

9

Working Paper
2024

KonsortSWD

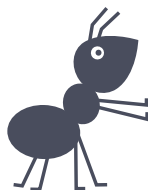


Konsortium für die
Sozial-, Verhaltens-, Bildungs- und
Wirtschaftswissenschaften

Langzeitarchivierung von Forschungsdaten

Einführung in das Thema zu Daten
der Sozial-, Verhaltens-, Bildungs- und
Wirtschaftswissenschaften in
Forschungsdatenzentren

Ute Hoffstätter, Anne Weber



April 2024

www.konsortswd.de

Langzeitarchivierung von Forschungsdaten

Einführung in das Thema zu Daten der Sozial-, Verhaltens-,
Bildungs- und Wirtschaftswissenschaften in
Forschungsdatenzentren

Ute Hoffstätter¹, Anne Weber¹

April 2024

<https://doi.org/10.5281/zenodo.10418834>

¹ Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW)

Abstract

Diese einführende Handreichung soll (angehende) Forschungsdatenzentren (FDZ) in den Bereichen Datenaufnahme und Langzeitarchivierung unterstützen. Ausgehend vom OAIS-Referenzmodell werden die Mindestanforderungen an ein Archiv dargestellt. Diese werden auf die Terminologie und Prozesse in Forschungsdatenzentren angewendet sowie mit Hinweisen und Beispielen aus der Praxis angereichert.

Keywords: Langzeitarchivierung (LZA), OAIS (Open Archival Information System), Forschungsdatenzentren (FDZ), Datenaufnahme

Inhaltsverzeichnis

1. Einleitung	3
2. Open Archival Information System (OAIS)-Referenzmodell	4
2.1. Zentrale Elemente.....	4
2.1.1. Rollen.....	5
2.1.2. Informationspakete.....	6
2.1.3. Funktionseinheiten.....	6
2.2. OAIS-Minimalanforderungen.....	7
3. Erfüllung der Archivierungsanforderungen in FDZ der Sozial-, Verhaltens-, Bildungs- und Wirtschaftsdaten	8
3.1. Verhandeln und Aufnehmen von Information	9
3.1.1. Scope und Auswahl/Priorisierung	9
3.1.2. Mindestanforderungen an die Daten und Metadaten	11
3.1.3. Datenübertragung und Eingangsprüfung.....	15
3.1.4. Aufnahme und Weiterverarbeitung.....	16
3.2. Ausreichende Kontrolle für die Langzeitarchivierung erlangen.....	17
3.3. Zielgruppe bestimmen	19
3.4. Sicherstellen der unmittelbaren Verstehbarkeit von Information.....	20
3.5. Bewährte Erhaltungspolicies und -abläufe einhalten.....	22
3.6. Verfügbarmachung und Bereitstellung.....	25
4. Zusammenfassung/Ausblick.....	26
Literaturverzeichnis.....	28

1. Einleitung¹

Kernaufgaben eines Forschungsdatenzentrums (FDZ) sind die Archivierung von Daten und der Datenzugang für Forschende.

*„In den FDZ werden Daten archiviert und über verschiedene Zugangswege unter Einhaltung des Datenschutzes für die Wissenschaft zugänglich gemacht.“
(Rat für Sozial- und Wirtschaftsdaten [RatSWD], 2022, S. 7)*

Nach der Handreichung zur Kernaufgabe des Datenzugangs (Hoffstätter & Linne, 2022) wird vorliegend der Archiv-Bereich in den Fokus genommen. Für eine dauerhaft gesicherte Datenbereitstellung ist eine Langzeitarchivierung ebendieser Daten und ihrer Metadaten² notwendig.

*„Immer mehr für die wissenschaftliche Sekundärnutzung potenziell interessante Daten sind in digitaler und damit leicht speicherbarer und weitergebarerer Form verfügbar. [...] Damit diese wertvollen Datensätze nicht verloren gehen, sind nachhaltige Konzepte der Langzeitarchivierung erforderlich.“
(Altenhöner & Oellers, 2012, S. 11)*

Langzeitarchivierung ist dabei mehr als die Sicherung der Dateien über eine bestimmte Zeit³ hinweg (Liegmann & Neuroth, 2010, Kap.1:3). Das heißt, eine rein physische Speicherung der Dateien ist die Grundlage, jedoch nicht ausreichend. Langzeitarchivierung ist die verantwortliche Entwicklung von Strategien, die den beständigen Wandel der digitalen Welt bewältigen können (Schwens & Liegmann, 2011, S. 567). Ziel sind der dauerhafte Erhalt sowie die technische und inhaltliche Interpretierbarkeit der digitalen Ressourcen, hier der Daten und Metadaten (Liegmann & Neuroth, 2010, Kap.1:3), wofür schon bei der Datenaufnahme wichtige Weichen gestellt werden müssen.

Das vorliegende Papier gibt einen Überblick über die besonders relevanten Aspekte, die bei der Langzeitarchivierung von Forschungsdaten, beginnend bei der Datenaufnahme, zu beachten sind. Um die bestehenden Anforderungen zu strukturieren und ein gemeinsames Vokabular zu nutzen, wird als konzeptioneller Rahmen das *Open Archival Information System (OAIS)*-Referenzmodell herangezogen und dessen zentrale Elemente sowie Minimal-Anforderungen an ein Archiv vorgestellt (2 Open Archival Information System (OAIS)-Referenzmodell). Darauf basierend werden anschließend Hilfestellungen gegeben, die FDZ im Bereich der Sozial-, Verhaltens-, Bildungs- und Wirtschaftsdaten bei der Erfüllung der im

¹ Vielen Dank an Daniel Buck, Dr. Jonas Recker und Dr. des. Henrike Schmidtchen für die fachkundigen und konstruktiven Reviewanmerkungen.

