweist.

Administrative Metadaten klären verwaltungsspezifische und rechtliche Rahmenbedingungen wie Urheberrecht, Weitergabe und Lizenzierung und liefern darüber hinaus auch Kontaktdaten.

Technische Metadaten geben z. B. Informationen zur Nutzung der Software und zur Einstellung der Hardware. Sie werden größtenteils automatisch durch die Erfassungssoftware erstellt. Bei einer digitalen Fotografie sind das unter anderem Auflösung, Verschlusszeit, Blende, geografische Position, Erstelldatum und -zeit sowie die verwendete Hard- und Software. Im Bereich 3D werden verwendete Scanner und deren Einstellungen nicht immer von allen Geräten aufgezeichnet und müssen daher manuell dokumentiert werden. Es ist empfehlenswert, verwendete Hilfsmittel wie Drehteller oder Stative, aber auch alle Bearbeitungsschritte mit Parametern und Ergebnissen zu erfassen, wie z. B. das Entfernen von Störungen oder das Anwenden von Transformationen.

Alle Metadaten werden für Objekttypgruppen in fachlicher und archivarischer Zusammenarbeit festgelegt, was üblicherweise Bestandteil der Übernahmevereinbarung ist.

Um alle Arten von Metadaten einfach maschinell weiterverarbeiten zu können, müssen diese möglichst in verbreiteten Metadatenschemata bereitgestellt werden. Im SLUBArchiv.digital werden beispielsweise alle Metadaten zum Archivgut durchweg als XML-Dateien festgehalten.

¹⁴⁸ <https://www.getty.edu/research/tools/vocabularies/aat/about.html>

¹⁴⁹ <https://iconclass.org/>

Metadaten schemata sind Zusammenstellungen von Elementen zur Beschreibung von Daten. Im Folgenden werden Metadaten schemata vorgestellt, die in der digitalen Langzeitarchivierung häufig verwendet werden:

- [PREMIS](#)¹⁵⁰ (Preservation Metadata: Implementation Strategies) hat sich zum De-facto-Standard für LZA-Metadaten entwickelt. Es wird insbesondere für technische und Provenienz-Metadaten verwendet. Das [PREMIS Data Dictionary](#)¹⁵¹ beschreibt dabei wesentliche, für die Langzeitarchivierung relevante semantische Einheiten. PREMIS ist implementierungsagnostisch, d. h. es kann neben XML auch in einer anderen Auszeichnungssprache wie z. B. RDF implementiert werden.
- Das Schema [METS](#)¹⁵² (Metadata Encoding and Transmission Standard) wird oft für die Beschreibung von Sammlungsstrukturen verwendet. Als sogenanntes [Containerformat](#)¹⁵³ kann es auch andere Schemata enthalten, beispielsweise:
 - [MODS](#)¹⁵⁴ (Metadata Object Description Schema) findet dabei Anwendung für umfangreiche beschreibende Metadaten,
 - das Schema [Dublin Core](#)¹⁵⁵ und das [DataCite Metadata Schema](#)¹⁵⁶ für einfache deskriptive Metadaten,
 - das Schema [TEI](#)¹⁵⁷ (Text Encoding Initiative) für Metadaten von Texteditionen,
 - [MEI](#)¹⁵⁸ (Music Encoding Initiative) für Metadaten musikalischer Werke,
 - [LIDO](#)¹⁵⁹ (Lightweight Information Describing Objects) für deskriptive und administrative Metadaten von 2D-Bildern, 3D-Digitalisaten und -Modellen und materiellen (nicht-digitalen) Objekten, und
 - [XMP](#)¹⁶⁰ (Extensible Metadata Platform) für technische Metadaten von 2D- oder 3D-Digitalisaten.

Schließlich sollten digitale Ressourcen und auch die in den Daten benannten nicht-digitalen Entitäten wie Personen, Körperschaften, Geografika, Werke, Begriffe etc. innerhalb der Metadaten über global eindeutige, dauerhaft gültige Identifikatoren, wie z. B. DOIs, Handles, ARKs oder URNs, verknüpft sein (→siehe Abschnitt [Persistente Identifikatoren](#)).

¹⁵⁰ <https://www.loc.gov/standards/premis/>

¹⁵¹ <https://www.loc.gov/standards/premis/v3/index.html>

¹⁵² <https://www.loc.gov/standards/mets/>

¹⁵³ <https://de.wikipedia.org/wiki/Containerformat>

¹⁵⁴ <https://www.loc.gov/standards/mods/>

¹⁵⁵ <https://www.dublincore.org/>

¹⁵⁶ <http://schema.datacite.org/>

¹⁵⁷ <https://tei-c.org/>

¹⁵⁸ <https://music-encoding.org/>

¹⁵⁹ <https://cidoc.mini.icom.museum/working-groups/lido/lido-overview/>

¹⁶⁰ <https://www.adobe.com/devnet/xmp.html>



Metadaten:

- NFDI4Culture-Handreichung: Kailus, A. (2023). *Guideline for a FAIR Cultural Studies Research Data Management* (1.0.3) [dataset]. Zenodo. <https://nfdi4culture.de/go/E3508>
- Altenhöner, R., Berger, A., Bracht, C., Klimpel, P., Meyer, S., Neuburger, A., Stäcker, T., & Stein, R. (2023). *DFG-Praxisregeln „Digitalisierung“. Aktualisierte Fassung 2022.* <https://doi.org/10.5281/ZENODO.7435724>
- Lindlar, M., Panitz, M., & Gadiraju, U. (2015). *Ingest and Storage of 3D Objects in a Digital Preservation System.* <https://doi.org/10.5281/zenodo.1115512>
- DURABLE ARchitectural Knowledge. (2015). *DURAARK Schemas for Ingestion and Storage of 3D Objects in a Digital Preservation System.* [Software]. <https://github.com/DURAARK/Schemas>

Normdaten:

- Gruß, M., & Mayer, D. (2023, Mai 25). *Was sind Normdaten?* <https://av.tib.eu/player/60986>

PREMIS:

- Caplan, P. (2021). *PREMIS verstehen (2021)* (T. Beinert, Übers.). Library of Congress Network Development and MARC Standards Office. https://www.loc.gov/standards/premis/understandingPREMIS_german_2021.pdf
- Lindlar, Micky & Rudnik, Pia. (2021). *Foliensatz: Crashkurs Digitale Langzeitarchivierung - Einführung in PREMIS.* <https://doi.org/10.5281/ZENODO.4761923>
- Rudnik, Pia. (2020). *Video: Crashkurs Digitale Langzeitarchivierung - Das Referenzmodell Open Archival Information System (OAIS).* <https://doi.org/10.5281/ZENODO.3984982>

Text Encoding Initiative (TEI):

- arthistoricum.net - Fachinformationsdienst Kunst. (2023). *TEI: Text Encoding Initiative.* <https://tei-c.org/>
- TEI Consortium. (2023). *Introducing the Guidelines – TEI: Text Encoding Initiative.* <https://tei-c.org/support/learn/introducing-the-guidelines/>

Music Encoding Initiative (MEI):

- Akademie der Wissenschaften und der Literatur. (2023). *Music Encoding Initiative.* <https://music-encoding.org/>
- Music Encoding Initiative. (2023). *MEI Tutorials.* <https://music-encoding.org/resources/tutorials.html>
- Music Encoding Initiative. (2023). *Guidelines.* <https://music-encoding.org/guidelines/v4/content/>

LIDO:

- ICOM CIDOC. (2023). *What is LIDO?* <https://lido-schema.org/>
- Knaus, G., Kailus, A., Stein, R., Deutsches Dokumentationszentrum für Kunstgeschichte – Bildarchiv Foto Marburg, & Bracht, C. (2022). *LIDO-Handbuch für die Erfassung und Publikation von Metadaten zu kulturellen Objekten*. arthistoricum.net. <https://doi.org/10.11588/ARTHISTORICUM.1026>
- Knaus, G., Stein, R., & Kailus, A. (2019). *LIDO-Handbuch für die Erfassung und Publikation von Metadaten zu kulturellen Objekten: Band 1: Graphik*. <https://doi.org/10.11588/ARTHISTORICUM.382.544>

SLUBRights:

- SLUB Dresden. (2019). *Schema zur Beschreibung von Rechteinformationen im SLUBArchiv (1.0) [XSD]*. SLUB Dresden. <https://slubarchiv.slub-dresden.de/fileadmin/groups/slubsite/slubarshiv/standards/rights/rights1.xsd>
- SLUB Dresden. (2020). *Rechteauszeichnung im SLUB Langzeitarchiv—Diskussionspapier*. SLUB Dresden. https://slubarchiv.slub-dresden.de/fileadmin/groups/slubsite/slubarchiv/SLUBArchiv_Diskussionspapier_Rechteauszeichnung_v1.1.3.pdf

Persistente Identifikatoren

Persistente Identifikatoren (Persistent Identifiers oder PIDs), wie z. B. DOIs, Handles, ARKs oder URNs, ermöglichen die dauerhafte Referenzierbarkeit von digitalen Objekten im Internet. Sie sind eine Art deskriptiver Metadaten (→siehe Abschnitt [Erfassen von Metadaten](#)) und werden von zentralen Institutionen oder Organisationen verwaltet. Sie spielen insbesondere bei der Online-Publikation von Forschungsdaten in der Wissenschaft sowie bei der Erstellung von Wissensgraphen (Datenbanken, die Informationen in Form von Entitäten und Beziehungen organisieren) eine zentrale Rolle. Sie sind ein wesentlicher Bestandteil des Semantic Web. Meist werden bei der Generierung von PIDs automatisch neben dem Ziellink weitere Metadaten (wie z. B. Titel, Autor, ISBN) hinterlegt.

Bei der digitalen Langzeitarchivierung hingegen sind Daten nicht der Dynamik des WWW ausgeliefert. Hier werden interne Identifikatoren vom digitalen Archiv vergeben, um eine präzise Zuordnung zu ermöglichen. Bei RADAR4Culture und SLUBArchiv.digital wird Material bei der Einlieferung mit einer ID versehen, mit der später Informationen über einen Webservice zum Stand und Zeitpunkt der Archivierung abgerufen werden können.

Idealerweise sollten Informationen zur digitalen Langzeitarchivierung auch mit den Forschungsdaten bei der Veröffentlichung und Katalogisierung bereitgestellt werden. Zu diesen „LZA-Metadaten“ sollten, ob, von welchem Archiv, unter welcher internen ID, bis wann sowie mit welcher Qualität diese Forschungsdaten dort archiviert sind, aber auch ggf. die Wichtig-

keit einzelner Bestandteile für die Nutzbarkeit der übergeordneten Publikation, gehören. Einerseits sollten Infrastrukturbetreibende wie Publikations- oder Katalogisierungsplattformen ein Erfassen dieser LZA-Metadaten ermöglichen. Andererseits können Forschende im Impressum ihrer Publikation diese LZA-Metadaten auch menschenlesbar angeben.

Zentrale Anlaufstelle zu PIDs in Deutschland ist das von der DFG geförderte Projekt „PID Network Deutschland – Netzwerk für die Förderung von persistenten Identifikatoren in Wissenschaft und Kultur“, gleichzeitig wird im Rahmen der NFDI seit 2023 der Aufbau eines Basisdienstes für Persistente Identifikatoren PID4NFDI gefördert.



Persistente Identifikatoren:

- Koster, L. (2020). Persistent identifiers for heritage objects. *The Code4Lib Journal*, 47. <https://journal.code4lib.org/articles/14978>  [Wayback](https://web.archive.org/web/20230602154447/https://journal.code4lib.org/articles/14978)¹⁶¹.
- *PID Network Deutschland*. (2023). <https://www.pid-network.de/>
- TIB Hannover. (2023). *Persistent Identifiers (PIDs)*. <https://projects.tib.eu/pid-service/persistent-identifiers/persistent-identifiers-pids/>
- Bodenschatz, E., & Fritzsche, F. (2022). *Forum Datenpublikation und -archivierung #5: Persistent Identifiers – Bericht*. <https://nfdi4culture.de/id/E3541>
- Helmholtz-Gemeinschaft. (2023). *PID4NFDI*. Helmholtz Open Science Office. <https://os.helmholtz.de/aktuelles/projekte/pid4nfdi/>



Metadatenerfassung und persistente Identifikatoren:

- Bereiten Sie **Metadaten** zu Ihren Forschungsdaten vor. Prüfen Sie, ob das Langzeitarchiv bereits **Vorlagen** dazu anbietet.
- Stellen Sie sicher, dass Verweise in Ihren Forschungsdaten möglichst **persistente Identifikatoren zu Normdaten** nutzen, anstatt einfacher Weblinks (URLs).

¹⁶¹ <https://web.archive.org/web/20230602154447/https://journal.code4lib.org/articles/14978>

3.6 Digitalisierung

Das Erstellen einer digitalen Aufnahme von einem analogen bzw. physischen Objekt wird Digitalisierung genannt. Eine digitale Repräsentation eines analogen Objekts geht dabei immer mit Informationsverlust einher. Daher ist das Erfassen von Informationen über den Erstellungsprozess des Digitalisats, aber auch über den Zustand des analogen Originals, unerlässlich (→siehe Abschnitt [Erfassen von Metadaten](#)). Außerdem können bestimmte Objekte unter Umständen nur einmal digitalisiert werden; wenn das Original sehr fragil ist oder sich in einem konservatorisch schlechten Zustand befindet (z. B. häufig bei analogen Filmen oder etwa mittelalterlichen Handschriften) oder nur ein einmaliger Zugriff auf das physische Original möglich ist (z. B. bei Ausstellungsstücken in Museen). Umso wichtiger ist es, bei der Planung der Digitalisierung die →[erhaltenswerten Eigenschaften](#) als Basis für die zukünftige Verfügbarkeit als digitales Objekt sorgfältig festzulegen.

Die Digitalisierung kann bei manchen Projekten auch in einer früheren Phase stehen, in manchen Fällen wird davon ausgegangen, dass bereits digitalisierte Objekte vorhanden sind.

Einen Spezialfall bilden sogenannte „born-digitals“. Dieser Begriff bezieht sich auf Inhalte oder Informationen, die von Anfang an in digitaler Form erstellt wurden, im Gegensatz zu solchen, die in analoger Form existieren und später digitalisiert werden. Beispiele für born-digital Inhalte sind E-Books, digitale Fotos, Videos, Musikdateien oder computergenerierte Kunstwerke. Besonders hier ist überlegenswert, wie viele Kopien man im Langzeitarchiv aufbewahren möchte.



Regeln bei der Digitalisierung:

- Altenhöner, R., Berger, A., Bracht, C., Klimpel, P., Meyer, S., Neuburger, A., Stäcker, T., & Stein, R. (2023). *DFG-Praxisregeln „Digitalisierung“*. Aktualisierte Fassung 2022. <https://doi.org/10.5281/ZENODO.7435724>
- Deutsches Institut für Normung (Ed.). (2023). DIN 33910, Information und Dokumentation-Objektschonende Digitalisierung von Archiv- und Bibliotheksgut (August 2023, p. 17). Beuth Verlag GmbH.

Metadaten zur Konservierung von analogem Film:

- iPRES2021 Organizing Committee. (2022). When Digital Remembers Analogue—Conservation Metadata for Analogue Film as Preservation Description Information in a Digital Archive. <https://doi.org/10.17605/OSF.IO/JT6XZ>
- TIB Hannover. (2020). *Conservation Metadata Data Dictionary*. <https://github.com/TIB-Digital-Preservation/FilmConservationMetadata>

3.7 Sonderfall Webarchivierung

Eine besondere Form der born-digital-Informationen sind Webseiten und Webinhalte. Nicht zuletzt durch die zunehmende Einbeziehung von Online-Quellen in die wissenschaftliche Forschung werden auch diese zu einem Teil der Forschungsdaten – oder sogar selbst zum Forschungsgegenstand – und müssen längerfristig erhalten werden. Jedoch unterliegt das Internet einer ständigen Veränderung: Inhalte von Webseiten werden geändert, umstrukturiert oder verschwinden ganz. Diese Phänomene werden als *content drift* (Änderung von Webinhalten), *link rot* (Verfall von Webseitenlinks) oder *reference rot* (Verfall von Zitierungen eines Webinhalts) bezeichnet. Sie betreffen das gesamte Spektrum der Online-Informationen, von einzelnen Websites, über Fachpublikationen (Coble & Karlin, 2023) bis hin zu Sammlungen von Rechtstexten (Zittrain et al., 2014). In einer Untersuchung von Artikeln aus den Bereichen Naturwissenschaft, Technik und Medizin wurde über den Zeitraum 1997 bis 2012 ein *link rot level* (Grad des Verfalls von Webseitenlinks) von etwa 70 % festgestellt (Klein et al. 2014).

Aufgabe der Webarchivierung ist es, sicherzustellen, dass Webinhalte auch in Zukunft erreichbar und lesbar sind. Das ermöglicht es, dauerhaft auf eine zitierte Quelle im Kontext zum Zeitpunkt der Zitation zu verweisen.

Während in dieser Handreichung bisher von einzelnen Medienarten wie Text, Bild, Video, oder 3D-Daten gesprochen wurde, für die es gut dokumentierte Lösungen zur digitalen Langzeitarchivierung gibt, ist zu beachten, dass es sich bei Webseiten auch um multimediale oder dynamisch erzeugte Objekte handeln kann. Die Langzeiterhaltung solcher interaktiven oder dynamisch generierten Webseiten ist oft nicht möglich. Das betrifft sowohl Frontends von Datenbanken, als auch nichtlineare Präsentationsformen wie interaktive Videos. Auch die Erhaltung von sozialen Medien wie Twitter / „X“, TikTok oder Diskussionsforen, Bulletin-Boards oder Chaträumen ist eine große Herausforderung.

Derzeit ist letztlich entscheidend, ob die Rohdaten der Inhalte extrahierbar sind und in archivfähige (statische) Dateiformate konvertiert werden können. So wird es bei einer interaktiven Textedition oder einem annotierten Video zunächst immer um die Archivierung der Texte (im [TEI-XML-Format](https://tei-c.org/release/doc/tei-p5-doc/en/html/)¹⁶²) oder der Videos (im [MP4-Format](https://www.loc.gov/preservation/digital/formats/fdd/fdd000155.shtml)¹⁶³) oder der Annotationen (z. B. im [Web Annotation Data Model Format](https://www.w3.org/TR/annotation-model/)¹⁶⁴ oder TEI-XML-Format) gehen. Die Weboberfläche, also das dynamische Webfrontend, wird unabhängig von konzeptionellem Aufwand oder Nutzerfreundlichkeit als nicht langzeitarchivfähig verstanden. Dynamisch generierte Inhalte

¹⁶² <https://tei-c.org/release/doc/tei-p5-doc/en/html/>

¹⁶³ <https://www.loc.gov/preservation/digital/formats/fdd/fdd000155.shtml>

¹⁶⁴ <https://www.w3.org/TR/annotation-model/>

oder nichtlineare, auf Nutzerinteraktion ausgerichtete Frontends können im Moment nur textlich beschrieben und mit Screenshots oder in Screencasts (d. h. kommentierte Bildschirmaufzeichnungen) linear dokumentiert werden. Mittelfristig sind Lösungen über eine Emulation, wie z. B. mit der [Emulation-as-a-Service Infrastructure \(EaaS\)](#)¹⁶⁵, vorstellbar.

Dennoch gibt es Möglichkeiten, Webinhalte zu archivieren. Vorreiter war das 1996 gegründete [Internet Archive](#)¹⁶⁶, das Webseiten automatisiert speichert und diese seit 2001 über die [Wayback Machine](#)¹⁶⁷ online verfügbar macht. Dafür wurde ein spezifisches Datenformat entwickelt, das „Web Archive Format“ („warc“, ISO 28500:2017). Dieses ermöglicht, verschiedene digitale Ressourcen und deren zugehörigen Metadaten in einer einzigen aggregierten Archivdatei zusammenzufassen. Das „warc“-Format ist inzwischen als Standard für Webarchivierung anerkannt.

Bei einer Webarchivierung wird eine Webseite von einem sogenannten „Web Crawler“ durchlaufen. Dieser prüft die Webseite auf Verweise auf externe und domainzugehörige Webinhalte sowie eingebettete Medien. Der Crawler stellt die gesammelten Informationen systematisch in einer „warc“-Datei oder einem komprimierten „wacz“-Paket zusammen. Das Crawlen kann voll- oder halbautomatisch erfolgen. Ein derzeit weit verbreiteter Crawler ist [Heritrix](#)¹⁶⁸, der vom Internet Archive gemeinsam mit den Nordic National Libraries entwickelt wurde. Ein anderes Tool, das das Aufzeichnen bestimmter Webseiten insbesondere für Endnutzer leichter zugänglich macht, ist der von Rhizome entwickelte [Webrecorder.io](#)¹⁶⁹, seit 2020 [Conifer](#)¹⁷⁰ genannt. Auch das Open-Source-Projekt [Webrecorder.net](#)¹⁷¹ arbeitet aktiv an der Entwicklung eines solchen Tools. Die Anzeige der Informationen aus den „warc“-Dateien erfolgt mit Viewern wie der Wayback Machine oder Open Source-Ablegern wie [Open Wayback](#)¹⁷². Mit Hilfsmitteln wie dem [Web Curator Tool](#)¹⁷³ oder Conifer ist es möglich, die Inhalte der „warc“-Dateien zu verwalten und die einzelnen Crawls zu steuern.

Für die Sicherung dynamischer Webinhalte gibt es mittlerweile eine ganze Reihe von Tools und Anwendungen, die aber neben den technischen Entwicklungen immer auch von den APIs der Plattformbetreiber oder auch deren Nutzungsbedingungen abhängig sind (wie vor allem das Beispiel Twitter / „X“ zeigt, woraus auch die Initiative des [DNB Twitter-Archivs](#)¹⁷⁴

¹⁶⁵ <https://www.softwarepreservationnetwork.org/emulation-as-a-service-infrastructure/>

¹⁶⁶ <https://archive.org/>

¹⁶⁷ <https://archive.org/web/>

¹⁶⁸ <https://github.com/internetarchive/heritrix3>

¹⁶⁹ <http://webrecorder.io/>

¹⁷⁰ <https://conifer.rhizome.org/>

¹⁷¹ <https://webrecorder.net/>

¹⁷² <https://github.com/iipc/openwayback>

¹⁷³ <https://webcuratortool.readthedocs.io/en/latest/guides/overview-history.html#introduction>

¹⁷⁴ https://www.dnb.de/DE/Professionell/Sammeln/Sammlung_Websites/twitterArchiv.html

 [Wayback](#)¹⁷⁵ entstanden ist). So finden sich Tools wie [Memespector](#)¹⁷⁶, [YouTube-Data-Tools](#)¹⁷⁷ und [4Cat](#)¹⁷⁸, die es erlauben, einen Großteil der Daten zu sichern. Bei jeder Sicherung von Webinhalten muss aber immer auch die Frage nach den →[erhaltenswerten Eigenschaften](#) gestellt werden, denn nur die für die jeweilige Forschungsfrage relevanten Daten sind (langzeit-)erhaltenswert.

Mittlerweile gibt es verschiedene, teils kommerzielle Initiativen, wie [Perma.cc](#)¹⁷⁹, die sich dem Archivieren von Webseiten widmen. Das an das Internet Archive angebundene [Archive-It](#)¹⁸⁰ archiviert für [eine Reihe von Institutionen und Projekten Webinhalte aus dem kulturellen Bereich](#)¹⁸¹, beispielsweise das [Collaborative ART Archive \(CARTA\)](#)¹⁸². Internationale Vereinigungen wie das [International Internet Preservation Consortium \(IIPC\)](#)¹⁸³ führen die Web Archiving Community zusammen, bieten [Fachinformationen](#)¹⁸⁴ an und tauschen sich beispielsweise auf der jährlich stattfindenden [Web Archiving Conference](#)¹⁸⁵ aus.

Im europäischen Kontext ist die schwierige Abklärung der Rechtslage der Inhalte, sowohl der Texte, als auch der eingeschlossenen Medien, ein wesentliches Hindernis für eine systematische Webarchivierung. Darüber hinaus kann für interaktive Webinhalte, die beispielsweise mit JavaScript gesteuert werden, von digitalen Langzeitarchiven nur der Erhalt der Datenintegrität und nicht der Nutzbarkeit garantiert werden. In Deutschland existieren eine Reihe von Webarchivierungs-Initiativen auf Bundes- ([Deutsche Nationalbibliothek – Webarchivierung](#)¹⁸⁶  [Wayback](#)¹⁸⁷) und Länderebene (Altenhöner, 2019), die zumeist im Rahmen des gesetzlichen Sammelauftrags der Bibliotheken und Archive agieren. Eine Einsicht in die archivierten Inhalte ist dabei aufgrund technischer und rechtlicher Beschränkungen oft nur an einem Computerterminal im Lesesaal vor Ort möglich.

¹⁷⁵ https://web.archive.org/web/20230625014937/https://www.dnb.de/DE/Professionell/Sammeln/Sammlung_Websites/twitterArchiv.html

¹⁷⁶ <https://github.com/jason-chao/memespector-gui>

¹⁷⁷ <https://github.com/bernorieder/YouTube-Data-Tools>

¹⁷⁸ <https://github.com/digitalmethodsinitiative/4cat>

¹⁷⁹ <https://perma.cc/>

¹⁸⁰ <https://archive-it.org/>

¹⁸¹ <https://archive-it.org/explore?fc=organizationType:museumsAndArtLibraries>

¹⁸² <https://communitywebs.archive-it.org/carta/>

¹⁸³ <https://netpreserve.org/>

¹⁸⁴ <https://netpreserve.org/web-archiving/>

¹⁸⁵ <https://netpreserve.org/events-archive/wac2023/>

¹⁸⁶ https://www.dnb.de/DE/Professionell/Sammeln/Sammlung_Websites/sammlung_websites_node.html

¹⁸⁷ https://web.archive.org/web/20230918153033/https://www.dnb.de/DE/Professionell/Sammeln/Sammlung_Websites/sammlung_websites_node.html

Für Webarchivierung im Kontext der NFDI4Culture muss festgestellt werden, dass bisher noch kein spezifisches Langzeitarchiv für Web-Ressourcen verfügbar ist. Es existieren jedoch einzelne Initiativen, wie beispielsweise [Saving Ukrainian Cultural Heritage Online](#)¹⁸⁸  [Wayback](#)¹⁸⁹, die Webinhalte themenorientiert sichern. Je nach Inhalt und Software (z. B. Telegram) muss beim Sichern von dynamischen Webinhalten sowohl auf skriptbasierte als auch auf manuelle Prozesse zurückgegriffen werden. Dabei sind die jeweils eingesetzten Datenbanken und Ablagesysteme oftmals sehr individuell angepasst und ihre Speicherung verlangt manuelle Eingriffe, wie das Beispiel des [ukrainischen „Telegram Archive of the War“](#)¹⁹⁰  [Wayback](#)¹⁹¹ zeigt. Neben dem Informationsangebot von [nestor](#)¹⁹² ist auch die von mehreren NFDIs ausgerichtete [Veranstaltungsreihe zu Social Media-Daten](#)¹⁹³ nennenswert. Unabhängig von der inhaltlichen Ausrichtung gibt es Vorschläge, wie Forschende Weblinks robuster machen können. Ein Vorschlag von Eld Zierau (Royal Danish Library) ist die Einführung eines PIDs für Webressourcen – ein [Persistent Web Identifier \(PWID\)](#)¹⁹⁴. Dieser muss jedoch erst zu einem Standard und breit akzeptiert werden. (Jones et al. 2021) schlagen vor, dass die Forschenden sich selbst aktiv um die Sicherung der von ihnen zitierten Webseiten in Webarchiven wie dem Internet Archive sorgen. Zusätzlich sollte in der Zitation neben der originalen URL auch die der archivierten Seite angegeben werden. Diesen Ansatz unterstützt auch diese Handreichung, indem bei potentiell instabilen Webressourcen zusätzlich ein Link auf die archivierte Seite in der Wayback Machine angeboten wird. Dabei werden Internetressourcen als (potentiell) stabil angesehen, wenn sie zu großen institutionalisierten Publikationsplattformen gehören (wie z. B. Zenodo), über persistente Identifier (DOI, Handle, etc.) angesprochen werden können oder zu etablierten digitalen Zeitschriften gehören (z. B. [DHQ](#)¹⁹⁵, [Code4Lib](#)¹⁹⁶, etc.).



Problematik vergänglicher Webressourcen:

- Klein, M., Sompel, H. V. de, Sanderson, R., Shankar, H., Balakireva, L., Zhou, K., & Tobin, R. (2014). Scholarly Context Not Found: One in Five Articles Suffers from Reference Rot. *PLOS ONE*, 9(12), e115253. <https://doi.org/10.1371/journal.pone.0115253>

¹⁸⁸ <http://www.sucho.org/>

¹⁸⁹ <https://web.archive.org/web/20230920222641/https://www.sucho.org/>

¹⁹⁰ <https://storymaps.arcgis.com/stories/0af72de4b008461bb441fc62fffb9f8d>

¹⁹¹ <https://web.archive.org/web/20230207160525/https://storymaps.arcgis.com/stories/0af72de4b008461bb441fc62fffb9f8d>

¹⁹² https://www.langzeitarchivierung.de/Webs/nestor/DE/Veranstaltungen_und_Termine/2023Webarchivierung.html

¹⁹³ <https://nfdi4culture.de/services/details/show-tell-social-media-data-in-research-practice.html>

¹⁹⁴ <https://digital.library.unt.edu/ark:/67531/metadc2143884/m1/1/>

¹⁹⁵ <http://www.digitalhumanities.org/dhq/>

¹⁹⁶ <https://journal.code4lib.org/>

- Zittrain, J., Albert, K., & Lessig, L. (2014). Perma: Scoping and Addressing the Problem of Link and Reference Rot in Legal Citations. *Legal Information Management*, 14(2), 88–99. <https://doi.org/10.1017/S1472669614000255>
- Coble, Z., & Karlin, J. (2023). Reference Rot in the Digital Humanities Literature: An Analysis of Citations Containing Website Links in DHQ. *Digital Humanities Quarterly*, 017(1). <http://www.digitalhumanities.org/dhq/vol/17/1/000662/000662.html>
- Beinert, T., Schmid, K., & Weimer, K. (2022). Infrastrukturen und Services für die wissenschaftliche Nutzung von Webarchiven: Ein Überblick. *o-bib. Das offene Bibliotheksjournal / Herausgeber VDB*, 9(3), Article 3. <https://doi.org/10.5282/o-bib/5821>
- Jones, S. M., Klein, M., & Sompel, H. V. de. (2021). Robustifying Links To Combat Reference Rot. *The Code4Lib Journal*, 50. <https://journal.code4lib.org/articles/15509>

Übersicht Anleitungen und Tools zur Webarchivierung:

- International Internet Preservation Consortium. (2023). *Awesome Web Archiving*. <https://github.com/iipc/awesome-web-archiving>

Weiterführende Informationen zum WARC-Format:

- International Internet Preservation Consortium. (2020, Januar 23). *The WARC Format*. <https://iipc.github.io/warc-specifications/specifications/warc-format/warc-1.1/>



Langzeitarchivierung von Social Media:

- Vorlesungsreihe zu Archivierungsbemühungen dynamischer Inhalte. <https://nfdi4culture.de/de/nachrichten/show-tell-social-media-daten-in-der-forschungspraxis-iii-sommer-2023.html>

Webarchivierung in Deutschland:

- Altenhöner, R. (2019). Noch immer am Anfang? Stand und Perspektiven der Webarchivierung in Deutschland 2019. In Simone Fühles-Ubach & Ursula Georgy (Hg.): *Bibliotheksentwicklung im Netzwerk von Menschen, Informationstechnologie und Nachhaltigkeit*. Festschrift für Achim Oßwald (S. 237–250). Bock und Herchen Verlag. https://publiccologne.th-koeln.de/frontdoor/deliver/index/docId/1623/file/Festschrift_Osswald.pdf



Webarchive in Deutschland (Auswahl):

- Deutsche Nationalbibliothek. (2023, September 14). *Webarchivierung*. https://www.dnb.de/DE/Professionell/Sammeln/Sammlung_Websites/sammlung_websites_node.html
- Bayerische Staatsbibliothek. (2020). *Langzeitarchivierung und Data Curation*. <https://www.digitale-sammlungen.de/de/langzeitarchivierung-und-data-curation>
- musiconn - Für vernetzte Musikwissenschaft. (2023). *Wunschbuch: Webinhalt*. <https://www.musiconn.de/services/wunschbuch3>

- nestor. (2023, August 24). *Webarchivierung – Praxis und Perspektiven [Workshop]*. nestor. https://www.langzeitarchivierung.de/Webs/nestor/DE/Veranstaltungen_und_Termine/2023Webarchivierung.html



Verweise auf Webinhalte:

- Prüfen Sie Ihre Verweise auf Internetquellen. Sind diese für Ihre Daten wesentlich? Identifizieren Sie **weniger stabile Ressourcen**, also solche, die z. B. NICHT bei Zenodo liegen oder NICHT über DOIs referenziert sind.
- Stellen Sie sicher, dass die für Sie wesentlichen Internetquellen in dem von Ihnen zitierten Zustand **in Webarchiven verfügbar** sind, z. B. im Internet Archive über die Wayback Machine.
- Fügen Sie in Ihren Texten **zusätzlich die Links zum Webarchiv** (Wayback Machine) mit ein.
- Wenn wesentliche Teile Ihrer Forschungsdaten aus dem Internet stammen, prüfen Sie, ob diese zusätzlich zu Ihrer **Publikation als Datensatz** veröffentlicht werden können.

4 Forschungsdaten im digitalen Langzeitarchiv

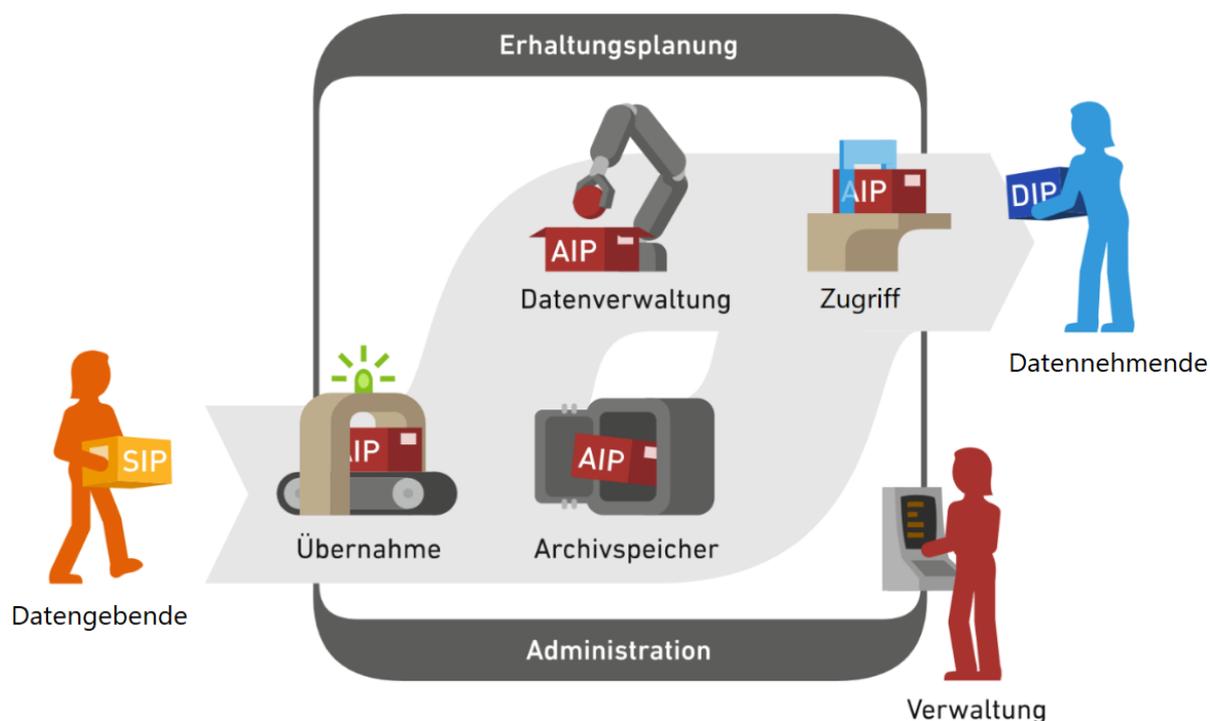
4.1 Das OAIS-Modell

Nachdem der Weg von Materialien in ein digitales Archiv von der Bestimmung der Archivwürdigkeit bis zur Vorbereitung der Einlieferung ins Archiv beschrieben wurde, werden nun die Prozesse genauer betrachtet, die INNERHALB des Langzeitarchivs ablaufen. Zu diesen Prozessen gehören neben der Einlieferung (*submission*) auch die Übernahme (*ingest*), die Datenverwaltung (*data management*), die Speicherung im Archiv (*archival storage*) und die Auslieferung (*dissemination*). Diese Prozesse werden durch sogenannte →[Archivinformationssysteme](#) umgesetzt.

Das 1999 von der NASA zur Archivierung anfangs noch eigener Weltraumaufnahmen entwickelte Referenzmodell [Open Archival Information System \(OAIS\)](#)¹⁹⁷ hat sich über die Zeit als Standard zur Beschreibung der Prozesse eines Langzeitarchivs einschließlich seiner Aufgaben und Schnittstellen etabliert.

Die folgende Abbildung gibt einen Überblick:

¹⁹⁷ <http://www.oais.info/>



Informationspakete (IP) im OAIS Modell



Abbildung 6: Das OAIS-Referenzmodell in vereinfachter Form. (Grafik: © [Andres Bucher](https://andresbucher.ch/cc/library.html)¹⁹⁸
[Wayback](https://web.archive.org/web/20221101103016/https://andresbucher.ch/cc/library.html)¹⁹⁹, [CC BY 4.0](http://creativecommons.org/licenses/by/4.0/)²⁰⁰)

Auf der einen Seite stehen die sogenannten **Datengebenden** (auch Produzenten bzw. Produzierende genannt), die Datenpakete erstellen und in das Archiv geben. Dies können beispielsweise eine Organisation oder eine Person sein, die digitale Objekte archivieren möchten.

Auf der anderen Seite stehen die **Datennehmenden** (auch Konsumenten bzw. Konsumierende oder Nutzer:innen bzw. Endnutzer:innen genannt), die auf Datenpakete zugreifen. Dies können Forscher, Wissenschaftler oder andere Interessenten sein.

Die **Verwaltung** ist dabei für die Organisation des Archivs zuständig. Dazu gehören die Festlegung von Regeln für die Auswahl, Übernahme, Speicherung sowie Zugriff auf die digitalen Objekte, ferner die Überwachung der Integrität gespeicherter Daten, die Planungen für Formatmigration und Emulation sowie schließlich die Zuweisung von Ressourcen, wie finanzieller Mittel, Personal oder der technischen Infrastruktur.