Alle im Dokument enthaltenen Links haben den Stand 09.04.2024.

² Metadaten sind Informationen, welche die Daten beschreiben und damit auch verständlich und interpretierbar machen.

³ In den Sozialwissenschaften und anderen Fachdisziplinen hat sich zwar eine zumindest zehnjährige Aufbewahrungsfrist der Forschungsdaten als Standard etabliert (s. etwa DFG-Kodex, Leitlinie 17 <https://wissenschaftliche-integritaet.de/kodex/archivierung/>); dennoch sollte eine Speicherung darüber hinaus angestrebt werden.

Modell gestellten zentralen Anforderungen mit Blick auf die Langzeitarchivierung unterstützen können (3 Erfüllung der Archivierungsanforderungen in FDZ der Sozial-, Verhaltens-, Bildungs- und Wirtschaftsdaten). Zuletzt erfolgt eine kurze Zusammenfassung sowie ein Ausblick auf weitere zentrale Elemente, die für eine funktionierende Langzeitarchivierung relevant sind (4 Zusammenfassung/Ausblick).⁴

2. Open Archival Information System (OAIS)-Referenzmodell

Das *Open Archival Information System* (OAIS)-Referenzmodell wurde Ende der 1990er Jahre von verschiedenen Raumfahrtorganisationen gemeinsam erarbeitet (The Consultative Committee For Space Data Systems, 2012). Es handelt sich dabei um ein sehr generisches Modell, welches Elemente, Prozesse, Aufgabenbereiche und Verantwortlichkeiten von Langzeitarchiven definiert und wichtige terminologische Festlegungen trifft (OAIS-Übersetzung/Terminologie, 2013, S. 1; Schumann, 2012, S. 40). „Ein OAIS ist ein Archiv, das aus einer Organisation, die Teil einer größeren Organisation sein kann, aus Menschen und Systemen besteht, das die Verantwortung übernommen hat, Information zu erhalten und sie einer vorgesehenen Zielgruppe zugänglich zu machen.“ (OAIS-Übersetzung/Terminologie, 2013, S. 2). Das Modell selbst ist sehr ausführlich. Hier werden nur einzelne Konzepte, die zentral für den Fokus der Handreichung als Einstieg in die Datenaufnahme und Langzeitarchivierung relevant sind, einführend vorgestellt.

2.1. Zentrale Elemente

Das OAIS-Referenzmodell (OAIS-Übersetzung/Terminologie (2013) (vgl. Abbildung 1) enthält:

- drei **Rollen** (in Abbildung grün links und rechts sowie unten ohne farbliche Markierung): Produzent*innen, Management, Endnutzer*innen
- drei **Informationspakete** (in Abbildung blau): SIP, AIP, DIP
- sechs **Funktionseinheiten** (in Abbildung orange): Übernahme, Archivspeicher, Datenverwaltung, Administration, Erhaltungsplanung, Zugriff

In einem OAIS können Produzent*innen Informationspakete an das Archiv übergeben. Das Management legt dabei die allgemeine Strategie des Archivs fest, wie diese Informationspakete in den Funktionseinheiten verarbeitet werden. Endnutzer*innen können dann Informationspakete von Interesse finden und erwerben.

⁴ Die Ausführungen dieser Handreichung setzen den Fokus auf quantitative Daten, werden jedoch möglichst generisch gehalten, sodass sie großteils auch für andere Datentypen gelten können. Für etwa qualitative Daten siehe insbesondere QualidataNet (<https://www.qualidatanet.com>).

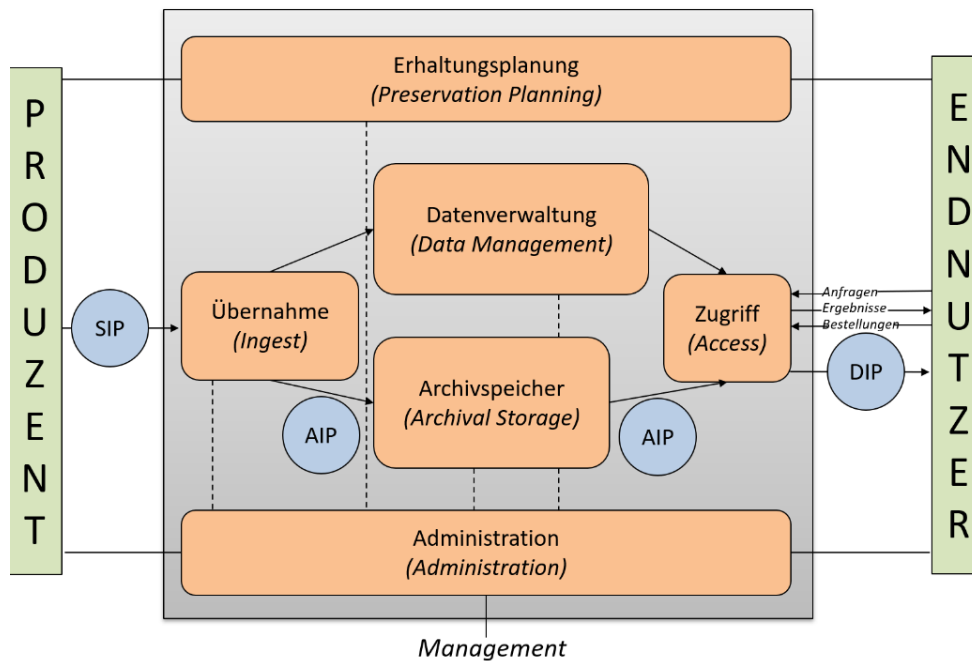


Abbildung 1: Umgebungsmodell: Übersicht über die zentralen Elemente des OAIS-Referenzmodells: Rollen, Informationspakete und Funktionseinheiten (Abbildung basierend auf OAIS-Übersetzung/Terminologie (2013))

Im Folgenden wird auf diese verschiedenen Elemente des OAIS-Referenzmodells genauer eingegangen.