Je nach Art des digitalen Langzeitarchivs werden unterschiedliche Erhaltungsmaßnahmen (→siehe Abschnitt [Content Preservation](#)) vorgenommen.

¹⁹⁸ <https://andresbucher.ch/cc/library.html>

¹⁹⁹ <https://web.archive.org/web/20221101103016/https://andresbucher.ch/cc/library.html>

²⁰⁰ <http://creativecommons.org/licenses/by/4.0/>



Grundsätzlich bewahren zwar alle digitalen Langzeitarchive Daten über längere Zeiträume auf, Unterschiede liegen jedoch in den **internen Verarbeitungsabläufen und Strategien**. In dieser Handreichung werden daher durchweg die Verarbeitungsprozesse von eingereichtem Archivmaterial anhand des OAIS-Modells mit Beispielen aus dem NFDI4Culture-Kontext erklärt.



OAIS-Modell:

- CCSDS Secretariat. (2012). *Recommendation for Space Data System Practices—Reference Model for an Open Archival Information System (OAIS)*. <https://public.ccsds.org/Pubs/650x0m2.pdf>  [Wayback²⁰¹](https://web.archive.org/web/20231123053249/https://public.ccsds.org/Pubs/650x0m2.pdf)
- nestor-Arbeitsgruppe. (2013, August 27). *Referenzmodell für ein Offenes Archiv-Informationssystem—Deutsche Übersetzung 2.0*. <http://d-nb.info/104761314X/34>
- Rudnik, Pia. (2020). *Video: Crashkurs Digitale Langzeitarchivierung - Das Referenzmodell Open Archival Information System (OAIS)*. <https://doi.org/10.5281/ZENODO.3984982>
- Schrimpf, S. (2014). *Das OAIS-Modell für die Langzeitarchivierung: Anwendung der ISO 14721 in Bibliotheken und Archiven*. Beuth.

Erhaltungsplanung

Die **Erhaltungsplanung** (auch „Preservation Planning“ genannt) befasst sich mit der Ermittlung geeigneter Erhaltungsmaßnahmen (→siehe Kapitel [5 Erhaltungsstrategien](#)) und betrifft, wie in der oberen Abbildung veranschaulicht, die Prozesse Übernahme, Archivspeicherung, Zugriff sowie Datenverwaltung.

Ausgangspunkt ist die Risikobewertung der Archivpakete. Diese kann auf zweierlei Arten vorgenommen werden: Bei der **Technology Watch** werden technologische Veränderungen in Bezug auf Dateiformate, Speichermedien und Archivsoftware systematisch beobachtet. Bei der **Community Watch** werden Anforderungen und Erwartungen der Zielgruppen beleuchtet, die sich im Laufe der Zeit auch verändern können. Technology und Community Watch sind komplementär zueinander. Beide können durch Umfragen und Beobachtung der Zielgruppe, der Teilnahme an einschlägigen Konferenzen und Workshops oder mithilfe entsprechender Webseiten umgesetzt werden.

Nach der Identifizierung von Risiken und der Ableitung von Maßnahmen wird ein Erhaltungsplan für die archivierten Daten entwickelt und durchgeführt. Diesen Vorgang wiederholt das Archiv zyklisch über einen gewissen Zeitraum hinweg und dokumentiert die Änderungen im Archiv.

²⁰¹ <https://web.archive.org/web/20231123053249/https://public.ccsds.org/Pubs/650x0m2.pdf>

Erstellung von Informationspaketen

Bei der **Abgabe** von Daten ins Archiv erstellen die Datengebenden ein Informationspaket – das **Submission Information Package** oder kurz **SIP**.

Es beinhaltet die Forschungsdaten und seine Metadaten. Die Struktur eines SIP ist vom Archiv vorgegeben, wobei das [BagIt-Format](#)²⁰²  [Wayback](#)²⁰³ weit verbreitet ist.

Intellektuelle Einheiten

Zusammenhängende Forschungsdaten und Metadaten werden als „intellektuelle Einheit“ (kurz „IE“) zusammengefasst. Beispielsweise bildet ein einzelner Zeitschriftenartikel mit seinen Metadaten eine intellektuelle Einheit. Archive wie das [SLUBArchiv.digital](#)²⁰⁴ bilden dabei eine IE jeweils in einem SIP ab. Jeder Zeitschriftenartikel würde hier als einzelnes SIP eingeliefert werden.

Erstellung von SIPs

Die Erstellung eines SIP kann durch die Datengebenden selbst erfolgen. Dazu werden von den Archiven verschiedene Tools bereitgestellt.

Bei [RADAR4Culture](#)²⁰⁵ und [heiARCHIVE](#)²⁰⁶ erfolgt die Erstellung von SIPs bspw. automatisch im Backend, sobald Nutzer:innen über ein Webseiten-Interface die Publikation mit impliziter Archivierung veranlassen. Das [SLUBArchiv.digital](#)²⁰⁷ stellt Datengebenden hingegen den [SLUB SIP Builder](#)²⁰⁸ bereit, ein Kommandozeilen-Tool, mit dem sie SIPs nach Vorgaben des SLUBArchiv.digital erstellen und im Anschluss über einen SFTP-Zugang einliefern können.

Unabhängig von den zuvor genannten Angeboten ist es auch möglich, einen Schritt in den Digitalisierungs-Workflow einzubauen, der automatisch ein SIP erzeugt, sobald die Digitalisierung und Strukturierung z. B. eines Buches abgeschlossen ist, und dieses dann ins Archiv sowie in die Präsentation fließen lässt. Es ist auch möglich, eine OAI-Schnittstelle im digitalen Archiv einzurichten, die standardisierte Anknüpfungspunkte bereitstellt. Damit können Nutzer:innen über den OAI-Harvester auf digitale Objekte zugreifen, ein SIP bauen, Metadaten zum digitalen Objekt ins SIP aufnehmen und dieses schließlich ans Archiv weitergeben.

²⁰² <https://dallibraries.atlassian.net/wiki/spaces/GLOS/pages/931725313/BagIt+File+Packaging+Format>

²⁰³ <https://web.archive.org/web/20230221164744/https://dallibraries.atlassian.net/wiki/spaces/GLOS/pages/931725313/BagIt+File+Packaging+Format>

²⁰⁴ <https://slubarchiv.slub-dresden.de/>

²⁰⁵ <https://www.radar-service.eu/>

²⁰⁶ <https://heiarhive.uni-heidelberg.de/>

²⁰⁷ <https://slubarchiv.slub-dresden.de/>

²⁰⁸ https://git.slub-dresden.de/digital-preservation/SLUB_SIP_Builder/-/tree/master



Infrastruktureinrichtungen, die eine Anbindung an die digitale Langzeitarchivierung erreichen möchten, sollten bereits bei der Wahl der Repository-Software darauf achten, dass diese nativ über OAI-Schnittstellen verfügt.

Einreichung beim Archiv

Wenn SIPs an das Archiv abgegeben werden, durchlaufen sie den Prozess der **Datenübernahme**, dem sogenannten Ingest. Dieser umfasst die Prüfung auf Vollständigkeit und Unversehrtheit anhand mitgelieferter Prüfsummen und Abgleich von Verzeichnisstrukturvorgaben des Archivs, die →[Datenformaterkennung](#), die →[Datenformatvalidierung](#), die Anreicherung mit technischen und administrative →[Metadaten](#), ggf. eine →[Formatkorrektur](#) sowie schließlich die →Umwandlung in Informationspakete.

Formaterkennung

Während der Datenübernahme müssen zunächst die Datenformate identifiziert werden.

Tools wie z. B. [DROID](#)²⁰⁹  [Wayback](#)²¹⁰ dienen der automatisierten Formaterkennung und sind meist in Archivinformationssystemen integriert.

Das PDF-Format umfasst bspw. eine Vielzahl an Subtypen und Versionen. So existiert eine Norm für die Langzeitarchivierung: das [PDF/A Format](#)²¹¹  [Wayback](#)²¹². Dieses Format umfasst wiederum diverse Subtypen. Liegt eine Datei mit der Endung „.pdf“ vor, lässt dies zunächst keine Aussage über die tatsächliche PDF-Version zu. Deswegen erfolgt eine automatisierte Formaterkennung über Signaturen, die im Bitstream innerhalb einer Datei enthalten sind.

Digitale Formatsignaturen entsprechen standardisierten IDs, die in Formatdatenbanken wie z. B. [PRONOM](#)²¹³  [Wayback](#)²¹⁴ dokumentiert sind. Dank standardisierter Signaturen können Archive genaue Aussagen darüber treffen, welche konkreten Formate und Format-Subtypen bei der Datenübernahme erfasst bzw. bereits in den Speichersystemen vorgehalten werden.

²⁰⁹ <https://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/>

²¹⁰ <https://web.archive.org/web/20231109214026/https://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/>

²¹¹ <https://www.pdflib.com/de/pdf-know-how/pdfa/die-pdf-a-standards/>

²¹² <https://web.archive.org/web/20230818131914/https://www.pdflib.com/de/pdf-know-how/pdfa/die-pdf-a-standards/>

²¹³ <https://www.nationalarchives.gov.uk/pronom/BasicSearch/proBasicSearch.aspx?status=new>

²¹⁴ <https://web.archive.org/web/20230511111748/https://www.nationalarchives.gov.uk/pronom/BasicSearch/proBasicSearch.aspx?status=new>



Tools zur Formaterkennung:

- The National Archives. (2023, März 30). Download DROID: file format identification tool [Text]. *The National Archives*. <https://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/>  [Wayback](https://web.archive.org/web/20230330120000/https://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/)²¹⁵
- Lehane, R. (2023, April 25). *Siegfried*. IT for Archivists. <https://www.itforarchivists.com/siegfried/>  [Wayback](https://web.archive.org/web/20230425120000/https://www.itforarchivists.com/siegfried/)²¹⁶

Formatvalidierung

Ist das Datenformat bekannt, wird die Datei gegen ihre jeweilige Formatspezifikation validiert, um sicherzustellen, dass die eingelieferten Daten auch tatsächlich deren Spezifikationen entsprechen. Nur dann kann der Erhalt der Datei und die erfolgreiche Durchführung von Erhaltungsmaßnahmen sichergestellt werden.

Bei PDF-Dateien lässt sich die Gültigkeit der Formatspezifikation z. B. mit den Validierungstools [JHOVE](https://www.jhove.org/)²¹⁷ oder [veraPDF](https://www.verapdf.org/)²¹⁸ prüfen. Bei TIFF-Dateien kann dies mit den Tools [checkit tiff](http://andreas-romeike.de/software.html#_checkit_tiff)²¹⁹  [Wayback](https://web.archive.org/web/20230528073250/http://andreas-romeike.de/software.html#_checkit_tiff)²²⁰ oder [DPF Manager](http://www.preforma-project.eu/dpf-manager.html)²²¹  [Wayback](https://web.archive.org/web/20230418172243/http://www.preforma-project.eu/dpf-manager.html)²²² geschehen.

In der folgenden Abbildung wird ein fehlerhafter HTML-Quellcode dargestellt: Zum einen ist das Element `<center>...</center>` nicht korrekt geschlossen, und zum anderen muss das schließende `</center>`-Tag vor dem schließenden `</h1>`-Tag stehen, denn eine Tag-Überschneidung ist in HTML nicht gestattet. Zwar kann der dargestellte HTML-Code mit aktuellen Browsern korrekt gerendert werden, die darin befindlichen Syntaxfehler könnten jedoch dazu führen, dass mit zukünftigen Browsern dieser HTML-Inhalt nicht mehr korrekt oder überhaupt nicht mehr dargestellt wird.

```
<html>
  <h1><center>Dieses HTML-Dokument ist nicht regelkonform,
    wird aber u.U. wie gewünscht dargestellt.</h1><center>
</html>
```

²¹⁵ <https://web.archive.org/web/20231109214026/https://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/>

²¹⁶ <https://web.archive.org/web/20231027015139/https://www.itforarchivists.com/siegfried/>

²¹⁷ <http://jhove.openpreservation.org/>

²¹⁸ <https://verapdf.org/>

²¹⁹ http://andreas-romeike.de/software.html#_checkit_tiff

²²⁰ https://web.archive.org/web/20230528073250/http://andreas-romeike.de/software.html#_checkit_tiff

²²¹ <http://www.preforma-project.eu/dpf-manager.html>

²²² <https://web.archive.org/web/20230418172243/http://www.preforma-project.eu/dpf-manager.html>

Abbildung 7: Beispiel einer fehlerhaften HTML-Datei.



Tools zur Formatvalidierung:

- JHOVE für PDF, PNG, WAV, XML:
Open Preservation Foundation. (2023, Oktober). *JHOVE | JSTOR/Harvard Object Validation Environment*. <http://jhove.openpreservation.org/>
- Bad Peggy für JPEG:
Hahn, M. (2022, August 24). *Coder's Lagoon*. <https://coderslagoon.com/>
- ImageMagick für GIF:
LLC, I. S. (2023, August). *ImageMagick*. <https://imagemagick.org/>
- checkit_tiff für TIFF:
Romeyke, A. (2023, Juni 20). *Checkit_tiff*. <https://andreas-romeike.de/software.html>  [Wayback](https://web.archive.org/web/20230528073250/https://andreas-romeike.de/software.html)²²³
- jpylyzer für JP2000:
National Library of the Netherlands & Open Preservation Foundation. (2022, Oktober 3). *Jpylyzer*. <https://jpylyzer.openpreservation.org/>
- X3D-Edit für X3D-Validierung:
Brutzman, D. (2023, August 2). *X3D-Edit 4.0 Authoring Tool for Extensible 3D (X3D) Graphics*. <https://savage.nps.edu/X3D-Edit/X3D-Edit.html>  [Wayback](https://web.archive.org/web/20230401230719/https://savage.nps.edu/X3D-Edit/X3D-Edit.html)²²⁴
- glTF-Validator für glTF:
The Khronos Group Inc. (2019). *glTF Validator*. Coder's Lagoon. <https://github.khronos.org/glTF-Validator/>  [Wayback](https://web.archive.org/web/20230505033232/https://github.khronos.org/glTF-Validator/)²²⁵
- veraPDF für PDF:
veraPDF Consortium. (2023). *VeraPDF*. <https://verapdf.org/software/>  [Wayback](https://web.archive.org/web/20230406004922/https://verapdf.org/software/)²²⁶
- pdfaPilot für PDF/A (kostenpflichtig):
callas software GmbH. (2023). *PdfaPilot*. Callas Software. <https://www.callassoftware.com/en/products/pdfapilot>  [Wayback](https://web.archive.org/web/20231103052656/https://www.callassoftware.com/en/products/pdfapilot)²²⁷
- XMLSpy für XML (kostenpflichtig):
Altova®. (2023, August 10). *XML Editor*. <https://www.altova.com/xml-spy-xml-editor/download>  [Wayback](https://web.archive.org/web/20230927092724/https://www.altova.com/xml-spy-xml-editor/download)²²⁸
- Oxygen XML Editor für XML (kostenpflichtig):
Syncro Soft. (2023). *Oxygen XML Editor*. The Complete Solution for XML Authoring, Development and Collaboration. <https://www.oxygen-xml.com/>

²²³ <https://web.archive.org/web/20230528073250/https://andreas-romeike.de/software.html>

²²⁴ <https://web.archive.org/web/20230401230719/https://savage.nps.edu/X3D-Edit/X3D-Edit.html>

²²⁵ <https://web.archive.org/web/20230505033232/https://github.khronos.org/glTF-Validator/>

²²⁶ <https://web.archive.org/web/20230406004922/https://verapdf.org/software/>

²²⁷ <https://web.archive.org/web/20231103052656/https://www.callassoftware.com/en/products/pdfapilot>

²²⁸ <https://web.archive.org/web/20230927092724/https://www.altova.com/xml-spy-xml-editor/download>

Extraktion technischer Metadaten

Im Anschluss erfolgt die Extraktion technischer Metadaten, die für die Archivierung und die Risikobewertung relevant sind. Wie zuvor am Beispiel von PDF-Dateien gezeigt, können Dateiformate in verschiedenen Subversionen vorliegen, die allein an der Dateiendung nicht erkannt werden können. Eine über einen Scanner erstellte TIFF-Bilddatei kann intern komprimiert oder unkomprimiert sein. Diese Information wird in den technischen Metadaten vermerkt. Da komprimierte Daten in Archiven häufig vermieden werden, fielen eine komprimierte TIFF-Datei nach der Extraktion der technischen Metadaten auf. Sollte sie nicht abgelehnt werden, fielen zumindest ihre Risikobewertung schlechter aus. Mit dem Wissen über die Ausprägung der Formate, die sich im Archiv befinden, kann das Archiv Metadaten für eine ganz spezifische Untermenge an Daten extrahieren und gezielte Erhaltungsmaßnahmen durchführen.

Formatkorrektur

Einige Archive, wie das SLUBArchiv, halten Datengebende zu Korrekturen ihrer Dateien an, bevor sie diese ins Archiv übernehmen. Ziel ist dabei, Abweichungen einer Datei gegenüber ihrer Formatspezifikation zu bereinigen, so dass die Datei von einem Validierungstool akzeptiert wird. So können etwa Fehler von TIFF-Bilddateien mit dem Tool [fixit tiff](#)²²⁹  [Wayback](#)²³⁰ oder von PDF/A-Dateien mit dem Tool [PDF/A-Pilot](#)²³¹  [Wayback](#)²³² korrigiert werden.



Formatwissensdatenbanken:

- Archives, T. N. (2006, März 23). *PRONOM | Simple Search*. <https://www.nationalarchives.gov.uk/PRONOM/BasicSearch/pro-BasicSearch.aspx?status=new>  [Wayback](#)²³³
- Library of Congress. (2023, August 21). *Alphabetical List of Format Descriptions—Sustainability of Digital Formats | Library of Congress* [Webpage]. Abgerufen 4. August 2023, von https://www.loc.gov/preservation/digital/formats/fdd/browse_list.shtml  [Wayback](#)²³⁴

²²⁹ http://andreas-romecke.de/software.html#_fixit_tiff

²³⁰ https://web.archive.org/web/20230528073250/http://andreas-romecke.de/software.html#_fixit_tiff

²³¹ <https://www.callassoftware.com/de/produkte/pdfapilot>

²³² <https://web.archive.org/web/20230407030146/https://www.callassoftware.com/de/produkte/pdfapilot>

²³³ <https://web.archive.org/web/20230511111748/https://www.nationalarchives.gov.uk/PRONOM/BasicSearch/proBasicSearch.aspx?status=new>

²³⁴ https://web.archive.org/web/20231123175644/https://www.loc.gov/preservation/digital/formats/fdd/browse_list.shtml

Format-Tool-Datenbanken:

- COPTR. (2023, August 9). *Tools Grid—COPTR*. https://coptr.digipres.org/index.php/Tools_Grid  [Wayback](https://web.archive.org/web/20231126215440/https://coptr.digipres.org/index.php/Tools_Grid)²³⁵
- nestor – Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen in Deutschland e. V. (2023, August 10). *nestor—Nestor-Toolbox*. https://www.langzeitarchivierung.de/Webs/nestor/DE/Services/nestor_toolbox/nestor_toolbox_node.html  [Wayback](https://web.archive.org/web/20230529234857/https://www.langzeitarchivierung.de/Webs/nestor/DE/Services/nestor_toolbox/nestor_toolbox_node.html)²³⁶

Speicherung im Archiv

Nachdem die SIPs ins Archiv übernommen worden sind, werden sie zu **Archivinformationspaketen**, kurz **AIPs**, umgewandelt. Diese Archivpakete enthalten Informationen über die Prozesse, die sie durchlaufen haben. Im Fall des SLUBArchiv.digital lässt sich so nachvollziehen, welche Erhaltungsmaßnahmen für ein bestimmtes Archivpaket ergriffen wurden.