2.1.1. Rollen

Die zentralen Rollen im OAIS-Referenzmodell sind:

- **Produzent*innen (Producer)**: Personengruppen oder Systeme, die die zu archivierende Information herstellen und zur Archivierung an das OAIS übergeben
- **Management (Management)**: Personengruppe, die das OAIS in seiner Zielsetzung verantwortet (entweder als eigenständiges Archiv oder als Teil einer größeren Organisation). Es legt die allgemeine Strategie des Archivs fest, wie die übergebenen Informationen verarbeitet werden.
- **Endnutzer*innen (Consumer)**: Personengruppen oder Systeme, die auf archivierte Information zugreifen und sie benutzen wollen. Mit Blick auf die Endnutzer*innen bestimmt das OAIS eine sogenannte vorgesehene Zielgruppe (*Designated Community*), welche die archivierten Informationen nutzen wird und der eine bestimmte Knowledge Base (z. B. ein bestimmter fachlicher Background) zugeschrieben wird.

2.1.2. Informationspakete

Im OAIS-Referenzmodell werden drei Informationspakete⁵ definiert, die nacheinander mit unterschiedlichen Zielen verarbeitet werden.

- Das **Übergabeinformationspaket** (*Submission Information Package – SIP*) ist die Version, die von Produzent*innen an das OAIS geliefert wird. Der genaue Inhalt wird vom Archiv festgelegt und in der Regel über eine Übergabvereinbarung (z. B. Vertrag) geregelt.
- Das **Archivinformationspaket** (*Archival Information Package – AIP*) ist die Version, die dauerhaft gespeichert wird und damit das zentrale Objekt von Interesse des Archivs. Es wird in der Regel im Archiv von Mitarbeitenden mit Informationen angereichert, die für die Langzeitarchivierung und Erhaltung nötig sind, etwa Metadaten, wie das Dateiformat, Erstellungsdatum, persistente Identifikatoren (z. B. DOI) (Schrimpf, 2012, S. 55). Die Unterscheidung zum SIP wird getroffen, da ggf. nicht alle Informationen der Produzent*innen relevant für das Archiv sind und die Dateien vor der Archivierung mit weiteren Informationen angereichert werden (Lavoie, 2014, S. 14). Gleichzeitig kann der Inhalt des SIP in unveränderter Form auch Teil des AIP sein.
- Das **Auslieferungsinformationspaket** (*Dissemination Information Package – DIP*) ist die Version, die den Endnutzer*innen bereitgestellt wird. Dieses Paket wird aus dem AIP generiert und kann davon abweichen, da das DIP etwa nur ein Subset an Informationen (beispielsweise aus datenschutzrechtlichen Gründen) enthält oder ein anderes Dateiformat hat. Auch die Metadaten, die für die Endnutzer*innen bereitgestellt werden, weichen vom AIP ab, da etwa die technischen Metadaten für sie nicht relevant sind (Schrimpf, 2012, S. 55). Dies ist im Regelfall die jeweils aktuelle Version, welche mit gängigen Systemen und Software genutzt werden kann.

2.1.3. Funktionseinheiten

Im OAIS-Referenzmodell werden sechs Funktionseinheiten eines OAIS definiert: Der Vorgang der Übernahme des SIP in das OAIS wird **Übernahme** (*Ingest*) genannt. Im Archiv wird das übernommene Objekt dann mit weiteren Informationen angereichert, um es dauerhaft verständlich und technisch verfügbar zu machen. Dieses angereicherte Objekt wird als AIP in den **Archivspeicher** (*Archival Storage*) überführt. Im Archivspeicher wird die langfristige Speicherung des AIP sichergestellt. Dazu gehören u. a. das Monitoring der Bestände, Integritätsprüfungen, Fehlerkontrolle sowie ein Prozess zur Notfallwiederherstellung der Dateien (OAIS-Übersetzung/Terminologie, 2012, S. 32). Bei der Übergabe des AIP an den Archivspeicher (*Archival Storage*) werden die Paketbeschreibungen (z. B. Titel, Autorschaft, ...) vom AIP in die Funktionseinheit **Datenverwaltung** (*Data Management*) überführt. Hier findet die Verwaltung der Metadaten der Archivbestände, also der Informationen, die die Archivbestände beschreiben, durch die Mitarbeitenden statt. In der Funktionseinheit

⁵ Mit Informationspaket ist dabei ein abstraktes Paket gemeint, d. h. die Dateien selbst müssen nicht alle an einem Ort gebündelt sein. Beispielsweise können die Metadaten zu den Daten, verknüpft über eine ID, in einer separaten Datenbank liegen.

Erhaltungsplanung (Preservation Planning) werden ganzheitliche Erhaltungsstrategien (wie Migration und Emulation) für die Archivbestände erarbeitet. Es werden u. a. der Technologiefortschritt und die Bedarfe der vorgesehenen Zielgruppe (*Designated Community*) beobachtet und konkrete Erhaltungsmaßnahmen geplant. Die Erhaltungsplanung ist übergreifend über die Prozesse von Übernahme, Datenverwaltung, Archivspeicher gelagert (Schrimpf, 2012, S. 57). In der Funktionseinheit **Administration (Administration)** werden Services zur Verwaltung des gesamten Archivs zur Verfügung gestellt, beispielsweise ein Muster für eine Übergabvereinbarung (*Submission Agreement*). Hier werden zum Beispiel Übergabvereinbarungen mit den Produzent*innen ausgehandelt, der Archivbetrieb überwacht, und Standards, Workflows und Policies festgelegt und weiterentwickelt (Schrimpf, 2012, S. 57). Für den **Zugriff (Access)** auf die Objekte durch Endnutzer*innen wird auf Grundlage des AIP ein DIP generiert. Die Endnutzer*innen können diese DIP recherchieren und über die in der Datenverwaltung hinterlegten Informationen (z. B. Titel, Schlagworte, ...) auffinden und bestellen bzw. beantragen (z. B. über einen Datenkatalog). Diese DIP werden den Endnutzer*innen dann bereitgestellt.

2.2. OAIS-Minimalanforderungen

Anknüpfend an das beschriebene OAIS-Referenzmodell ergeben sich sechs Minimalanforderungen, die verbindlich erfüllt sein sollten, um ein OAIS zu betreiben (OAIS-Übersetzung/Terminologie, 2013, S. 28). Das OAIS sollte:

1. **mit Produzenten*innen über Information verhandeln** und diese entsprechend annehmen.
2. genügend **Kontrolle** über die angebotene Information bekommen, in dem Maß, das benötigt wird, um deren Langzeiterhaltung sicherzustellen
3. bestimmen, entweder allein oder zusammen mit anderen, welche Gruppen zur vorgesehenen **Zielgruppe** gehören sollten und deswegen fähig sein sollten die angebotene Information zu verstehen, um dadurch ihr Grundwissen zu definieren.
4. **sicherstellen**, dass die zu erhaltende Information für die vorgesehene Zielgruppe unmittelbar verstehbar ist. Insbesondere sollte die vorgesehene Zielgruppe befähigt sein, die **Information** ohne den Gebrauch spezieller Hilfsmittel wie die Hilfe von Expert*innen, die die Information erstellt haben, **zu verstehen**.
5. **dokumentierten Richtlinien und Abläufen** folgen, die sicherstellen, dass die Informationen gegen alle vorstellbaren Gefahren **geschützt** sind, einschließlich der Schließung eines Archivs, sicherstellend, dass sie niemals gelöscht wird, außer wenn es als Bestandteil einer erprobten Strategie gestattet wird. Es sollte keine ad-hoc Löschungen geben.
6. die archivierten Informationen der vorgesehenen Zielgruppe **verfügbar machen** und die Auslieferung der Information ermöglichen, mit Belegen für ihre Authentizität

3. Erfüllung der Archivierungsanforderungen in FDZ der Sozial-, Verhaltens-, Bildungs- und Wirtschaftsdaten

Forschungsdatenzentren für den Bereich der Sozial-, Verhaltens-, Bildungs- und Wirtschaftsdaten können an die im vorangegangenen Kapitel dargestellten allgemeinen Betrachtungen zu einem OAIS grundsätzlich unmittelbar anknüpfen: (Interne oder externe) Datengebende (Produzent*innen/*Producer*) geben Dateien (digitale Objekte), z. B. Forschungsdaten und Dokumentationsmaterial, an ein FDZ (OAIS). Es wird ein Datenüberlassungsvertrag (Übergabvereinbarung / *Submission Agreement*) zwischen den Datengebenden (Produzent*innen/*Producer*) und dem FDZ (OAIS) geschlossen. Die überlassenen Dateien und Informationen (Übergabeinformationspaket / *Submission Information Package – SIP*) werden im Bereich Datenaufnahme (Übernahme/*Ingest*) aufgenommen und verarbeitet. Sie werden dort geprüft, ggf. verändert (z. B. anonymisiert) und mit Informationen (insb. Metadaten) angereichert. Diese angereicherten Dateien und Informationen (Archivinformationspaket / *Archival Information Package – AIP*) werden zur langfristigen Sicherung (Archivspeicher / *Archival Storage*) gegeben. Im Metadatenkatalog (Datenverwaltung / *Data Management*) werden die deskriptiven Metadaten zu den Daten für die Community bzw. Communities an Datennutzer*innen (vorgesehenen Zielgruppe / *Designated Community*) recherchierbar und bestellbar gemacht. Die Daten, die Datennutzende (Endnutzer*innen / *Consumer*) bestellen können (Auslieferungsinformationspaket / *Dissemination Information Package – DIP*), sind ggf. noch einmal eine abgewandelte Version davon (z. B. weil nicht alle Variablen zur Verfügung gestellt werden dürfen). Im Bereich Datenzugang (Zugriff/*Access*) werden die Anfragen der Datennutzenden geprüft und den Datennutzenden dann die Daten (Auslieferungsinformationspaket / *Dissemination Information Package – DIP*) über im FDZ (OAIS) verfügbaren Bereitstellungswege (z. B. Download, Secure Remote Access oder Gastwissenschaftsarbeitsplatz) bereitgestellt. Die FDZ-Leitung (Management/*Management*) gestaltet dabei die Standards und Policies des FDZ (OAIS) (z. B. Verträge, Überwachung der Infrastruktur, Gestaltung der Arbeitsprozesse) sowie deren Pflege. Insgesamt muss das FDZ (OAIS) dabei die sechs definierten Minimalanforderungen an ein OAIS verbindlich erfüllen.

Im Folgenden werden nun Hilfestellungen gegeben, wie die allgemein formulierten Minimalanforderungen von FDZ im Bereich der Sozial-, Verhaltens-, Bildungs- und Wirtschaftsdaten konkret erfüllt werden können. Der Fokus liegt dabei auf den ersten fünf Anforderungen, da sich diese unmittelbar auf die Archivierungskernaufgabe beziehen, während die sechste Anforderung eher auf die Kernaufgabe des Zugangs abzielt, zu der bereits eine Handreichung veröffentlicht wurde (Hoffstätter & Linne, 2022).