Der **Archivspeicher**, auch Permanentspeicher genannt, sorgt für die physische Erhaltung der AIPs. Dazu werden AIPs georedundant an mindestens zwei **Standorten** gespeichert. Außerdem werden regelmäßige Prüfungen der Datenintegrität durchgeführt. Hierzu werden **Prüfsummen** anhand des Bitstreams immer wieder neu berechnet und mit den gespeicherten Prüfsummen verglichen. Sind die Prüfsummen verschieden, liegt ein Datenfehler vor und es muss eine Datenwiederherstellung eingeleitet werden.



²³⁵ https://web.archive.org/web/20231126215440/https://coptr.digipres.org/index.php/Tools_Grid

²³⁶ https://web.archive.org/web/20230529234857/https://www.langzeitarchivierung.de/Webs/nestor/DE/Services/nestor_toolbox/nestor_toolbox_node.html

Abbildung 8: Eine Magnetbandkassette, ein Magnetbandlaufwerk sowie eine Tape Library.

Üblicherweise werden Archivdaten auf **Magnetbändern** (siehe Abbildung oben), seltener auf Festplatten gespeichert, da erstere deutlich kostengünstiger sind.

RADAR4Culture und heiARCHIVE bewahren dabei alle Daten einer →[intellektuellen Einheit](#) als gepackte „tar“-Datei, das SLUBArchiv.digital bewahrt seine Daten stets ungepackt auf Magnetbändern auf.

Um dem Verfall der Datenträger entgegenzuwirken, werden alle Daten im Archiv regelmäßig umkopiert. In großen Archiven werden diese mechanischen Arbeiten von speziellen Tape-Robotern übernommen.

Für eine schnelle **Präsentation** der Daten für Endnutzer:innen werden oft zusätzlich datenreduzierte Versionen erzeugt, die separat auf Festplatten gespeichert und abgerufen werden, da der Zugriff auf Daten auf Magnetbändern mehr Zeit erfordert.

Das Archiv ist auch in der Lage, **Updates an AIPs** entgegenzunehmen. Wurden bspw. Metadaten verändert oder korrigiert, wird ein spezielles SIP ins Archiv eingeliefert. Ob ein komplett neues Paket eingeliefert werden muss oder nur eines, das die veränderten Daten enthält (sog. „Delta-Paket“), hängt vom jeweiligen Archivinformationssystem ab. Ungeachtet dessen werden Änderungen als neues AIP hinzugefügt, das wiederum interne Informationen zur Fehlerkorrektur enthält.

Fallbeispiel: Umsetzung eines AIP-Updates

Eine digitalisierte mittelalterliche Handschrift wird als AIP im digitalen Langzeitarchiv gespeichert. Später wird festgestellt, dass ein weiterer Bestandteil zum Objekt hinzugefügt werden muss, z. B. weil bei einem Buchscan eine Seite überblättert wurde oder aus einer Bestandsrückführung ein fehlendes Teil ergänzt werden konnte. Mit dem AIP-Update (auch „Ergänzungslieferung“ genannt) wird das AIP aktualisiert. Die vorherige Version des AIP bleibt dabei erhalten.



Einlieferung ins Archiv:

- Boyko, A., & Kunze, J. (2012). *The BagIt File Packaging Format (V0.97)*. <https://www.digitalpreservation.gov/documents/bagitspec.pdf> 
[Wayback](https://web.archive.org/web/20231121213005/https://www.digitalpreservation.gov/documents/bagitspec.pdf)²³⁷
- SLUB Dresden. (2021). *SLUB SIP Builder* [Software]. https://git.slub-dresden.de/digital-preservation/SLUB_SIP_Builder

²³⁷ <https://web.archive.org/web/20231121213005/https://www.digitalpreservation.gov/documents/bagitspec.pdf>

- TIB Hannover. (2023). CSV-Ingest generisch [Software]. https://github.com/TIB-Digital-Preservation/CSV-Ingest_generisch  [Wayback](#)²³⁸
- TIB Hannover. (2023). Pre-Ingest-Analyzer [Software]. <https://github.com/TIB-Digital-Preservation/pre-ingest-analyzer>  [Wayback](#)²³⁹

Abrufen von Arbeitspaketen

Beim **Zugriff** auf archivierte Daten, auch Access genannt, wird auf Anfrage von Datennehmenden ein Original bzw. ein Archivpaket aus dem Permanentenspeicher ausgelesen und weitergegeben. Dabei muss der komplette Versionsverlauf bis hin zum Original bzw. der Ersteinlieferung zur Verfügung gestellt werden können. Das digitale Archiv hat nicht nur Zugriff auf alle jemals von dem AIP erzeugten Versionen, sondern kann diese auch auf Wunsch wieder ausgeben. Diese Fähigkeit kann auch als Grundlage für die Bewertung der Vertrauenswürdigkeit eines digitalen Langzeitarchivs dienen (→siehe Abschnitt [Zertifizierung](#)).

Datennehmende erhalten beim Abrufen von Materialien aus dem Archiv sogenannte **Dissemination Information Packages** oder kurz **DIPs**, die üblicherweise den SIPs sehr ähneln. Ob das kumulierte AIP oder alle einzelnen Delta-Pakete innerhalb eines DIPs ausgeliefert werden, hängt vom Archivinformationssystem und seinen Einstellungen ab. Im SLUBArchiv.digital werden beispielsweise nur die Metadaten ausgespielt, die bereits im SIP enthalten waren. Metadaten, die intern generiert und im AIP gespeichert wurden, werden im DIP nicht mit ausgespielt.



Herausgabe aus dem Archiv:

- SLUB Dresden. (2020). *DIP Spezifikation für automatischen Access SLUBArchiv*. SLUB Dresden. https://slubarchiv.slub-dresden.de/fileadmin/groups/slubsite/slubarchiv/SLUBArchiv_Konsumenten_DIP_Spezifikation_v1.0.pdf  [Wayback](#)²⁴⁰

²³⁸ https://web.archive.org/web/20221225110221/https://github.com/TIB-Digital-Preservation/CSV-Ingest_generisch

²³⁹ <https://web.archive.org/web/20221225204323/https://github.com/TIB-Digital-Preservation/pre-ingest-analyzer>

²⁴⁰ https://web.archive.org/web/20221208074957/https://slubarchiv.slub-dresden.de/fileadmin/groups/slubsite/slubarchiv/SLUBArchiv_Konsumenten_DIP_Spezifikation_v1.0.pdf

4.2 Archivinformationssysteme

Archivinformationssysteme (auch „Digital Preservation Systems“ genannt) sind Softwarelösungen, um digitale Inhalte unabhängig von sich ändernden Speichermedien und Technologien langfristig zu bewahren und zugänglich zu machen. Fast alle in der LZA-Community verbreiteten Archivinformationssysteme implementieren das → [OAIS-Modell](#) und können ein Dark Archive umsetzen. Folgende Tabelle gibt eine Übersicht:

Produktname	Hersteller	Land	Open Source	Kostenlos
Archivemata ²⁴¹	Artefactual Systems Inc.	Kanada	Ja	Ja
Rosetta ²⁴²	Ex Libris	Israel	Nein	Nein
Preservica ²⁴³	Preservica	Vereinigtes Königreich	Nein	Nein
Arkivum ²⁴⁴	Arkivum Ltd.	Vereinigtes Königreich	Nein	Nein
LOCKSS ²⁴⁵	Stanford University	USA	Ja	Ja
LIBNOVA ²⁴⁶	LIBNOVA, S.L.	Spain	Nein	Nein
Portico ²⁴⁷ (für e-Journals)	ITHAKA	USA	Nein	Nein

Tabelle 3: Überblick verbreiteter Archivinformationssysteme in der LZA-Community.

Darüberhinaus gibt es auch Entwicklungen an Archivinformationssystemen im deutschsprachigen Raum, wie z. B. [DIMAG](#)²⁴⁸  [Wayback](#)²⁴⁹ (Eigenentwicklung), [EWIG](#)²⁵⁰ (basierend auf Archivemata) oder [DA NRW](#)²⁵¹ (Eigenentwicklung), die jedoch nicht als nachnutzbare Software konzipiert sind.

²⁴¹ <https://www.archivemata.org/>

²⁴² <https://exlibrisgroup.com/products/rosetta-digital-asset-management-and-preservation/>

²⁴³ <https://preservica.com/>

²⁴⁴ <https://arkivum.com/>

²⁴⁵ <https://www.lockss.org/>

²⁴⁶ <https://www.libnova.com/>

²⁴⁷ <https://www.portico.org/>

²⁴⁸ <https://dimag-wiki.la-bw.de/xwiki/bin/view/Main/>

²⁴⁹ <https://web.archive.org/web/20230909145816/https://dimag-wiki.la-bw.de/xwiki/bin/view/Main/>

²⁵⁰ <https://ewig.zib.de/>

²⁵¹ <https://www.danrw.de/>

5 Erhaltungsstrategien

Die digitale Langzeitarchivierung setzt sich deutlich von der kurzzeitigen Datensicherung (Backup) ab (→siehe Abschnitt [Langzeitarchivierung vs. Backup](#)). Unterschiedliche Strategien der digitalen Langzeitarchivierung werden als „Erhaltungsstrategien“ bezeichnet. Diese Erhaltungsstrategien werden in der folgenden schematischen Darstellung abgebildet.

Zu den Erhaltungsstrategien gehören die „Bitstream Preservation“ und die „Content Preservation“. Die Content Preservation wird dabei weiter in „Logical Preservation“ und „Semantic Preservation“ unterteilt.

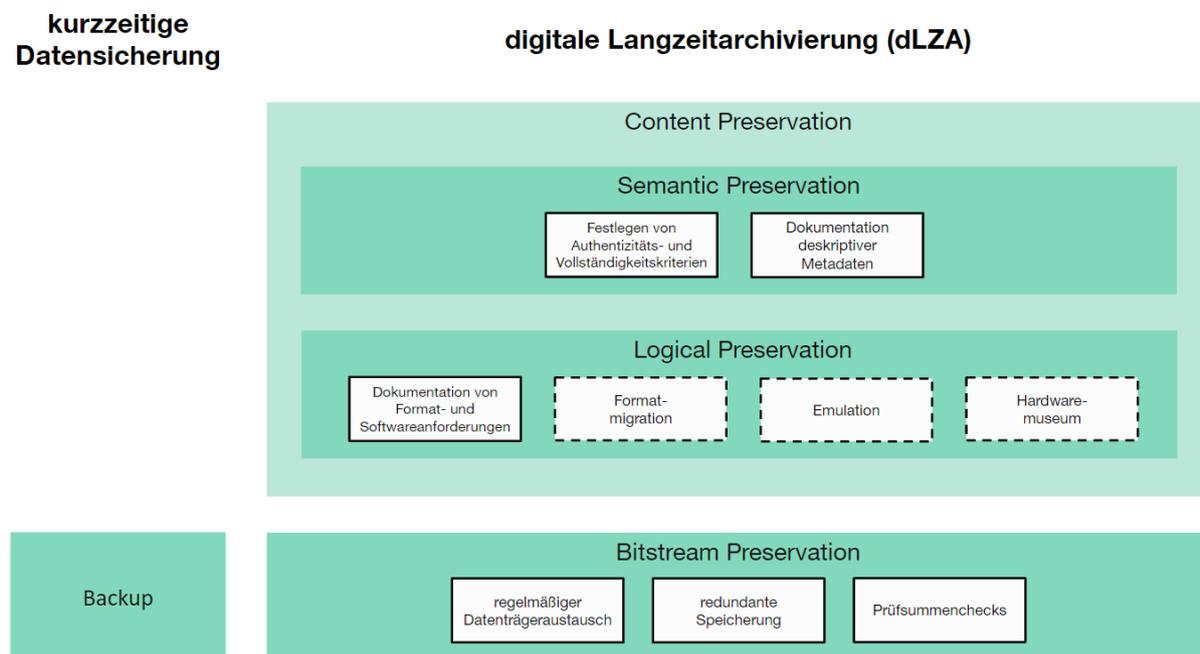


Abbildung 9: Kurzzeitige Datensicherung und digitale Langzeitarchivierung im Überblick. Innerhalb der Erhaltungsstrategien werden **Aktivitäten** (weiße Kästen) angewandt, die obligatorisch (dicker Rahmen) oder optional (gestrichelter Rahmen) sind.



3-Ebenen-Modell von K. Thibodeau & Levels of Digital Preservation der NDSA:

- Rudnik, P., & Lindlar, M. (2021, Februar 26). Foliensatz: Crashkurs Digitale Langzeitarchivierung - Ebenen-Modelle der LZA (3-Ebenen-Modell von K. Thibodeau & Levels of Digital Preservation der NDSA). <https://doi.org/10.5281/zenodo.4761975>
- Rudnik, P., & Lindlar, M. (2021). Video: Crashkurs Digitale Langzeitarchivierung - Ebenen-Modelle der LZA (3-Ebenen-Modell von K. Thibodeau & Levels of Digital Preservation der NDSA). <https://doi.org/10.5281/ZENODO.4761919>

5.1 Bitstream Preservation

Die Bitstream Preservation beschreibt den **Erhalt der Datenintegrität**. Sie ist eine der wichtigsten Erhaltungsstrategien der digitalen Langzeitarchivierung. Dies bedeutet, dass die Lesbarkeit der Sequenz aller gespeicherter Nullen und Einsen unverfälscht und vollständig auf Datenträgern über Technologiewechsel hinaus garantiert werden muss.

Die Bitstream Preservation umfasst eine Reihe von **Aktivitäten**. Die wichtigsten davon beziehen sich auf die Erhaltung der gesamten IT-Infrastruktur (also Hardware und Software), auf der die Informationen gespeichert werden.

Eine Herausforderung bei der Bitstream Preservation ist die Haltbarkeit der Datenträger. Während in Stein gemeißelte Inschriften über Tausende von Jahren erhalten bleiben können, ist die Haltbarkeit von digitalen Datenträgern wie Festplatten oder Disketten nur auf einige Jahre begrenzt; selbst bei professionell gepressten DVDs oder Blu-ray-Discs beschränkt sie sich auf wenige Jahrzehnte. Der Bitstream, also die Sequenz aus Nullen und Einsen auf dem Datenträger, aus welchem sich die eigentlichen Dateien ergeben, wird zudem durch äußere Umwelteinflüsse (wie Strahlung, Magnetfelder, usw.), Hardware-Ausfälle oder Software-Fehler bedroht. Daher ist es notwendig, diese technische Umgebung, die zum Betreiben des Archivs benötigt wird, zu pflegen und Hardware-Komponenten, insbesondere Datenträger, auszutauschen, bevor sie das Ende ihrer Lebensdauer erreicht haben.

Andererseits müssen Vorkehrungen getroffen werden, um eventuell auftretende Fehler in den Daten erkennen zu können. Dazu werden **Prüfsummen** erstellt, die anhand verschiedener Datenmerkmale berechnet und zusätzlich auf dem Datenträger abgelegt werden. Im Rahmen von Prüfsummenchecks, auch „Fixity-Checks“ genannt, werden gespeicherte und neu berechnete Prüfsummen miteinander verglichen. Ergibt sich eine abweichende Prüfsumme, haben sich die Daten geändert, was auf eine Beschädigung hinweist.

Um solchen Defekten, aber auch Katastrophenfällen wie Bränden oder Kriegen entgegenzutreten zu können, werden Forschungsdaten mit ihren zugehörigen Metadaten **redundant gespeichert**. In der Regel sind das drei Kopien an mindestens zwei unterschiedlichen Standorten. Einige Archive bieten darüber hinaus die Möglichkeit eine vierte Kopie (bspw. aus Bestandserhaltungsgründen) zu archivieren, wodurch die Speicherung noch robuster und zuverlässiger wird.

Eine Bitstream Preservation ohne Content Preservation wird bei Daten angewandt, für die eine Content Preservation noch nicht durchgeführt werden kann. Beispiele hierfür sind ausführbare Computercodes wie Quellcodes, Anwendungen bzw. Computerspiele. Darüber hinaus eignet sich die Bitstream Preservation ohne Content Preservation auch für die schnelle und einfache Erhaltung von Daten, wie etwa im Falle von RADAR4Culture.

Schließlich sei bemerkt, dass der Betrieb einer Bitstream Preservation mit Content Preservation deutlich kostenintensiver ist als ohne.



Digitale Langzeitarchive innerhalb von NFDI4Culture, die Bitstream Preservation anbieten:

- [RADAR4Culture](#)²⁵²
- [heiARCHIVE](#)²⁵³
- [TIB Preservation-as-a-Service](#)²⁵⁴
- [SLUBArchiv.digital](#)²⁵⁵ (nur Content Preservation)



Weitere Informationen zur Lebensdauer verschiedener Datenträger:

- Weber, A., & Piesche, C. (2021). 4.2 Datenspeicherung, -kuration und Langzeitverfügbarkeit. In *Praxishandbuch Forschungsdatenmanagement* (S. 327–356). De Gruyter Saur. <https://www.degruyter.com/document/doi/10.1515/9783110657807-019/html> [Wayback](#)²⁵⁶
- forschungsdaten.info. (2023, Juli 24). *Formate erhalten | Veröffentlichen und Archivieren*. <https://forschungsdaten.info/themen/rechte-und-pflichten/forschungsdaten-veroeffentlichen/>
- Deutsche Nationalbibliothek. (2014, August 31). *Bitstream preservation & Speichermedien—Nestor—Deutsche Nationalbibliothek—Wiki*. <https://wiki.dnb.de/pages/viewpage.action?pageId=95651771> [Wayback](#)²⁵⁷

²⁵² <https://www.radar-service.eu/>

²⁵³ <https://heiarchive.uni-heidelberg.de/de>

²⁵⁴ <https://www.tib.eu/de/publizieren-archivieren/digitale-langzeitarchivierung>

²⁵⁵ <https://slubarchiv.slub-dresden.de/>

²⁵⁶ <https://web.archive.org/web/20230616143542/https://www.degruyter.com/document/doi/10.1515/9783110657807-019/html>

²⁵⁷ <https://web.archive.org/web/20231115082324/https://wiki.dnb.de/pages/viewpage.action?pageId=95651771>

5.2 Content Preservation

Eine weitere Strategie der digitalen Langzeitarchivierung ist der **Erhalt der Nutzbarkeit gespeicherter Informationen**. Dieser wird durch die Content Preservation erreicht und schließt neben der technischen Interpretierbarkeit (logische Perspektive) auch die inhaltliche Interpretierbarkeit (semantische Perspektive) ein.

Digitale Langzeitarchive definieren ihre Erhaltungsstrategien nicht einheitlich. Während etwa die TIB Hannover in der [TIB Preservation Policy](https://www.tib.eu/de/die-tib/policies/preservation-policy)²⁵⁸  [Wayback](https://web.archive.org/web/20231125054955/https://www.tib.eu/de/die-tib/policies/preservation-policy)²⁵⁹ drei „Preservation Levels“ festlegt: Bitstream Preservation, Logical Preservation und Semantic Preservation. Die SLUB Dresden fasst die letzteren beiden zur Content Preservation zusammen (→siehe Abbildung 9). Die Content Preservation wird oft auch „funktionaler Erhalt“ genannt.