Die genaue Ausgestaltung der Erfüllung der Anforderungen ist immer von den jeweiligen Spezifika (Datentypen, Anzahl der digitalen Objekte, Ressourcen, Mission, Ressourcen etc.) eines FDZ abhängig. Dennoch ist es möglich, besonders wichtige und allgemeingültige Aspekte zu nennen und Beispiele zu geben, die für FDZ ein hilfreicher Ansatzpunkt sein können und ggf. dabei helfen, eigene Lücken zu identifizieren und zu füllen.

Im Folgenden wird eher die Terminologie der FDZ-Community genutzt, wobei einzelne englische Begriffe auch in der FDZ-Community genutzt werden, wie *SIP*, *DIP* sowie *AIP* oder *Designated Community*.

3.1. Verhandeln und Aufnehmen von Information

Die erste Anforderung bezieht sich unmittelbar auf den Prozess der Datenaufnahme bzw. dessen Vorbereitung, die in der Praxis häufig als *Pre-Ingest* bezeichnet wird. Meist handelt es sich dabei um einen iterativen Prozess zwischen FDZ und Datengebenden, weshalb die im Folgenden näher erläuterten Punkte in der Praxis nicht als trennscharf und strikt lineare Abfolge zu sehen sind. Eine transparente Darstellung der Services und Kosten hilft FDZ und Datengebenden bei einer zuverlässigen Planung.⁶

3.1.1. Scope und Auswahl/Priorisierung

Möchten Forschende ihre Daten an ein FDZ übergeben, so muss zunächst geprüft werden, ob die Daten in den *Scope* des FDZ fallen. Um den Prozess für Datengebende möglichst transparent zu gestalten, sollte ein FDZ vorab seinen *Scope* definieren. Hierfür wird oft eine *Collection Policy* eingesetzt. Dieses Dokument umreißt die Grundsätze des Bestands und der (geplanten) Entwicklung des FDZ. So kann der Bestand eines FDZ kohärent aufgebaut und weiterentwickelt werden.

Als Beispiel einer *Collection Policy* können die Dokumente folgender Institutionen dienen:

- UKDS⁷: <https://ukdataservice.ac.uk/app/uploads/cd234-collections-appraisal.pdf>
- ICPSR⁸: <https://www.icpsr.umich.edu/web/pages/datamanagement/policies/colldev.html>
- GESIS⁹: https://www.gesis.org/fileadmin/upload/Datenservices/Collection_Policy/2023-05-25_Collection_policy_dt.pdf
- VerbundFDB¹⁰: <https://www.forschungsdaten-bildung.de/collectionpolicy>
- FDZ Bildung¹¹: <https://www.fdz-bildung.de/collection-policy-fdz>
- AUSSDA¹²: https://aussda.at/fileadmin/user_upload/p_aussda/Documents/AUSSDA_Data_Collection_Policy.pdf

Es kann zudem sein, dass zwischen der Aufnahme von verschiedenen Daten, die alle Kriterien des Scopes erfüllen, abgewogen und eine Auswahl getroffen bzw. eine Priorisierung (*Appraisal*) vorgenommen werden muss, wenn beispielsweise aus Ressourcengründen des FDZ nicht alle Daten zu einem bestimmten Zeitpunkt aufgenommen werden können. Eine Auswahl oder Priorisierung von Daten in einem FDZ sollte integer, konsistent, transparent und

⁶ Siehe beispielsweise ZPID: <https://rdc-psychology.org/de/service-katalog-fdm> oder GESIS: <https://www.gesis.org/datenservices/daten-teilen>

⁷ United Kingdom Data Service: <https://ukdataservice.ac.uk/>

⁸ Inter-University Consortium for Political and Social Research: <https://www.icpsr.umich.edu>

⁹ GESIS – Leibniz-Institut für Sozialwissenschaften: <https://www.gesis.org/>

¹⁰ Verbund Forschungsdaten Bildung: <https://www.forschungsdaten-bildung.de/>

¹¹ FDZ Bildung: <https://www.fdz-bildung.de/home>

¹² The Austrian Social Science Data Archive: <https://aussda.at/>

nachvollziehbar sein. Daher sollten die Kriterien und der Prozess vorab klar definiert sein. Grundsätzlich sollten immer die Zielsetzung der Institution (z. B. Satzung, Leitbild, ...) und die Zielgruppe(n) sowie deren Erwartungen und Bedürfnisse berücksichtigt werden (Mauer, 2016). Ein solches Konzept sollte demnach nicht vom FDZ allein erstellt werden. Die verschiedenen Stakeholder (Datennutzende, Beirat, Fachgesellschaften, Forschungsförderer, Forschungspolitik, ...) sollten in die Entwicklung des Auswahlprozesses einbezogen werden. Whyte und Wilson (2010) geben Hinweise, welche Aspekte bei der Erstellung eines Konzepts zu Bewertung und Auswahl berücksichtigt werden sollten:

1. Bedeutung in Bezug auf den eigenen Auftrag (Relevance to Mission)
2. Wissenschaftlicher oder historischer Wert (Scientific or Historical Value)
3. Einmaligkeit (Uniqueness)
4. Möglichkeiten und Grenzen der Zurverfügungstellung (Potential for Redistribution)
5. Nicht-Reproduzierbarkeit (Non-Replicability)
6. Wirtschaftlicher Aspekt (Economic Case)
7. Ausführliche Dokumentation (Metadaten) (Full Documentation)

Auch Weber und Piesche (2021) haben eine Checkliste als Hilfestellung zur Prüfung der Archivwürdigkeit von Forschungsdaten zusammengestellt (Weber & Piesche, 2021, 335–336, 346–347):