Eine Herausforderung bei der Content Preservation ist vor allem der **Alterungsprozess** der Informationen. Wie analoge Inschriften in Sprachen verfasst sein können, die sich in heutiger Zeit nicht mehr entschlüsseln lassen (wie etwa [Kretische Hieroglyphen](https://de.wikipedia.org/wiki/Kretische_Hieroglyphen)²⁶⁰  [Wayback](https://web.archive.org/web/20230220144302/https://de.wikipedia.org/wiki/Kretische_Hieroglyphen)²⁶¹ oder die Inka-Schrift [Quipu](https://de.wikipedia.org/wiki/Quipu)²⁶²  [Wayback](https://web.archive.org/web/20231015010213/https://de.wikipedia.org/wiki/Quipu)²⁶³), können Dateiformate im Laufe der Jahre veralten und somit unlesbar werden.

Wenn Langzeitarchive eine Content Preservation anbieten, bauen sie zumeist auf der Bitstream Preservation auf, übernehmen aber zusätzlich die Erhaltungsplanung. Diese beinhaltet eine Überwachung und Risikoanalyse von Dateiformaten, in der regelmäßig überprüft wird, ob ein Dateiformat weiterhin zukunftssicher ist oder ggf. Maßnahmen eingeleitet werden müssen (→siehe Abschnitt [Erhaltungsplanung](#)).

Einige Langzeitarchive, wie z. B. das SLUBArchiv, betreiben eine strikte Content Preservation: Beispielsweise können hier „nur statische [HTML-]Seiten ohne ausführbaren [JavaScript-]Code“ ([Langzeitarchivfähige Dateiformate](https://slubarchiv.slub-dresden.de/fileadmin/groups/slubsite/slubarchiv/SLUBArchiv_langzeitarchivfaehige_Dateiformate_v2.0.pdf)²⁶⁴  [Wayback](https://web.archive.org/web/20230605033654/https://slubarchiv.slub-dresden.de/fileadmin/groups/slubsite/slubarchiv/SLUBArchiv_langzeitarchivfaehige_Dateiformate_v2.0.pdf)²⁶⁵, S. 9) eingeliefert werden. Denn bisher ist eine Formatmigration für JavaScript-Code und somit eine Garantie der Nutzbarkeit beim Ausspielen der Daten Jahrzehnte später aus dem Archiv nicht möglich (→siehe Abschnitt [Sonderfall Webarchivierung](#)).

²⁵⁸ <https://www.tib.eu/de/die-tib/policies/preservation-policy>

²⁵⁹ <https://web.archive.org/web/20231125054955/https://www.tib.eu/de/die-tib/policies/preservation-policy>

²⁶⁰ https://de.wikipedia.org/wiki/Kretische_Hieroglyphen

²⁶¹ https://web.archive.org/web/20230220144302/https://de.wikipedia.org/wiki/Kretische_Hieroglyphen

²⁶² <https://de.wikipedia.org/wiki/Quipu>

²⁶³ <https://web.archive.org/web/20231015010213/https://de.wikipedia.org/wiki/Quipu>

²⁶⁴ https://slubarchiv.slub-dresden.de/fileadmin/groups/slubsite/slubarchiv/SLUBArchiv_langzeitarchivfaehige_Dateiformate_v2.0.pdf

²⁶⁵ https://web.archive.org/web/20230605033654/https://slubarchiv.slub-dresden.de/fileadmin/groups/slubsite/slubarchiv/SLUBArchiv_langzeitarchivfaehige_Dateiformate_v2.0.pdf

Logical Preservation

Die Logical Preservation ist die Erhaltungsstrategie der Content Preservation, bei welcher der Erhalt der technischen Interpretierbarkeit bzw. die **Ausführbarkeit der Objekte durch Software** sichergestellt wird.

Potentiell auftretende Probleme sind hier die langfristige Verfügbarkeit von Software (so lässt sich etwa das früher gängige Picture Exchange-Bildformat (.pcx) mit heute gängigen Programmen nicht mehr öffnen), aber auch Sicherheitsrisiken (wie beispielsweise der Adobe Flash Player, der in der Vergangenheit verstärkt Malware-Attacken zum Opfer fiel).

Voraussetzung für die Erhaltungsplanung ist bei der Logical Preservation eine **Dokumentation von** software- und dateiformatspezifischen **Anforderungen** in Form von technischen und administrativen Metadaten. So lässt sich eine künftige Fehlinterpretation bzw. falsche Darstellung eines Objekts durch Software vermeiden.

Im Wesentlichen werden **drei Erhaltungsmaßnahmen** unterschieden:

- Formatmigration
- Emulation
- Hardware Preservation



Digitale Langzeitarchive innerhalb NFDI4Culture, die Content Preservation anbieten:

- [SLUBArchiv.digital](https://slubarchiv.slub-dresden.de/)²⁶⁶
- [TIB Preservation-as-a-Service](https://www.tib.eu/de/publizieren-archivieren/digitale-langzeitarchivierung)²⁶⁷

Formatmigration

Ziel der Formatmigration ist es, Dateien im Archiv in aktuelle und zukunftssichere Formate bzw. Versionen zu überführen. Sie wird vom Archiv im Rahmen der Erhaltungsplanung innerhalb des Prozesses der digitalen Langzeitarchivierung durchgeführt.

Fast alle Computersysteme verwenden heutzutage zum Beispiel die ASCII-Codierung für Textdateien. Diese beschreibt, dass beispielsweise ein Zeichen mit der Nummer 65 (bzw. binär codiert 0100 0001) ein großes „A“ darstellt. Dadurch sind Textdateien seit den 1960er- Jahren für Menschen nahezu plattformunabhängig und ohne weitere Hilfsmittel darstellbar. Die meisten Word-DOC-Dokumente werden schon in ein paar Jahren veraltet sein und dann mit gängiger Software nicht mehr korrekt dargestellt werden können. Warum ist

²⁶⁶ <https://slubarchiv.slub-dresden.de/>

²⁶⁷ <https://www.tib.eu/de/publizieren-archivieren/digitale-langzeitarchivierung>

das so? Eine ASCII-Textdatei ist ein anwendungsunabhängiges Dateiformat, während moderne Office-Dokumente ihre Form und Erscheinung erst durch die Interpretation einer speziellen Software erhalten. Kompliziert ist es, wenn das Word-Dokument nicht nur aus einer, sondern aus verschiedenen verknüpften Dateien besteht, wie es bei Webseiten oder Computerspielen der Fall ist.

Um ein Format in ein neues Zielformat übertragen zu können, muss der Aufbau des Ausgangsformats bekannt sein. Ist das Ausgangsformat herstellerspezifisch und nicht-offen, also proprietär, können die Daten nicht gelesen und verändert werden; eine Migration ist dann nicht möglich (zum Beispiel können alte Windows 3.1-Anwendungen nicht migriert werden, da es keine Formatkonverter gibt). Hier ist nur eine →[Emulation](#) möglich.

Bei einer Formatmigration kann es zu Datenveränderungen oder sogar zu Informationsverlust kommen. Bekanntes Beispiel in der digitalen Langzeitarchivierung sind WordStar-Dokumente, in denen Tabellen nach der Migration auf eine neue Version nicht mehr angezeigt werden konnten, oder auch PDF-Dateien, bei denen sich durch die Migration die Seitennummerierung änderte.

Um Informationsverlust zu vermeiden, sollten die Originaldateien bzw. die einzelnen Archivpakete (AIPs) vor der Migration ebenfalls aufbewahrt werden.

Die Formatmigration arbeitet üblicherweise mit einzelnen Dateien. Jedoch existiert nicht für jeden vorliegenden Dateityp ein geeignetes Zielformat. Im Falle komplexer, verknüpfter Objekte ist es nach einer Migration in der Regel nur noch möglich, die einzelnen Dateien statt das Verbunddokument zu betrachten. Eine funktionierende Verknüpfung von Daten ist hier wesentlich.

Fallbeispiel: Formatmigration von Adobe Flash-Dateien

Bis Ende 2020 war es üblich, Adobe Flash-SWF-Dateien im digitalen Langzeitarchiv aufzubewahren. Der vorher sehr verbreitete und in Webseiten zur Darstellung multimedialer Inhalte genutzte Adobe Flash Player wurde Ende 2020 jedoch vom Hersteller selbst aufgrund kostenintensiver Wartung eingestellt. Wurde Flash von den im Archiv befindlichen Materialien nur zur Darstellung von Text und Bildern genutzt, konnten diese in HTML-Webseiten überführt werden. Handelte es sich jedoch um interaktive Anwendungen, konnten diese nur mit anderen Erhaltungsmaßnahmen, wie Emulation oder Hardware Preservation, erhalten werden.



Digitale Langzeitarchive innerhalb von NFDI4Culture, die Content Preservation mit Formatmigration anbieten:

- [SLUBArchiv.digital](https://slubarchiv.slub-dresden.de/)²⁶⁸
- [heiARCHIVE](https://heiarchive.uni-heidelberg.de/)²⁶⁹
- [TIB Preservation-as-a-Service](https://www.tib.eu/de/publizieren-archivieren/digitale-langzeitarchivierung)²⁷⁰

Emulation

Als Emulation wird die funktionelle Nachbildung eines Systems bezeichnet. Die Ausgangsdateien bzw. der Bitstream bleiben hierbei unverändert. Bei der Emulation werden die Laufzeitumgebung bzw. die benötigten Tools erhalten, um die archivierten Daten lesbar zu machen und geeignet darzustellen. Für einige Anwendungsbereiche ist das die einzige Erhaltungsmöglichkeit, wie folgendes Beispiel zeigt.

Fallbeispiel: Logical Preservation mittels Emulation

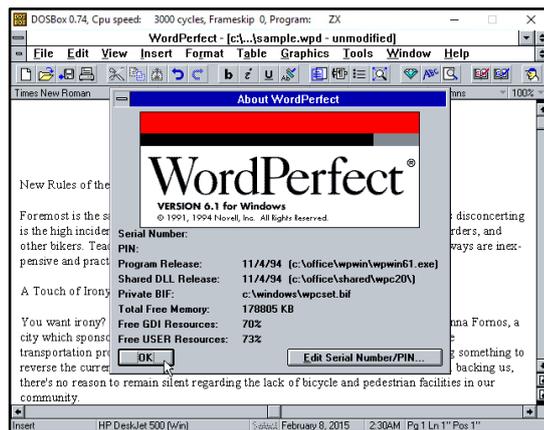
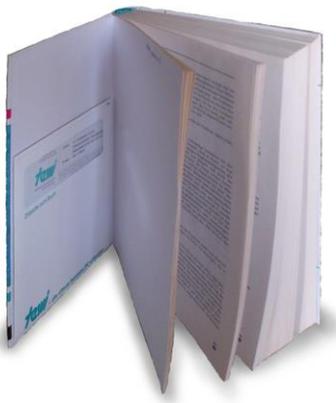


Abbildung 10: Buch mit 5,25-Zoll-Diskette aus den 1990ern, die u. a. WordPerfect-Dokumente enthält. (Links: © Matthias Razum, CC BY 4.0 International; rechts: © [WinWorld](https://winworldpc.com/)²⁷¹)

Bis Anfang der 2000er Jahre wurden Büchern manchmal Datenträger wie im linken Bild der oberen Abbildung beigelegt. Die Diskette im Buch enthält WordPerfect-Dokumente mit Tabellen und Anwendungen für Windows 3.1. Um diese unter modernen Betriebssystemen zu öffnen und auszuführen, müssen meist Emulatoren genutzt werden, wie im rechten Bild der Abbildung dargestellt ist: Hier läuft Windows 3.1 mit WordPerfect im [DOSBox-Emulator](https://www.dosbox.com/)²⁷² unter Windows 10.

²⁶⁸ <https://slubarchiv.slub-dresden.de/>

²⁶⁹ <https://heiarchive.uni-heidelberg.de/>

²⁷⁰ <https://www.tib.eu/de/publizieren-archivieren/digitale-langzeitarchivierung>

²⁷¹ <https://winworldpc.com/screenshot/c38ac2b7-02e2-8099-27c3-be11c3a4efbf>

²⁷² <https://www.dosbox.com/>

Die größte Schwierigkeit bei der Emulation ist jedoch die rechtliche Lage, vor allem hinsichtlich der Lizenzen kleinerer Software-Bestandteile (→siehe auch Abschnitt [Klärung rechtlicher Aspekte](#)). In der EU hängt das Urheberrecht am Datenträger, etwa einer CD-ROM, deren geschützte Daten nicht kopiert werden dürfen. Da in den USA hingegen das „fair use“-Prinzip gilt, ist Emulation dort verbreiteter.



Deutsche digitale Langzeitarchive, die Expertise mit Emulation haben:

- [Zentrum für Kunst und Medien](#)²⁷³  [Wayback](#)²⁷⁴
- [Albert-Ludwigs-Universität Freiburg](#)²⁷⁵

Hardware Preservation

In der digitalen Langzeitarchivierung verwendet man oft den Begriff „Hardware Museum“, um den Prozess der Hardware Preservation zu beschreiben. Im Folgenden wird der Prozess „Hardware Preservation“ genannt, um sich von Einrichtungen zu unterscheiden, die Hardware bewahren und ausstellen. Bei der Hardware Preservation werden digitale Objekte auf ihren originalen Datenträgern mit ihrer originalen Hardware- und Software-Umgebung erhalten. Dabei steht der Erhalt der digitalen Objekte, nicht der Erhalt der Hardware im Mittelpunkt.

Diese Erhaltungsmaßnahme der Content Preservation ist charakteristisch für die Frühphase eines digitalen Archivs, in der noch keine technische Möglichkeit zum Transfer bestimmter Daten auf aktuelle Plattformen bzw. zur Formatmigration und Emulation erarbeitet wurde. Die Hardware Preservation bietet zwar einen Einblick in die ursprüngliche Nutzungserfahrung digitaler Objekte, jedoch stellen der Platzbedarf, der eingeschränkte Zugriff für Nutzende und die Wartung der Geräte eine große Herausforderung dar.



Digitale Langzeitarchive mit Hardware Preservation:

- [Zentrum für Kunst und Medien](#)²⁷⁶  [Wayback](#)²⁷⁷

²⁷³ <https://zkm.de/de/sammlung-archiv/archive/zkm-archiv>

²⁷⁴ <https://web.archive.org/web/20230512065631/https://zkm.de/de/sammlung-archiv/archive/zkm-archiv>

²⁷⁵ <https://www.rz.uni-freiburg.de/de>

²⁷⁶ <https://zkm.de/de/sammlung-archiv/archive/zkm-archiv>

²⁷⁷ <https://web.archive.org/web/20230512065631/https://zkm.de/de/sammlung-archiv/archive/zkm-archiv>

Fallbeispiel: Grenzen der Emulation

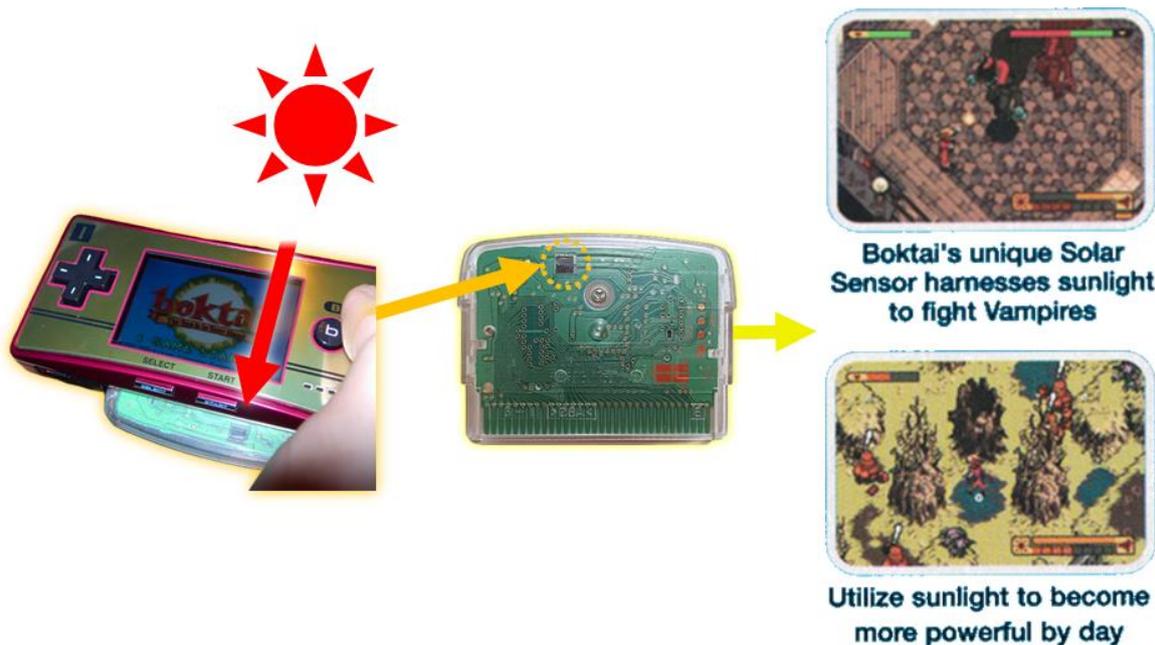


Abbildung 11: Gameboy-Advance-Spiel „Boktai: The Sun is in Your Hand“ von Konami Computer Entertainment Japan, 2003. Die vom Sonnensensor im oberen Teil der Karte erfasste Stärke des Sonnenlichts wird in Echtzeit im Spiel reflektiert. (Foto links: Gameboy Mikro mit Boktai-Spiel, © Autor unbenannt, via [Flickr](#), CC BY-SA 2.0; Mitte: Sonnensensor der Boktai-Spielkassette, © Tobias Kleinmann, CC0; rechts: Bildschirmausschnitte der Spielverpackung, © Autor unbenannt, via [Tumblr](#))

Die Archivierung interaktiver dynamischer Anwendungen und Spiele ist eine besondere Herausforderung. Im Beispiel soll das Spiel [Boktai: The Sun is in Your Hand](#)²⁷⁸ [Wayback](#)²⁷⁹ im Kontext eines Computerspielmuseums archiviert und Besucher:innen zugänglich gemacht werden.

Das Spiel wurde als Spielkassette für den [Nintendo Gameboy Advance](#)²⁸⁰ [Wayback](#)²⁸¹ veröffentlicht. Auf der Kassette ist ein spezieller Lichtsensor angebracht, der zum Spielen benötigt wird. Die erfasste Stärke des Sonnenlichts wird dabei in Echtzeit in der Spielwelt reflektiert: Das Sonnenlicht lässt den Protagonisten des Spiels „Django“ „tagsüber kraftvoller werden“ und kann „genutzt werden, um Vampire zu bekämpfen“ (siehe obere Abb. rechts).

Während das Spiel als solches emuliert werden kann, ist das bei diesem Sensor mittels heutiger Emulatoren nur über spezielle Umwege möglich. Das bedeutet, dass der Bitstream der Kassette zwar ausgelesen und emuliert werden könnte, das Spiel aber nicht, wie ursprünglich konzipiert, nutzbar ist. Es kann daher nur mittels Original-Hardware in vollem Umfang zugänglich gemacht werden. Zusätzlich kann der Bitstream der Kassette ins digitale Archiv geliefert werden.

²⁷⁸ https://en.wikipedia.org/wiki/Boktai:_The_Sun_Is_in_Your_Hand

²⁷⁹ https://web.archive.org/web/20231016121837/https://en.wikipedia.org/wiki/Boktai:_The_Sun_Is_in_Your_Hand

²⁸⁰ https://en.wikipedia.org/wiki/Game_Boy_Advance

²⁸¹ https://web.archive.org/web/20231123030719/https://en.wikipedia.org/wiki/Game_Boy_Advance

Fallbeispiel: Logical Preservation mittels Erhalt von Original-Hardware



Abbildung 12: Schreibmaschine KX W1500 von Panasonic, 1989. (Foto: © Autor ohne Namen)

Einige Systeme nutzen proprietäre Formate, was sowohl Hardware als auch Software umfassen kann. So kann zwar der „rohe“ Bitstream, der von einer [Panasonic KX W1510-Schreibmaschine](#)²⁸²  [Wayback](#)²⁸³ auf eine 3,5-Zoll-Diskette geschrieben wurde, als Ganzes ausgelesen werden. Der Inhalt einzelner Dokumente jedoch kann mit herkömmlichen Mitteln nicht ausgelesen werden, da das Dateisystem der Disketten proprietär ist.

Da es für diese Schreibmaschine keine Emulatoren gibt und im konkreten Fall eine an den Hersteller gerichtete Anfrage zur Formatspezifikation erfolglos blieb, können die gespeicherten Dokumente nur mittels der originalen Hardware zugänglich gemacht werden.

²⁸² <https://typewriterdatabase.com/1990-panasonic-w1505a.12551.typewriter>

²⁸³ <https://web.archive.org/web/20210122093507/https://typewriterdatabase.com/1990-panasonic-w1505a.12551.typewriter>

Semantic Preservation

Die Semantic Preservation ist die Erhaltungsstrategie der Content Preservation, bei der die langfristige **Interpretierbarkeit auf inhaltlicher Ebene** sichergestellt wird. Ziel ist, Authentizität, Vollständigkeit und Kontext eines Objektes zu bewahren sowie dessen Verlust zu vermeiden. Hier muss nicht nur vom Archiv sichergestellt werden, dass eine bestimmte Anzahl an Dateien geliefert wurde, sondern dass diese auch tatsächlich ihrem erwarteten Inhalt entsprechen. Dies ist insbesondere bei der Archivierung großer Konvolute (d. h. in Zusammenhang stehende Sammlungen von Dokumenten) relevant, die direkt von den Datengebern (z. B. Publishern) ans Archiv geliefert werden.

Um eine Semantic Preservation zu erreichen, werden Objekte mit deskriptiven Metadaten inhaltlich beschrieben, wobei die **Aktualität dieser Kontextinformationen regelmäßig** im Rahmen der Erhaltungsplanung **überprüft** wird.

Während bei der Bitstream Preservation in erster Linie darauf geachtet wird, dass technisch verlinkte Informationen, wie nicht eingebettete Schriftarten oder Grafiken mitgeliefert werden, wird im Rahmen der Semantic Preservation geprüft, ob verknüpfte Internet-Ressourcen noch erreichbar sind oder sich verändert haben (→siehe Abschnitt [Sonderfall Webarchivierung](#)). Das ist beispielsweise der Fall, wenn etwa Zeitschriften auf zusätzliche externe Dateien verweisen, diese aber nicht mit ins Archiv einreichen wollen.

In semantischen digitalen Archiven werden Materialien zudem oft mit [Linked Open Data \(LOD\)](#)²⁸⁴-Informationen angereichert. LOD bezeichnet strukturierte, maschinenlesbare Daten, die über das Internet verfügbar sind und miteinander verknüpft werden können. Im [DURAARK-Projekts](#)²⁸⁵  [Wayback](#)²⁸⁶ von 2016 wurde diese Methode genutzt, um 3D-Daten aus der Architektur semantisch zu bereichern. Konkret wurden →[signifikante Eigenschaften](#), wie architektonische Strukturen, Bauelemente und Konstruktionsweisen als Linked Open Data hinzugefügt und bewahrt. Dies ermöglichte nicht nur eine umfassende Kontextualisierung der archivierten 3D-Daten, sondern auch eine dynamische Anreicherung mit aktuellen Informationen über diese architektonischen Elemente. Damit wurde die langfristige Interpretierbarkeit, Nutzbarkeit und Authentizität der digitalen Archivmaterialien entscheidend verbessert.

Analog zur →[Formatmigration](#) auf logischer Ebene können Metadaten auf kontextueller Ebene migriert und versioniert werden. Voraussetzung hierfür ist der Gebrauch von Metadatenschemata und standardisiertem Vokabular (→siehe Abschnitt [Erfassen von Metadaten](#)).

²⁸⁴ <https://lod-cloud.net/>

²⁸⁵ <https://www.tib.eu/de/forschung-entwicklung/projektuebersicht/projektsteckbrief/duraark>

²⁸⁶ <https://web.archive.org/web/20230323152045/https://www.tib.eu/de/forschung-entwicklung/projektuebersicht/projektsteckbrief/duraark>



Semantic Preservation:

- Eisner, P. (2022, März 21). Vor der Archivierung: Kontrolle der Vollständigkeit bei Wiley DEAL Journals. *TIB-Blog*. <https://blogs.tib.eu/wp/tib/2022/03/21/vor-der-archivierung-kontrolle-der-vollstaendigkeit-bei-wiley-deal-journals/>  Wayback²⁸⁷
- Beetz, J., Blümel, I., Dietze, S., Fetahui, B., Gadiraju, U., Hecher, M., Krijnen, T., Lindlar, M., Tamke, M., Wessel, R., & Yu, R. (2016). Enrichment and Preservation of Architectural Knowledge. In S. Münster, M. Pfarr-Harfst, P. Kuroczyński, & M. Ioannides (Hrsg.), *3D Research Challenges in Cultural Heritage II* (Bd. 10025, S. 231–255). Springer International Publishing. https://doi.org/10.1007/978-3-319-47647-6_11

5.3 Exit-Strategie

Um die Verfügbarkeit der langzeitarchivierten Daten auch bei einem Ausfall des zugrundeliegenden Archivinformationssystems sicherzustellen, ist es notwendig, rechtzeitig Vorsorge zu treffen. Das dazu entwickelte Vorgehen wird „Exit-Strategie“ genannt. Diese stellt sicher, dass die zu den AIPs erfassten Metadaten, durchgeführten Erhaltungsmaßnahmen (z. B. Formatmigrationen) und Dokumentation der Speicherung im Archiv so umfassend und standardkonform (z. B. PREMIS) sind, dass sie auch ohne zugrundeliegende Archivsoftware wieder zusammengesetzt werden können.



Exit-Strategie:

- SLUB Dresden. (2022). *Exit Strategie Rosetta*. https://slubarchiv.slub-dresden.de/fileadmin/groups/slubsite/slubarchiv/SLUBArchiv_Exit_Strategie_v1.5.pdf  Wayback²⁸⁸
- TIB Hannover. (2023). *Export und Exit-Szenario*. <https://wiki.tib.eu/confluence/display/lza/Export+und+Exit-Szenario>

²⁸⁷ <https://web.archive.org/web/20230325135227/https://blogs.tib.eu/wp/tib/2022/03/21/vor-der-archivierung-kontrolle-der-vollstaendigkeit-bei-wiley-deal-journals/>

²⁸⁸ https://web.archive.org/web/20230605024403/https://slubarchiv.slub-dresden.de/fileadmin/groups/slubsite/slubarchiv/SLUBArchiv_Exit_Strategie_v1.5.pdf

5.4 Auffindbarkeit langzeitarchivierter Objekte

Objekte, die sich in digitaler Langzeitarchivierung befinden, werden in der Regel auch in Bibliothekskatalogen nachgewiesen. Manche Objekte befinden sich in Repositorien und sind für die dLZA vorgesehen, werden aber noch für einen Zugriff über das Internet angeboten. Wie können Nutzende herausfinden, ob Forschungsdaten langzeitarchiviert oder zumindest dafür vorgesehen sind?

In Bibliothekssystemen kann ein entsprechender Vermerk im Katalogeintrag hinterlegt werden. In einzelnen Fällen wird das bereits getan. Ein Beispiel hierfür ist der [K10plus](#)²⁸⁹, einer der überregionalen Online-Katalog im [deutschen Bibliotheksverbund](#)²⁹⁰. K10plus nutzt im [MARC 21 Format for Bibliographic Data](#)²⁹¹  [Wayback](#)²⁹² das [Feld „583 - Action Note“](#)²⁹³  [Wayback](#)²⁹⁴ zur Anzeige solcher Langzeitarchivierungsinformationen.

/ Felder \ ISBD \ MARC21 (FL_924) \ Citavi, Referencemanager (RIS) \ Endnote Tagged Format \ BibTex-Format \ RDF-Format \	
K10plusPPN:	1685375715 
Titel:	Katastrationsprotokolle über sämtliche zur Stadt Zschopau gehörigen Gebäude Cat. Nr. 1 bis mit 638 - 455
Erschienen:	Zschopau, [1853]
Umfang:	550 Blätter
Sprache(n):	Deutsch
Anmerkung:	Früher unter der Archivnummer: Altarchiv Abt. I Abschn. 12b Nr. 119 Archivierung/Langzeitarchivierung gewährleistet. SLUB Dresden
Sonstige Nummern:	OCoLC: 1137073182 
Sekundärausgabe:	Online-Ausgabe
Erschienen:	Dresden : SLUB ; Zschopau : Stadtarchiv, 2019
Gesamttitle:	Handschriften aus dem Stadtarchiv Zschopau / Stadtarchiv (Zschopau) Bürgerbücher sächsischer Städte Digitale Sammlungen. Saxonica
Standort der Vorlage:	Stadtarchiv Zschopau <455>
Link zum Volltext:	Elektronische Ressource: Zugang zum Digitalisat (Lizenzangabe: Kostenfrei zugänglich ohne Registrierung) URN: http://nbn-resolving.de/urn:nbn:de:bsz:14-db-id16853757150 Rechteinformation und Access Status: PDM 1.0 Open Access
Art und Inhalt:	Handschrift

Abbildung 13: „Felder“-Ansicht des [K10plus-Datensatz mit PPN 1685375715](#)²⁹⁵.

²⁸⁹ <https://opac.k10plus.de/>

²⁹⁰ <https://www.bibliotheksverband.de>

²⁹¹ <https://www.loc.gov/marc/bibliographic/>

²⁹² <https://web.archive.org/web/20231110223156/https://www.loc.gov/marc/bibliographic/>

²⁹³ <https://www.loc.gov/marc/bibliographic/bd583.html>

²⁹⁴ <https://web.archive.org/web/20230922142941/https://www.loc.gov/marc/bibliographic/bd583.html>

²⁹⁵ <https://opac.k10plus.de/DB=2.299/PPNSET?PPN=1685375715&PRS=HOL&HILN=888&INDEXSET=21>

```

000 $0 cam a22 4500
001 $01685375715
003 $0DE-627
005 $020200907124304.0
007 $0cr uuu---uuuuu
008 $0191212s1853 xx |||||o 00| ||ger c
024 7 $aurn:nbn:de:bsz:14-db-id16853757150$2urn
035 $a(DE-627)1685375715
035 $a(DE-599)KXP1685375715
035 $a(OCOLC)1137073182
040 $aDE-627$bger$cDE-627$erakwb
041 $ager
245 10$aKatastrationsprotokolle über sämtliche zur Stadt Zschopau gehörigen Gebäude Cat. Nr. 1 bis mit 638 - 455
264 1$aZschopau$c[1853]
300 $a550 Blätter
336 $aText$btxt$2rdacontent
337 $aComputermedien$bc$2rdamedia
338 $aOnline-Ressource$bcr$2rdacarrier
500 $aFrüher unter der Archivnummer: Altarchiv Abt. I Abschn. 12b Nr. 119
506 0 $qDE-14$aOpen Access$eControlled Vocabulary for Access Rights$uhttp://purl.org/coar/access_right/c_abf2
533 $aOnline-Ausgabe$bZschopau$bDresden$cStadtarchiv$cSLUB$d2019$e1 Online-Ressource$fSaxonica / Sächsische Landesbibliothek - Staats- und Universitätsbibliothek Dresden$fBürgerbücher sächsischer Städte$fHandschriften aus dem Stadtarchiv Zschopau$7|2019|||||||||
535 1 $aStadtarchiv Zschopau$3455
540 $qDE-14$aPublic Domain Mark 1.0$fPDM 1.0$2cc$uhttps://creativecommons.org/publicdomain/mark/1.0/
583 1 $aArchivierung/Langzeitarchivierung gewährleistet$2pdager$5DE-14
655 7$<$aHandschrift$>$0(DE-588)4023287-6$0(DE-627)10457187X$0(DE-576)208948376$2gnd-content
776 1 $nSLUB$aStadtarchiv (Zschopau)$tHandschriften aus dem Stadtarchiv Zschopau$gDresden$h20XX$w(DE-627)168209166X
776 1 $nSLUB$tBürgerbücher sächsischer Städte$gDresden$h20XX$w(DE-627)1683497651
776 1 $nSLUB$tDigitale Sammlungen. Saxonica$gDresden$h2007$w(DE-627)1333876785$w(DE-576)263876780
856 40$uhttp://digital.slub-dresden.de/id1685375715$xDigitalisierung$zkostenfrei$3Volltext
924 1 $a3564033351$bDE-14$914$cBSZ$dd
924 1 $a3567550926$bLFRER$9LFRER$cBSZ$dd$khhttp://digital.slub-dresden.de/id1685375715
951 $aBO

```

Abbildung 14: „MARC 21“-Ansicht des [K10plus-Datensatz mit PPN 1685375715](https://opac.k10plus.de/DB=2.299/PPNSET?PPN=1685375715&PRS=HOL&HILN=888&INDEXSET=21)²⁹⁶.

In der „MARC 21“-Ansicht im Feld „583“ des vorgestellten Beispiels findet man folgende Zeile:

```
583 1 $aArchivierung/Langzeitarchivierung gewährleistet$2pdager$5DE-14
```

Hier zeigt zunächst die „1“ an, dass die Daten öffentlich sind. Weiterhin wird im [Subfeld Code](#)²⁹⁷  [Wayback](https://www.wayback.org/)²⁹⁸ „\$a“ festgehalten, dass die „Archivierung/Langzeitarchivierung gewährleistet“ ist, was laut „\$2“ auf den normierten Begriff entsprechend der deutschsprachigen „[Dokumentation von Bestandsschutzmaßnahmen in MARC 583 Unterfeld \\$a – Terminogieliste](#)“²⁹⁹ formuliert ist. Alternativ sind hier auch die Hinweise „Archivie-

²⁹⁶ <https://opac.k10plus.de/DB=2.299/PPNSET?PPN=1685375715&PRS=HOL&HILN=888&INDEXSET=21>

²⁹⁷ <https://www.loc.gov/marc/bibliographic/bd583.html>

²⁹⁸ <https://web.archive.org/web/20230922142941/https://www.loc.gov/marc/bibliographic/bd583.html>

²⁹⁹ <https://www.agkva.org/919503035.html>

rung/Langzeitarchivierung geplant“ oder „Archivierung/Langzeitarchivierung nicht möglich“ möglich. Der Code „5“ verweist auf die [ISIL-Nummer](#)³⁰⁰  [Wayback](#)³⁰¹ der archivierenden Einrichtung, in diesem Fall steht „DE-14“ für die Sächsische Landesbibliothek – Staats- und Universitätsbibliothek Dresden (SLUB).

Auch andere Bibliotheksverbände oder Datenbanken bieten entsprechende Informationen, beispielsweise die [Zeitschriftendatenbank \(ZDB\)](#)³⁰².

Bisher werden Informationen zur Langzeitarchivierung noch nicht systematisch in bibliothekarischen Nachweisinstrumenten vorgehalten. Für Nutzende ist es daher schwierig, herauszufinden, ob ein digitales Objekt für eine Langzeitarchivierung vorgesehen oder bereits langzeitarchiviert ist. In der Regel ist eine direkte Suche nach dieser Information über Benutzerschnittstellen von Katalogsystemen für Nutzende nicht möglich.

Eine zumindest deutschlandweit einheitliche Aufnahme von Hinweisen zur digitalen Langzeitarchivierung in den Katalogen wäre ebenso wünschenswert wie die Möglichkeit einer Suchoption auf Nutzendenseite.



Feld 583 in MARC 21:

- Library of Congress (2022). *583 - Action Note*.
<https://www.loc.gov/marc/bibliographic/bd583.html>  [Wayback](#)³⁰³

³⁰⁰ <https://sigel.staatsbibliothek-berlin.de/startseite>

³⁰¹ <https://web.archive.org/web/20231122160533/https://sigel.staatsbibliothek-berlin.de/startseite>

³⁰² <https://zdb-katalog.de/title.xhtml?idn=02750879X&view=full>

³⁰³ <https://web.archive.org/web/20230922142941/https://www.loc.gov/marc/bibliographic/bd583.html>

6 Nachwort

Die vorliegende Handreichung zur digitalen Langzeitarchivierung für die NFDI4Culture Community soll Forscher:innen als erster Einstieg in die Thematik dienen und auch die Weiterverarbeitung der Daten im Archiv vorstellen. Neben der Erläuterung von Grund- und Fachbegriffen beschreibt sie die wichtigsten Erhaltungsstrategien sowie einzelne Schritte, die für eine digitale Langzeitarchivierung durchlaufen werden sollten.

Wesentlich ist, dass Forscher:innen möglichst frühzeitig, also spätestens zu Beginn der Konzeptionsphase von Projekten, sich über Möglichkeiten der Langzeitarchivierung der zu erwartenden Forschungsdaten informieren. Auch wenn grundsätzlich gilt: „Die Langzeitsicherung und Archivierung digitaler Inhalte ist technisch und organisatorisch gelöst, gut durch Modelle beschrieben und anhand von Standards prüfbar“ (DFG 2023, S. 35), „gibt es zum jetzigen Zeitpunkt keine verallgemeinerbare Lösung, die für alle Objektgattungen geeignet ist“ (DFG 2023, S. 10).

Während es für einzelne Objektgattungen, wie Text, Bild, Audio oder Video gut dokumentierte Lösungen für die digitale Langzeitarchivierung gibt, ist zu beachten, dass sich das in der Regel auf die Rohdaten bezieht. Entsprechend ist bei einer Langzeitarchivierung von Präsentationsoberflächen, Frontends, Designs u.ä. davon auszugehen, dass diese eher als *ephemer* betrachtet werden. Nur in komplexeren Emulationen können auch interaktive und dynamische Elemente erhalten werden. Es ist für Forschungsprojekte daher wesentlich, möglichst klar zwischen den Ebenen Forschungsdaten und Präsentation unterscheiden zu können.

Im Zuge der Erstellung der Handreichung standen der Austausch und die Abstimmung zwischen den NFDI4Culture-Langzeitarchiven im Vordergrund. Da Prozesse und Konzepte in der Langzeitarchivierung jedoch noch nicht flächendeckend in den Kulturwissenschaften gefestigt sind, wird die Handreichung nicht alle Aspekte abdecken können. Der Austausch mit der Community ist essentiell für die Erstellung einer solchen Handreichung. Daher heißen die Autor:innen jede Art von Feedback herzlich willkommen und möchten alle Leser:innen ermuntern, sich mitzuteilen. Bitte wenden Sie sich dazu an den [NFDI4Culture Helpdesk](#)³⁰⁴ und wählen Sie dort als Thema den Bereich „Datenpublikation und Langzeitarchivierung“ aus.

Die Autor:innen freuen sich auf Ihr Feedback!