- Bestehen Vorgaben Dritter (Fördergeber, Datenpolicies, Richtlinien der Forschungseinrichtung), die es notwendig machen, die Daten langfristig aufzubewahren?
- Hat man die notwendigen Nutzungsrechte an den Daten? Unter welchen Bedingungen besitzt man die Daten?
- Sind die erhobenen Daten einmalig und nicht reproduzierbar oder sind die Kosten der Reproduktion höher als die Kosten der Langzeitaufbewahrung?
- Liefert die Datenerhebung durch den technologischen Fortschritt voraussichtlich keine besseren Ergebnisse?
- Gibt es ein hohes Nachnutzungsinteresse an den Forschungsdaten?
- Wurden die Daten noch nicht vollständig wissenschaftlich untersucht?
- Sind die Daten charakteristisch oder untypisch für ein Forschungsgebiet bzw. handelt es sich um einmalige Forschungsergebnisse?
- Haben die Daten möglicherweise eine allgemeine oder regionale historische Bedeutung?
- Ist die Datenqualität technisch und inhaltlich gut?
- Sind deskriptive Metadaten vollständig vorhanden oder können generiert werden?
- Können die notwendigen Erhaltungsmetadaten (Referenz-, Provenienz-, Kontext- und Persistenz-Informationen sowie Angaben zu Zugriffsrechten) geliefert werden?

Die Bewertung bezieht sich nach den rechtlichen und dokumentarischen Mindestvoraussetzungen insbesondere auf die Einmaligkeit der Daten und der potenzielle Nutzen für die Forschung. Auch eventuelle zeitliche Verpflichtungen zur Veröffentlichung

seitens der Datengebenden sollten berücksichtigt werden, um eine fristgerechte Veröffentlichung zu prüfen bzw. zu ermöglichen.

Beispiele¹³:

- ICPSR: <https://www.icpsr.umich.edu/web/pages/datamanagement/lifecycle/selection.html>
- UK Data Service: <https://www.ukdataservice.ac.uk/media/455175/cd234-collections-appraisal.pdf>
- FDZ am ZPID: <https://rdc-psychology.org/de/datenauswahl>
- IQB: https://www.iqb.hu-berlin.de/fdz/Datenuebergabe/Template_Ingest_.pdf
(In dieser Checkliste sind auch Punkte enthalten, die auf die Prüfung der Erfüllung der Mindestanforderungen bei Dateneingang abzielen, vgl. Kapitel 3.1.2 und 3.1.3)

Sofern sich ergibt, dass ein Datenbestand gar nicht oder nicht zeitnah aufgenommen werden kann, kann auf andere Archivierungsmöglichkeiten hingewiesen werden, beispielsweise auf andere FDZ oder Verbünde (vgl. Kapitel 4) oder ggf. auch auf Self-Service Forschungsdatenrepositorien (z. B. Repositorien wie GESIS Archivierung Basis¹⁴ oder RADAR¹⁵).

3.1.2. Mindestanforderungen an die Daten und Metadaten

Ein FDZ sollte zur Sicherstellung der Datenqualität und der Einhaltung rechtlicher Vorgaben bestimmte Mindestanforderungen an die Daten und Metadaten, die für eine Übernahme erfüllt sein müssen, klar definieren. Diese sollten – idealerweise schon während der Antragsstellung bzw. Planung eines Projekts inkl. eines Datenmanagementplans¹⁶ – den Datengebenden transparent kommuniziert werden, damit einer effizienten und rechtssicheren Datenaufnahme und Sekundärnutzung nichts entgegensteht. Zu denken ist an rechtliche, technische und inhaltliche Mindestanforderungen.

Beispiele für Mindestanforderungen:

- GESIS: <https://www.gesis.org/datenservices/ueber-die-datenservices/standards-und-workflows-datenservices/vorbereitung-datenuebergabe-pre-ingest>
- AUSSDA: https://aussda.at/fileadmin/user_upload/p_aussda/Documents/DataDeposit-Guideline_SUF_v2_0.pdf
- IQB: https://www.iqb.hu-berlin.de/fdz/Datenuebergabe/Template_Ingest_.pdf

¹³ Weitere Ressourcen: *CESSDA Data Archiving Guide*: <https://dag.CESSDA.eu/Chapter-3/3-Data-review-and-appraisal> sowie *DCC How-To-Guide*: <https://www.dcc.ac.uk/guidance/how-guides/appraise-select-data#5>

¹⁴ <https://data.gesis.org/sharing> als sozialwissenschaftliches Repositorium

¹⁵ <https://www.radar-service.eu/> als generisches Repositorium

¹⁶ beispielweise <https://forschungsdaten.info/themen/informieren-und-planen/datenmanagementplan/> bzw. speziell für die Bildungsforschung <https://www.forschungsdaten-bildung.de/stamp-nutzen> sowie auch Netscher und Jensen (2019b, S. 38–40) zur Planung von Forschungsdatenmanagement und Uwe Jensen (2012).

a) Rechtlich

Die Erfüllung der rechtlichen Mindestanforderungen eines FDZ an eine Datenaufnahme entscheidet über die grundsätzliche Archivierbarkeit der Daten und Metadaten. Besonders zentral ist dabei, dass die Vereinbarungen mit allen Projektbeteiligten (Zuwendungsgeber*in, Auftraggeber*in, Untersuchungspersonen ...) so gestaltet sein müssen, dass einer Weitergabe der Daten in ein FDZ rechtlich nichts entgegensteht. Da die rechtlichen Aspekte für Datengebende oft eine Herausforderung darstellen, kann es sich für FDZ lohnen, dabei Hilfestellung zu leisten.¹⁷