³⁰⁴ <https://nfdi4culture.de/id/E2409>

7 Anhang

Abkürzungen

4Culture	Kurz für <i>NFDI4Culture</i> . Siehe auch → 4Culture-Disziplinen.
4Culture-Disziplinen	Sammelbegriff für die im NFDI-Konsortium „NFDI4Culture“ vertretenen Fachdisziplinen: Architektur, Kunstgeschichte, Medien-, Musik-, Film-, Theater- und Tanzwissenschaften
AIP	Ein <i>Archive Information Package</i> ist ein Paket digitaler Daten und Metadaten, das Informationen zur langfristigen Archivierung enthält.
ARK	Ein <i>Archival Resource Key</i> ist eine eindeutige Kennung, die zur Identifizierung von archivischen Ressourcen verwendet wird.
ASCII	Der <i>American Standard Code for Information Interchange</i> ist ein Zeichencodierungsstandard, der zur Darstellung von Text in Computern verwendet wird.
CARE	<i>Collective benefit, Authority to control, Responsibility</i> und <i>Ethics</i> sind Prinzipien, die die Grundlage für ethisches Datenmanagement bilden.
CIDOC	Das <i>Comité international pour la documentation</i> ist ein internationales Komitee für die Dokumentation von Kulturgütern.
CRM	Ein <i>Conceptual Reference Model</i> ist ein Referenzmodell, das abstrakte Konzepte zur Beschreibung von Informationen und Datenmodellen bietet.
CSV	<i>Comma Separated Values</i> ist ein Dateiformat, das zur Speicherung tabellarischer Daten verwendet wird und Werte durch Kommata trennt.
DAE	<i>Digital Asset Exchange</i> ist ein Dateiformat, das für den Austausch von 3D-Modellen und Szenen verwendet wird.
dLZA	digitale Langzeitarchivierung. Siehe → LZA
DIP	Ein <i>Dissemination Information Package</i> ist ein Paket, das Informationen enthält und zur Verbreitung von digitalen Inhalten dient.
DOI	Ein <i>Digital Object Identifier</i> ist eine eindeutige Kennung zur Identifizierung von digitalen Objekten, wie wissenschaftlichen Publikationen.
FAIR	<i>Findable, Accessible, Interoperable</i> und <i>Reusable</i> sind Prinzipien für die Verwaltung von Forschungsdaten, um deren Auffindbarkeit, Zugänglichkeit, Interoperabilität und Wiederverwendbarkeit sicherzustellen.
FIZ Karlsruhe	Fachinformationszentrum Karlsruhe – Leibniz-Institut für Informationsinfrastruktur GmbH
glTF	Das <i>Graphics Language Transmission Format</i> ist ein Dateiformat für 3D-Modelle und Szenen.
GND	Die <i>Gemeinsame Normdatei</i> ist ein Normdatensatz für bibliografische Informationen in Deutschland.

ISO	Die International Organization for Standardization ist eine internationale Organisation, die Standards für verschiedene Branchen entwickelt.
JPEG	Joint Photographic Experts Group ist ein weitverbreitetes Dateiformat zur (meist verlustbehafteten) Komprimierung von digitalen Bildern.
LIDO	Lightweight Information Describing Objects ist ein Datenmodell zur Beschreibung von Kulturobjekten.
LOD	Linked Open Data bezieht sich auf die Verknüpfung von frei zugänglichen strukturierten Daten im Web, um ihre Interoperabilität zu fördern.
LZA	Langzeitarchivierung ist die langfristige Aufbewahrung von digitalen Daten und Informationen zur Sicherstellung ihrer Verfügbarkeit über einen langen Zeitraum.
MEI	Die Music Encoding Initiative ist ein Standard zur Kodierung von musikalischen Notationen in digitaler Form.
METS	Der Metadata Encoding and Transmission Standard ist ein Standard für die Kodierung und Übertragung von Metadaten in digitalen Bibliotheken.
MODS	Das Metadata Object Description Schema ist ein Schema zur Beschreibung von Metadaten für digitale Ressourcen.
NFDI	Die Nationale Forschungsdateninfrastruktur ist eine deutsche Initiative zur Förderung und Bereitstellung von Forschungsdateninfrastrukturen.
NFDI4Culture	Konsortium für Forschungsdaten materieller und immaterieller Kulturgüter in der Nationalen Forschungsdateninfrastruktur (NFDI) →siehe 4Culture-Disziplinen
OAIS	Das Open Archival Information System ist ein Referenzmodell für digitale Archivierung und Langzeitarchivierung.
OCR	Optical Character Recognition bezieht sich auf die automatische Texterkennung aus Bildern oder gescannten Dokumenten.
PCX	Personal Computer Exchange ist ein älteres Dateiformat für Rastergrafiken, das in den 1980er und 1990er Jahren auf PCs verwendet wurde.
PID	Ein Persistent Identifier ist eine dauerhafte Kennung zur eindeutigen Identifizierung von digitalen Objekten oder Ressourcen.
SFTP	Das Secure File Transfer Protocol ist ein sicheres Protokoll zur Übertragung von Dateien über das Internet.
SIP	Ein Submission Information Package ist ein Datenpaket, das Informationen für die Einreichung digitaler Objekte in ein Archiv enthält.
SLUB Dresden	Sächsische Landesbibliothek – Staats- und Universitätsbibliothek Dresden
STL	Stereolithographie ist ein Dateiformat für 3D-Druck.
TAR	Tape Archive ist ein Archivierungsformat, das zur Zusammenfassung von Dateien und Verzeichnissen in einer einzelnen Datei verwendet wird.

TEI	<i>Text Encoding Initiative</i> ist eine Richtlinie für die Kodierung von Texten in digitaler Form.
TIB Hannover	Technische Informationsbibliothek Hannover
TIFF	Das <i>Tagged Image File Format</i> ist ein Dateiformat zur Speicherung von Bilddateien.
TRUST	<i>Transparency, Responsibility, User Focus, Sustainability</i> und <i>Technology</i> sind Prinzipien für vertrauenswürdigen Datenmanagement eines digitalen Langzeitarchivs.
UB Heidelberg	Universitätsbibliothek Heidelberg
URI	Ein <i>Uniform Resource Identifier</i> ist eine Zeichenfolge zur Identifizierung von Ressourcen im Internet.
URN	<i>Uniform Resource Name</i> ist eine spezielle Form des URI, die zur dauerhaften Identifizierung von Ressourcen verwendet wird.
WWW	Das <i>World Wide Web</i> ist ein über das Internet abrufbares System von Webseiten bzw. Hypertext-Dokumenten.
X3D	<i>Extensible 3D</i> ist ein Dateiformat für 3D-Grafiken und Szenen.
XML	<i>Extensible Markup Language</i> ist eine Auszeichnungssprache zur Strukturierung von Daten.
XMP	<i>Extensible Metadata Platform</i> ist ein Format zur Speicherung von Metadaten in Dateien.

Synonyme

Bisher werden im Bereich digitale Langzeitarchivierung viele Begriffe gleichbedeutend verwendet. Um eine Vereinheitlichung innerhalb der NFDI4Culture Community zu erreichen, folgt in diesem Abschnitt eine Übersicht gleichbedeutender Begriffe. **Bevorzugte Begriffe** werden dabei **fett** hervorgehoben und vorangestellt.

Deutscher Begriff	Englischer Begriff
<u>Archivinformationssystem</u> ⁴ = <u>Archivsystem</u> ¹ = <u>Langzeitarchivierungssoftware</u> ¹ = <u>LZA-Software</u> ¹ = <u>Langzeitarchivierungssystem</u> ¹ = <u>LZA-System</u> ¹	<u>archive system</u> = <u>digital preservation software</u>
<u>Informationspaket</u> ^{4,8}	<u>information package</u> ⁸
<u>Archivinformationspaket</u> = <u>Archivpaket</u> ^{7,4}	<u>archival information package</u>
<u>(Archiv-)Speicherung</u> = <u>Archivspeicher</u> ⁸ = <u>Archivablage</u> ⁷ = <u>Magazin</u> ¹	<u>archival storage</u>
<u>(Daten-)Abgabe</u> = <u>(Daten-)Übergabe</u> ¹	<u>submission</u> ¹
<u>(Daten-)Auslieferung</u>	<u>dissemination</u> ¹
<u>(Daten-)Übernahme</u> ¹ = <u>(Daten-)Einlieferung</u> = <u>Anlieferung</u> = <u>Ablieferung</u> = <u>Aufnahme</u> ⁷	<u>ingest</u> ¹
<u>(Daten-)Verwaltung</u>	<u>data management</u>
<u>(Daten-)Zugriff</u> ¹ = <u>Zugang</u>	<u>access</u> ¹
<u>Bitfehler</u>	<u>bit error</u> = <u>bit flip</u>
<u>Content Preservation</u> ⁴ = <u>funktionaler Erhalt</u> ⁴ = <u>klassische Langzeitarchivierung</u> ⁴	<u>content preservation</u>
<u>Datengebende</u> ³ = <u>Datengeber</u> = <u>(Daten-)Produzent</u> ¹ = <u>(Daten-)Einliefernder</u>	<u>producer</u>
<u>Datennehmende</u> ³ = <u>Datennehmer</u> = <u>(Daten-)Konsument</u> ⁴ = <u>Endnutzer</u> ¹	<u>consumer</u>
<u>digitale Langzeitarchivierung</u> = <u>Langzeitarchivierung</u> ^{1,6} = <u>Langzeitverfügbarkeit</u> ⁵ = <u>Langfristverfügbarkeit</u> ⁵ = <u>Langzeiterhaltung</u> = <u>Langzeitsicherung und digitale Bestandserhaltung</u> ⁵ = <u>Langzeitsicherung und Archivierung digitaler Inhalte</u> ⁵	<u>digital preservation</u> = <u>(digital) long term preservation</u> = <u>(digital) long term archiving</u>
<u>erhaltenswerte Eigenschaft</u> = <u>signifikante Eigenschaft</u> ⁷	<u>significant property</u> ⁷

<u>Erhaltungsplanung</u> ⁸ = <u>Planung der Langzeitarchivierung</u> ¹	<u>preservation planning</u> ⁸
<u>Erhaltungsmaßnahme</u> ¹ = <u>Langzeiterhaltungsmaßnahmen</u> ⁷	<u>preservation action</u>
<u>Erhaltungsstrategie</u> ¹ = <u>Erhaltungsmodell</u> = <u>Preservation Level</u> ² = <u>Preservation Strategie</u> ¹	<u>preservation strategy</u>
<u>Formaterkennung</u> ¹ = <u>Formatidentifizierung</u> ⁴	<u>format identification</u> = <u>file format identification</u>
<u>Handreichung</u> = <u>Guideline</u>	<u>guideline</u> ¹
<u>Hardware Preservation</u> ¹ = <u>Hardware-Museum</u> ⁴	<u>hardware preservation</u>
<u>intellektuelle Einheit</u> ⁴ = <u>Informationsobjekt</u> ⁷ = <u>Payload</u> ³	<u>intellectual entity</u> ⁷ = <u>payload</u> ³
<u>Nutzungspaket</u> ⁷ = <u>Auslieferungsinformationpaket</u> ⁸	<u>dissemination information package</u> ⁷
<u>Prüfsumme</u>	<u>check sum</u>
<u>Prüfsummencheck</u>	<u>fixity check</u>
<u>Transferpaket</u> ⁷	<u>submission information package</u> ⁷

Tabelle 4: Gegenüberstellung deutscher und englischer Fachbegriffe der digitalen Langzeitarchivierung.

¹ von [nestor](#)³⁰⁵ verwendet

² von TIB verwendet

³ von [NFDI4Culture, Task Area 4](#) vorgeschlagen ³⁰⁶

⁴ von [SLUB](#)³⁰⁷ verwendet

⁵ von DFG 2016 verwendet

⁶ von [forschungsdaten.info](#)³⁰⁸ verwendet

⁷ von [DIN 31644](#)³⁰⁹ verwendet

⁸ von [OAIS](#)³¹⁰ verwendet

³⁰⁵ <http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:0008-2010071949>

³⁰⁶ <https://nfdi4culture.de/id/E1824>

³⁰⁷ https://slubarchiv.slub-dresden.de/fileadmin/groups/slubsite/slubarchiv/SLUBArchiv_Glossar_v1.3.pdf

³⁰⁸ <https://forschungsdaten.info/>

³⁰⁹ <https://www.beuth.de/de/-/147058907>

³¹⁰ <http://d-nb.info/104761314X/34>

Glossare

- forschungsdaten.info. (2023, Juli 24). Glossar | Praxis kompakt. <https://forschungsdaten.info/praxis-kompakt/glossar/#c269821>
- Otto-von-Guericke-Universität. (2023). Glossar zum Forschungsdatenmanagement. <http://www.fdm.ovgu.de/home/Kurz+erkl%C3%A4rt/Glossar.html>
- Center for Open Science. (2023). OSF. <https://osf.io/rynmf/>
- California Digital Library. (2023). Glossary of Digital Library Terms. <https://cdlib.org/resources/technologists/glossary-of-digital-library-terms/>
- DIN Deutsches Institut für Normung e. V. (2012). *DIN 31644:2012-04, Information und Dokumentation—Kriterien für vertrauenswürdige digitale Langzeitarchive* (S. 23–35). Beuth Verlag GmbH. <https://doi.org/10.31030/1854419>
- ISO. (2012, September). *ISO 14721:2012(en), Space data and information transfer systems—Open archival information system (OAIS)—Reference model*. <https://www.iso.org/obp/ui/en/#iso:std:iso:14721:ed-2:v1:en>
- SLUB Dresden. (2023). *Glossar SLUBArchiv*. SLUB Dresden. https://slubarchiv.slub-dresden.de/fileadmin/groups/slubsite/slubarchiv/SLUBArchiv_Glossar_v1.3.pdf
- Digital Curation Centre. (2023). *Glossary | DCC*. <https://www.dcc.ac.uk/about/digital-curation/glossary>
- Kussmann, C. (2023, September 29). *Research Guides: Digital preservation program at the Libraries: Glossary*. <https://libguides.umn.edu/dp/glossary>
- Wyk, C. S.-V. (2021, Januar 14). *LibGuides: Digital preservation: Glossary*. <https://ufs.libguides.com/c.php?g=1113411&p=8118655>
- Digital Preservation Coalition. (2023). *Glossary—Digital Preservation Handbook*. <https://www.dpconline.org/handbook/glossary>
- Society of American Archivists. (2023). “Dictionary of Archives Terminology.” <https://dictionary.archivists.org/index.html>

Datensatz: Übersicht Wikidata–GND

Im Rahmen dieser Handreichung wurden Begriffe zur digitalen Langzeitarchivierung in Wikidata nachgetragen und mit existierenden GND-Einträgen verknüpft. Diese Wikidata-Begriffe wurden dabei u. a. durch Literaturquellen, Standards, Erhaltungsstrategien und -maßnahmen, Langzeitarchive, Archivinformationssoftware, persistente Identifikatoren und Synonyme ergänzt. Diese Übersicht ist unter <https://doi.org/10.5281/zenodo.10281971> als CSV-Tabelle abrufbar.

Datensatz: Bibliographie

Die für diese Handreichung herangezogenen Quellen wurden in einer Zotero-Gruppe „Handreichung_dLZA“ online publiziert: <https://www.zotero.org/groups/5329239>. Als Export in verschiedene bibliographische Formate (BibTeX, CSV, MODS XML und RIS) ist die Bibliographie unter <https://doi.org/10.5281/zenodo.10281971> abrufbar.

8 Referenzen

Letzte Prüfung aller Weblinks: 19.09.2023

Zitierstil: APA 7th ed.

Zotero-Gruppe: <https://www.zotero.org/groups/5329239>

Albrecht-Hohmaier, M., Bergmann, K., Eggersglüß, C., Polywka, A., & Stark, A. (2022). *Kulturwissenschaftliche Forschungsdaten – eine Einführung* (1.01). <https://nfdi4culture.de/go/E3508>

Altenhöner, R. (2019). Noch immer am Anfang? Stand und Perspektiven der Webarchivierung in Deutschland 2019. In Simone Fühles-Ubach & Ursula Georgy (Hg.): *Bibliotheksentwicklung im Netzwerk von Menschen, Informationstechnologie und Nachhaltigkeit*. Festschrift für Achim Oßwald (S. 237–250). Bock und Herchen Verlag. https://publiscologne.th-koeln.de/frontdoor/deliver/index/docId/1623/file/Festschrift_Osswald.pdf

Altenhöner, R., Berger, A., Bracht, C., Klimpel, P., Meyer, S., Neuburger, A., Stäcker, T., & Stein, R. (2023). *DFG-Praxisregeln „Digitalisierung“*. Aktualisierte Fassung 2022. <https://doi.org/10.5281/ZENODO.7435724>

Beetz, J., Blümel, I., Dietze, S., Fetahui, B., Gadiraju, U., Hecher, M., Krijnen, T., Lindlar, M., Tamke, M., Wessel, R., & Yu, R. (2016). Enrichment and Preservation of Architectural Knowledge. In S. Münster, M. Pfarr-Harfst, P. Kuroczyński, & M. Ioannides (Hrsg.), *3D Research Challenges in Cultural Heritage II* (Bd. 10025, S. 231–255). Springer International Publishing. https://doi.org/10.1007/978-3-319-47647-6_11

Beinert, T., Schmid, K., & Weimer, K. (2022). Infrastrukturen und Services für die wissenschaftliche Nutzung von Webarchiven: Ein Überblick. *o-bib. Das offene Bibliotheksjournal / Herausgeber VDB*, 9(3), 1–15. <https://doi.org/10.5282/o-bib/5821>

Bodenschatz, E., & Fritzsche, F. (2022). *Forum Datenpublikation und -archivierung #5: Persistent Identifiers – Bericht*. <https://nfdi4culture.de/id/E3541>

Boktai: The Sun Is in Your Hand. (2023). In *Wikipedia*. <https://en.wikipedia.org/w/index.php?title=Boktai: The Sun Is in Your Hand&oldid=1160701127>

Caplan, P. (2021). *PREMIS verstehen (2021)* (T. Beinert, Übers.; 2009 erstmals veröffentlicht; überarbeitet in 2017 und 2021). Library of Congress Network Development and MARC Standards Office. https://www.loc.gov/standards/premis/understandingPREMIS_german_2021.pdf

Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J. D., Anderson, J., & Hudson, M. (2019). *Die CARE-Prinzipien für indigene Data Governance* (deutsche Übersetzung). <https://doi.org/10.5281/zenodo.5995059>

Chiquet, S. (2001). Was heisst eigentlich archivwürdig? *Schweizerische Zeitschrift für Geschichte = Revue suisse d'histoire = Rivista storica svizzera*, 51(2001), 470–486. <https://doi.org/10.5169/seals-107875>

Coble, Z., & Karlin, J. (2023). Reference Rot in the Digital Humanities Literature: An Analysis of Citations Containing Website Links in DHQ. *Digital Humanities Quarterly*, 017(1). <http://www.digitalhumanities.org/dhq/vol/17/1/000662/000662.html>

CSV-Ingest generisch. (2023). TIB Hannover. https://github.com/TIB-Digital-Preservation/CSV-Ingest_generisch

Data Archiving Solution to Preserve Your Digital Assets For Decades. (2023). Arkivum. Data Archiving & Digital Preservation Solutions. <https://arkivum.com/>