Basieren die Forschungsdaten auf der Abgabe von Informationen von Personen (z. B. Online-Survey, Interview, ...), sind oft Fragen des Datenschutzes¹⁸ zu klären, da auch personenbezogene Daten erhoben werden. Die Datenerhebung erfolgt in der Regel auf Basis einer informierten Einwilligung der Untersuchungspersonen (*Informed Consent* oder auch Einverständniserklärung), berechtigtem Interesse, oder einer speziellen Rechtsvorschrift bei amtlichen Daten. Ein *Informed Consent* ist dahingehend zu prüfen, ob er den gesetzlichen Bestimmungen entspricht, die für die *Datenerhebung* gegolten haben (z. B. der DSGVO, BDSG oder LDSG bei personenbezogenen Daten). Zudem darf eine Sekundärnutzung der Daten nicht ausgeschlossen sein. Darüber hinaus können auch weitere Aspekte, insbesondere ethische Abwägungen, eine Rolle spielen, die für jeden Datenbestand zu berücksichtigen sind. Informierte Einwilligungen müssen auf diese rechtlichen und ethischen Aspekte hin geprüft werden, um die Rechte der Untersuchungspersonen bzw. im Kontext des Datenschutzes der Betroffenen zu wahren. Hinweise an die Anforderungen an *Informed Consents* finden sich etwa beim Verbund Forschungsdaten Bildung (2018) und bei Meyermann und Porzelt (2019). Zugleich sind die Interessen der Forschung zu berücksichtigen, dies auch unter Berücksichtigung der Privilegien, die ihr die Datenschutzgesetze einräumen. Das FDZ muss im Rahmen der Datenaufnahmen beide Perspektiven angemessen berücksichtigen.

An dieser Stelle seien exemplarisch Differenzierungen bzgl. der Nachnutzungsmöglichkeiten von anonymisierten und personenbezogenen Daten für Forschungszwecke genannt. Im Fall von Datenerhebungen auf Basis von *Informed Consents* sollte die Weitergabe der Daten an Dritte (personenbezogen oder anonymisiert) mindestens nicht explizit ausgeschlossen werden. Wird die Nachnutzung von anonymisierten Daten nicht explizit ausgeschlossen, so ist deren Nachnutzung rechtlich erlaubt, sofern keine weiteren Gründe dagegensprechen. Forschungsethisch begründet sollte eine geplante Nachnutzung gegenüber den untersuchten Personen im *Informed Consent* aber auch bei einer Nachnutzung nur der anonymisierten Daten kommuniziert werden. Aussagen zum Ausschluss der Nutzung von anonymisierten Daten können dagegen im Rahmen eines *Informed Consents* vertragsrechtlich interpretiert werden und so eine Nachnutzung ausschließen (Meyermann & Porzelt, 2019, S. 17). Die Nachnutzung

¹⁷ Konkret zu rechtlichen Fragen zur Datenaufnahme siehe auch das Erläuterungsdokument von Schallaböck, Kreuzer, Hoffstätter und Buck (2023b). Weitere Ressourcen zur Unterstützung finden sich insb. zur Bildungsforschung bei <https://www.forschungsdaten-bildung.de> und bei <https://rdm-compas.org/>.

¹⁸ Siehe hierzu auch insbesondere Watteler und Ebel (2019) (Kapitel 4.3.1 und 4.3.2).

von personenbezogenen Daten ist dagegen zunächst regulär nur mit entsprechender Erklärung im *Informed Consent* rechtlich möglich (Art. 7 DSGVO; ausschließlich datenschutzrechtlich bewertend bezieht sich die DSGVO nur auf personenbezogene Daten). Zugleich sollte deren Sekundärnutzung forschungsethisch kritisch geprüft werden. Zusätzlich sind für den Forschungskontext insbesondere Art. 89 DSGVO und § 27 BDSG zu benennen, die Garantien sowie Ausnahmen bezüglich der Verarbeitung personenbezogener Daten für im öffentlichen Interesse liegende wissenschaftliche, historische Forschungs- und Archivzwecke sowie statistische Zwecke regeln. Deren Anwendung erfordert aber eine gründliche Abwägung zwischen den Interessen der Forschung und der Betroffenen bzw. der Untersuchungspersonen.

Häufig ergibt sich aus der rechtlichen Prüfung, dass eine Weitergabe zur Archivierung und/oder Sekundärnutzung nur für anonymisierte¹⁹ Daten möglich ist, sodass entsprechende Anonymisierungsarbeiten bei der Datenaufbereitung anfallen (vgl. dazu auch Abschnitt c).

Des Weiteren sollte sichergestellt sein, dass mit der Übergabe der Daten und Metadaten keine Rechte Dritter verletzt und die gute wissenschaftliche Praxis etwa bzgl. der Zitation von nachgenutztem Material eingehalten wird.²⁰ So haben die Datengebenden dafür Sorge zu tragen, dass in den Daten, im Erhebungsinstrument oder anderen Unterlagen eventuell verwendete Frageformulierungen, Skalen, Itembatterien u. Ä. korrekt zitiert werden und ggf. bestehende Nutzungsbedingungen (bzw. Lizenzbedingungen) eingehalten werden. Häufig werden sich diese im wissenschaftlichen Kontext auf Zitationspflichten begrenzen. In Ausnahmefällen können mit der Nutzung aber auch beispielsweise Zahlungsverpflichtungen oder Verbote einer öffentlichen Weitergabe und Vervielfältigung einhergehen. Dies kann etwa bei Erhebungsinstrumenten der Fall sein, die auch für die psychologische Diagnostik genutzt werden. Aufgrund des erst wachsenden Bewusstseins in der Community empfiehlt es sich hier als FDZ, die Datengebenden darauf hinzuweisen und zu unterstützen. Die Beurteilung, ob an Forschungsdaten und/oder begleitenden Metadaten (z. B. Dokumentationsmaterialien wie Erhebungsinstrumente) Urheberrechte bestehen, kann von Fall zu Fall variieren.²¹ Zentral für die rechtssichere Datenaufnahme ist, dass – sollten urheberrechtlich relevante Bestandteile vorliegen – die entsprechenden Nutzungsrechte über eine Übergabevereinbarung an das FDZ übertragen werden (vgl. Abschnitt 3.2).