- Deppe, A. (2020). FAIR, CARE und mehr. Prinzipien für einen verantwortungsvollen Umgang mit Forschungsdaten. In Matthias Schulze (Hrsg.): *Historisches Erbe und zeitgemäße Informationsinfrastrukturen: Bibliotheken am Anfang des 21. Jahrhunderts* (S. 299–312). kassel university press. <http://dx.doi.org/doi:10.17170/kobra-202010131934-299>
- Deutsche Forschungsgemeinschaft (Hrsg.). (2013). *Sicherung guter wissenschaftlicher Praxis. Denkschrift* (ergänzte Auflage). Wiley. <https://doi.org/10.1002/9783527679188.oth1>
- Deutsche Forschungsgemeinschaft. (2022a). *Leitlinien zur Sicherung guter wissenschaftlicher Praxis. Kodex*. (Version 1.1). <https://doi.org/10.5281/zenodo.6472827>
- Deutsche Forschungsgemeinschaft. (2022b). *Guidelines for Safeguarding Good Research Practice. Code of Conduct*. <https://doi.org/10.5281/ZENODO.3923601>
- Deutsches Institut für Normung (Hrsg.). (2023). DIN 33910, Information und Dokumentation—Objektschonende Digitalisierung von Archiv- und Bibliotheksgut (August 2023, S. 17). Beuth Verlag GmbH.
- Digital Asset Management: Rosetta*. (2023). Ex Libris. <https://exlibrisgroup.com/products/rosetta-digital-asset-management-and-preservation/>
- Eisner, P. (2022, März 21). Vor der Archivierung: Kontrolle der Vollständigkeit bei Wiley DEAL Journals. *TIB-Blog*. <https://blogs.tib.eu/wp/tib/2022/03/21/vor-der-archivierung-kontrolle-der-vollstaendigkeit-bei-wiley-deal-journals/>
- Film Conservation Metadata Data Dictionary* (3.0). (31.08.2020). [XSD]. TIB Hannover. <https://github.com/TIB-Digital-Preservation/FilmConservationMetadata>
- Fischer, V., & Petri, G. (2022). *Bildrechte in der kunsthistorischen Praxis – ein Leitfaden*. (Zweite, überarbeitete und erweiterte Auflage). Heidelberg University Library. <https://doi.org/10.11588/ARTDOK.00007769>
- Gerlach, R., Rex, J., Lang, K., Neute, N., & Schröter, A. (2020). *Fact Sheet: Personal Data*. <https://doi.org/10.5281/zenodo.4035992>
- GIDA - Global Indigenous Data Alliance. (2018). *Die CARE-Prinzipien für indigene Data Governance*. <https://www.gida-global.org/s/Die-CARE-Prinzipien-fur-indigene-Data-Governance.pdf>
- GO FAIR Initiative. (2022, Januar 21). *FAIR Principles*. GO FAIR. <https://www.go-fair.org/fair-principles/>
- Imeri, S., & Rizzolli, M. (2022). CARE Principles for Indigenous Data Governance: Eine Leitlinie für ethische Fragen im Umgang mit Forschungsdaten? *o-bib. Das offene Bibliotheksjournal / Herausgeber VDB*, 9(2), 1–14. <https://doi.org/10.5282/o-bib/5815>
- Jones, S. M., Klein, M., & Sompel, H. V. de. (2021). Robustifying Links To Combat Reference Rot. *The Code4Lib Journal*, 50. <https://journal.code4lib.org/articles/15509>
- Kailus, Angela. (2023). *Handreichung für ein FAIRes Management kulturwissenschaftlicher Forschungsdaten* (Albrecht-Hohmaier, Martin, Eggersgluß, Christoph, & Podschwadek, Frodo, Hrsg.). <https://doi.org/10.5281/zenodo.7716941>
- Keitel, C., & Schoger, A. (Hrsg.). (2013). *Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644* (1. Aufl., S. 1–112). Beuth Verlag GmbH. <https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=2033134>
- Kindling, M., & Schirnbacher, P. (2013). „Die digitale Forschungswelt“ als Gegenstand der Forschung. *Information - Wissenschaft & Praxis*, 64(2–3), 127–136. <https://doi.org/10.1515/iwp-2013-0017>
- Klein, M., Sompel, H. V. de, Sanderson, R., Shankar, H., Balakireva, L., Zhou, K., & Tobin, R. (26.12.2014). Scholarly Context Not Found: One in Five Articles Suffers from

- Reference Rot. *PLOS ONE*, 9(12), e115253. <https://doi.org/10.1371/journal.pone.0115253>
- Klimpel, P. (2021). *Urheberrechtsreform 2021: Neue Chancen für das kulturelle Erbe*. Digitales Deutsches Frauenarchiv. <https://nbn-resolving.org/urn:nbn:de:0297-zib-84315>
- Knaus, G., Kailus, A., & Stein, R. (2022). *LIDO-Handbuch für die Erfassung und Publikation von Metadaten zu kulturellen Objekten: Band 2: Malerei und Skulptur* (Deutsches Dokumentationszentrum Für Kunstgeschichte – Bildarchiv Foto Marburg & C. Bracht, Hrsg.). arthistoricum.net. <https://doi.org/10.11588/ARTHISTORICUM.1026>
- Knaus, G., Stein, R., & Kailus, A. (2019). *LIDO-Handbuch für die Erfassung und Publikation von Metadaten zu kulturellen Objekten: Band 1: Graphik*. arthistoricum.net. <https://doi.org/10.11588/arthistoricum.382.544>
- Koçer, K. K. (o. J.). *URN:NBN Resolver für Deutschland und Schweiz*. Abgerufen 3. August 2023, von <https://nbn-resolving.org/html/urn:nbn:de:0008-2010071949>
- Kommission Archäologie und Informationssysteme im Verband der Landesarchäologen Deutschlands. (2017). *Themenblätter zur Archivierung digitaler Daten – Archivwürdigkeit*. https://www.landesarchaeologien.de/fileadmin/mediamanager/004-Kommissionen/Archaeologie-und-Informationssysteme/Archivierung/TB_Archivwu_erdig-keit.pdf
- Koster, L. (2020). Persistent identifiers for heritage objects. *The Code4Lib Journal*, 47. <https://journal.code4lib.org/articles/14978>
- Laagland, H., Alloing, S., Gilissen, V., Steeman, M., & Swagemakers, W. (2020). *White paper on preferred formats*. <https://doi.org/10.5281/zenodo.4518486>
- Lang, K., Assmann, C., Neute, N., Gerlach, R., & Rex, J. (2023, Februar 27). *FAIR Assessment Tools Overview (Version 2.1)*. 3. Sächsische FDM-Tagung, Leipzig. <https://doi.org/10.5281/zenodo.7701941>
- Lin, D., Crabtree, J., Dillo, I., Downs, R. R., Edmunds, R., Giaretta, D., De Giusti, M., L'Hours, H., Hugo, W., Jenkyns, R., Khodiyar, V., Martone, M. E., Mokrane, M., Navale, V., Petters, J., Sierman, B., Sokolova, D. V., Stockhouse, M., & Westbrook, J. (2020). The TRUST Principles for digital repositories. *Scientific Data*, 7(1), 144. <https://doi.org/10.1038/s41597-020-0486-7>
- Lindlar, Michelle, Panitz, M., & Gadiraju, U. (2015). *Ingest and Storage of 3D Objects in a Digital Preservation System*. <https://doi.org/10.5281/zenodo.1115512>
- Lindlar, Michelle, & Schwab, F. (2018, September 21). *203.5 All that work ... for what? Return on investment for trustworthy archive certification processes – a case study*. IPRES2018. <https://doi.org/10.17605/OSF.IO/8A3SC>
- Lindlar, Micky, & Rudnik, P. (Regisseure). (2021a). *Video: Crashkurs Digitale Langzeitarchivierung - Einführung in PREMIS*. <https://doi.org/10.5281/zenodo.4761823>
- Lindlar, Micky, & Rudnik, P. (2021b). *Deutsche Übersetzung der 2019 NDSA Levels of Digital Preservation Matrix*. <https://doi.org/10.5281/zenodo.4718662>
- Music Encoding Initiative*. (2023). <https://music-encoding.org/>
- nestor ad-hoc WG TRUST. (2020). *Nestor endorsement of TRUST Principles*. <https://doi.org/10.5281/ZENODO.4120927>
- nestor-Arbeitsgruppe. (2013, August 27). *Referenzmodell für ein Offenes Archiv-Informationssystem—Deutsche Übersetzung 2.0*. <http://d-nb.info/104761314X/34>
- nestor-Arbeitsgruppe Standards für Metadaten, Transfer von Objekten in digitale Langzeitarchive und Objektzugriff (Hrsg.). (2009). *Wege ins Archiv: Ein Leitfaden für die Infor-*

- mationsübernahme in das digitale Langzeitarchiv.* (Version 1-zur öffentlichen Kommentierung). nestor - Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit Digitaler Ressourcen für Deutschland. <https://doi.org/10.18452/1527>
- nestor-Arbeitsgruppe Zertifizierung. (2019). *Erläuterungen zum nestor-Siegel für vertrauenswürdige digitale Langzeitarchive, Version 2.1.* Deutsche Nationalbibliothek. <http://dnb.info/1189191830/34>
- Neuroth, H., Oßwald, A., Scheffel, R., Strathmann, S., & Huth, K. (Hrsg.). (2010). *Nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.3.* <http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:0008-2010071949>
- O'Donohoe, E. (2021). *Server-Side Web Archiving* (S. 31). <https://doi.org/10.5281/zenodo.4889460>
- Owens, T. (2014, Februar 27). What Do you Mean by Archive? Genres of Usage for Digital Preservers. *The Signal: Digital Preservation*. <http://blogs.loc.gov/digitalpreservation/2014/02/what-do-you-mean-by-archive-genres-of-usage-for-digital-preservers/>
- Pampel, H., Weisweiler, N. L., Strecker, D., Witt, M., Vierkant, P., Elger, K., Bertelmann, R., Buys, M., Ferguson, L. M., Kindling, M., Kotarski, R., & Petras, V. (2023). Re3data – Indexing the Global Research Data Repository Landscape Since 2012. *Scientific Data*, 10(1), 571. <https://doi.org/10.1038/s41597-023-02462-y>
- PDFlib GmbH. (2023). *PDFlib: Die PDF/A-Standards.* <https://www.pdfliib.com/de/pdf-know-how/pdfa/die-pdfa-standards/>
- Pohlkamp, S., & Zarnitz, M. (2022). Die lange Sicht...: Nestor – Das Kompetenznetzwerk für Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen. *Bibliotheksdienst*, 56(5), 274–284. <https://doi.org/10.1515/bd-2022-0044>
- PRONOM - file format registry. (2006, März 23). The National Archives. <https://www.nationalarchives.gov.uk/PRONOM/BasicSearch/proBasicSearch.aspx?status=new>
- Puhl, J., Andorfer, P., Höckendorff, M., Schmunk, S., Stiller, J., & Thoden, K. (2015). Diskussion und Definition eines Research Data LifeCycle für die digitalen Geisteswissenschaften. *DARIAH-DE Working Papers*, 11, 53. <http://resolver.sub.uni-goettingen.de/purl/?dariah-2015-4>
- Research Data Alliance FAIR Data Maturity Model Working Group. (2020). *Das FAIR Data Maturity Model. Spezifikation und Leitlinien (deutsche Übersetzung).* <https://doi.org/10.5281/zenodo.5834115>
- Romeyke, A. (2013, August 27). Art1Pirat Blog: Baseline TIFF – Versuch einer Rekonstruktion, Teil1. *Art1Pirat Blog*. <https://web.archive.org/web/20220125221158/https://art1pirat.blogspot.com/2013/08/baseline-tiff-versuch-einer.html>
- Rudnik, P. (Regisseur). (2020). *Video: Crashkurs Digitale Langzeitarchivierung - Dateiformate.* <https://doi.org/10.5281/zenodo.3985075>
- Rudnik, P., & Lindlar, M. (2021a). *Foliensatz: Crashkurs Digitale Langzeitarchivierung - Ebenen-Modelle der LZA (3-Ebenen-Modell von K. Thibodeau & Levels of Digital Preservation der NDSA).* <https://doi.org/10.5281/zenodo.4761975>
- Rudnik, P., & Lindlar, M. (Regisseure). (2021b). *Video: Crashkurs Digitale Langzeitarchivierung - Ebenen-Modelle der LZA (3-Ebenen-Modell von K. Thibodeau & Levels of Digital Preservation der NDSA).* <https://doi.org/10.5281/zenodo.4761919>
- Rudnik, Pia (Regisseur). (2020). *Video: Crashkurs Digitale Langzeitarchivierung - Das Referenzmodell Open Archival Information System (OAIS).* <https://doi.org/10.5281/zenodo.3984982>

- Schoger, A., Dobratz, S., Altenhöner, R., & nestor-Arbeitsgruppe Vertrauenswürdige Archive - Zertifizierung (Hrsg.). (2008). *Kriterienkatalog vertrauenswürdige digitale Langzeitarchive* (Version 2). nestor - Kompetenznetzwerk Langzeitarchivierung. <http://nbn-resolving.de/urn:nbn:de:0008-2008021802>
- Schrimpf, S. (2013). *Das OAIS-Modell für die Langzeitarchivierung: Anwendung der ISO 14721 in Bibliotheken und Archiven*. Beuth Verlag GmbH. <https://www.beuth.de/de/publikation/das-oais-modell-fuer-die-langzeitarchivierung/186561465>
- Soltau, K. (2023). *RADAR4Culture: Handreichung zu personenbezogenen Daten [deutsch]*. <https://doi.org/10.5281/ZENODO.8221496>
- Soltau, K., & Goeller, S. (2023). *RADAR4Culture: Quickstart-Guide für Datengeberinnen und Datengeber [deutsch]*. <https://doi.org/10.5281/ZENODO.8221341>
- The British Museum. (2017, Juli 14). *Everything you ever wanted to know about the Rosetta Stone*. <https://www.britishmuseum.org/blog/everything-you-ever-wanted-know-about-rosetta-stone>
- The British Museum. (2022, Oktober 13). *Hieroglyphs: Unlocking ancient Egypt*. <https://www.britishmuseum.org/exhibitions/hieroglyphs-unlocking-ancient-egypt>
- The British Museum (@britishmuseum). (2017, Juli 11). *The Rosetta Stone -3D Model*. Sketchfab.Com. <https://sketchfab.com/models/1e03509704a3490e99a173e53b93e282/embed?autostart=1>
- The Library of Congress. (2023, August 31). *PREMIS: Preservation Metadata Maintenance Activity (Library of Congress)*. <https://www.loc.gov/standards/premis/>
- The National Archives. (2023, März 30). Download DROID: file format identification tool [Text]. *The National Archives*. <https://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/>
- TIB Hannover. (2016, Januar 31). *DURAARK*. <https://www.tib.eu/de/forschung-entwicklung/projektuebersicht/projektsteckbrief/duraark>
- TIB Hannover. (2022, Februar 18). *Preservation Policy*. <https://www.tib.eu/de/die-tib/policies/preservation-policy>
- TIB Hannover. (2023a). *Digitale Langzeitarchivierung*. <https://www.tib.eu/de/publizieren-archivieren/digitale-langzeitarchivierung>
- TIB Hannover. (2023b). *Persistent Identifiers (PIDs)*. <https://projects.tib.eu/pid-service/persistent-identifiers/persistent-identifiers-pids/>
- TIB Hannover. (2023c). *Pre-Ingest-Analyzer*. <https://github.com/TIB-Digital-Preservation/pre-ingest-analyzer>
- Vettermann, O., & Petri, G. (2023). Should I CARE about FAIR? – Ein juristischer Blick auf Prinzipien des Forschungsdatenmanagements. *RuZ - Recht und Zugang*, 4(1), 5–29. <https://doi.org/10.5771/2699-1284-2023-1-5>
- Web, C., Pearson, D., & Koerben, P. (2013). „Oh, you wanted us to preserve that?!“ Statements of Preservation Intent for the National Library of Australia’s Digital Collections. *D-Lib Magazine*, 19(1/2). <https://doi.org/10.1045/january2013-webb>
- Weber, A., & Piesche, C. (2021). 4.2 Datenspeicherung, -kuration und Langzeitverfügbarkeit. In *Praxishandbuch Forschungsdatenmanagement*, edited by Markus Putnings, Heike Neuroth and Janna Neumann (S. 327–356). De Gruyter Saur. <https://doi.org/10.1515/9783110657807-019>
- Wilkinson, C., Woodruff, S. D., Brohan, P., Claesson, S., Freeman, E., Koek, F., Lubker, S. J., Marzin, C., & Wheeler, D. (2011). Recovery of logbooks and international marine

data: The RECLAIM project. *International Journal of Climatology*, 31(7), 968–979.
<https://doi.org/10.1002/joc.2102>

Zittrain, J., Albert, K., & Lessig, L. (2014). Perma: Scoping and Addressing the Problem of Link and Reference Rot in Legal Citations. *Legal Information Management*, 14(2), 88–99. <https://doi.org/10.1017/S1472669614000255>

Impressum



NFDI4Culture – Konsortium für Forschungsdaten zum materiellen und immateriellen Kulturerbe, <https://nfdi4culture.de/>

Autor:innen:

Heseler, Jörg; ORCID: [0000-0002-1497-627X](https://orcid.org/0000-0002-1497-627X); Sächsische Landesbibliothek – Staats- und Universitätsbibliothek Dresden (SLUB Dresden); ROR ID: <https://ror.org/03wf51b65>

Büttner, Alexandra; ORCID: [0000-0002-4950-0941](https://orcid.org/0000-0002-4950-0941); Ruprecht-Karls-Universität Heidelberg, Universitätsbibliothek; ROR ID: <https://ror.org/038t36y30>

Arnold, Matthias; ORCID: [0000-0003-0876-6177](https://orcid.org/0000-0003-0876-6177); Ruprecht-Karls-Universität Heidelberg, Universitätsbibliothek; ROR ID: <https://ror.org/038t36y30>

Zitierung:

Heseler, Jörg; Büttner, Alexandra; & Arnold, Matthias. (2024). Grundlagen der digitalen Langzeitarchivierung: Eine Handreichung zur digitalen Langzeitarchivierung aus Perspektive der NFDI4Culture Community. Aufgabenbereich Datenpublikation und Langzeitarchivierung (TA4). <https://doi.org/10.5281/zenodo.10281971>.

Schlagwörter:

[NFDI4Culture](#); [Archivierung](#); [digitale Langzeitarchivierung](#); [Forschungsdaten](#); [RADAR4Culture](#); [SLUBArchiv.digital](#); [heiARCHIVE](#); [TIB Preservation-as-a-Service](#); [Datenübernahme](#); [OAIS-Modell](#); [Erhaltungsstrategie](#); [Bitstream Preservation](#); [Content Preservation](#); [Formatmigration](#); [Emulation](#); [Web-Archivierung](#); [Signifikante Eigenschaften](#)

Finanzierung:

NFDI4Culture und diese Untersuchung werden durch die Deutsche Forschungsgemeinschaft (DFG) unter der Projektnummer [441958017](#) gefördert.

DFG: <https://www.dfg.de/>; ROR ID: <https://ror.org/018meiw64>

Lizenzen:



Dieses Werk ist lizenziert unter einer [Creative Commons Namensnennung 4.0 International Lizenz \(CC BY 4.0\)](#).

© Copyright der Texte: Die Autor:innen, 2024.

Für die enthaltenen Materialien gelten, sofern angegeben, andere Lizenzen bzw. sind alle Rechte vorbehalten. Alle Markennamen werden in dieser Handreichung nur deskriptiv verwendet. Alle Rechte liegen bei den Markeninhabern.