In der Übergabevereinbarung (vgl. Kapitel 3.2) sollten zudem auch weitere Aspekte geklärt werden, beispielsweise die Bereitstellungswege der Daten. Sofern das FDZ verschiedene Datenbereitstellungswege für Datennutzende anbietet, sollte in diesem Rahmen auch geklärt

¹⁹ Siehe zur Anonymität von Forschungsdaten auch Erläuterungsdokument Version 2.0.0 (Frage 1.1) (Schallaböck, Hoffstätter, Buck & Linne, 2023).

²⁰ Dies schließt auch nachgenutzte und übergebene gemeinfreie Daten ohne Urheberrechte und Lizenzen ein (Lauber-Rönsberg (2021, S. 93–96).

²¹ "Zusammenfassend lässt sich feststellen, dass es für das FDM problematisch ist, dass die Schutzfähigkeit einzelner Forschungsdaten in der Regel nur im Einzelfall und selbst dann nicht mit hinreichender Rechtssicherheit beurteilt werden kann." schreiben Lauber-Rönsberg, Krahn und Baumann (2018, S. 3) in einem Gutachten zu den rechtlichen Rahmenbedingungen des Forschungsdatenmanagements. Siehe dazu insbesondere auch das Erläuterungsdokument von Schallaböck, Kreuzer, Hoffstätter und Buck (2023a), in dem rechtliche Fragen von Forschungsdatenzentren beantwortet werden.

werden, welche Bereitstellungswege für die Daten und Metadaten gewählt werden sollten. Daten mit sensiblen Angaben (z. B. geografische Informationen zu Befragten) sollten etwa über restriktivere Datenbereitstellungswege zur Verfügung gestellt werden. Eine einführende Handreichung dazu, wie Daten zugänglich gemacht werden können, findet sich bei Hoffstätter und Linne (2022).

b) Technisch-formal

Neben grundlegenden technisch-formalen Mindestanforderungen, wie der Vollständigkeit, Intaktheit und Virenfreiheit aller übermittelten Dateien sowie der Maschinenlesbarkeit der Daten, sind besonders zentral die zugelassenen Dateiformate. Für die Langzeitarchivierung sowie für den Erhalt von Interpretierbarkeit und Nutzbarkeit der Dateien sollten möglichst offene Formate genutzt werden, da diese über die Zeit beständiger sind. Die Forschungsdaten können dabei von den Datengebern in Dateiformaten, die in der Community weit verbreitet sind, übergeben werden. Das FDZ sollte dazu eine Liste an akzeptierten Dateiformaten veröffentlichen.²² Zentral ist, dass das FDZ mit den entsprechenden Dateiformaten vertraut ist und sie ggf. auch in andere Formate (Migration) oder technische Umgebungen (Emulation) überführen kann. Darüber hinaus können auch Vorgaben zur Dateibenennung²³ sinnvoll sein.

c) Inhaltlich

Um die Daten sinnvoll und langfristig nachnutzen und interpretieren zu können, sind auch inhaltliche Mindestanforderungen im Hinblick auf die Datenaufbereitung und Metadaten notwendig. Bei der Datenaufbereitung von quantitativen Umfragedaten sind dabei beispielsweise die Vergabe von korrekten und verständlichen Variablennamen sowie Variablen- und Wertelabels, die Codierung offener Angaben und fehlender Werte (Missings), die Datenkontrolle (z. B. Plausibilitäts-/Konsistenzprüfung) und -bereinigung, die Generierung neuer Variablen sowie die Anonymisierung (z. B. durch Aggregation von Werten oder Löschung von Variablen) zu nennen.

Zentrale Metadaten zur inhaltlichen Verständlichkeit der Daten sind beispielsweise das Erhebungs-/ Messinstrument (z. B. Fragebogen, Interviewleitfaden), ein Daten- und Methodenbericht (mit Informationen zu z. B. Inhalt/Anlage der Studie, Grundgesamtheit, Stichprobenziehung/Sampling, Durchführung der Erhebung, Rücklauf, vorgenommene Datenaufbereitungsschritte),²⁴ Codebuch/Variablenreport (bei quantitativen Erhebungen)

²² Beispiele und Hilfestellungen finden sich bei AUSSDA (Butzlaff, 2022, S. 8), GESIS (<https://www.gesis.org/datenservices/ueber-die-datenservices/standards-und-workflows-datenservices/vorbereitung-datenuebergabe-pre-ingest>), UKDS (<https://ukdataservice.ac.uk/learning-hub/research-data-management/format-your-data/recommended-formats/>), Verbund Forschungsdaten Bildung (<https://www.forschungsdaten-bildung.de/dateiformate>) sowie Weber und Piesche (2021, S. 347–349).

²³ Siehe dazu beispielsweise zu Dateioorganisation Recker und Brislinger (2019) oder auch <https://www.forschungsdaten-bildung.de/dateien-benennen#Benennung-von-Dateien>

²⁴ s. beispielsweise Watteler (2010).

sowie ggf. verschiedene strukturierte beschreibende Metadaten²⁵ (abhängig von den verwendeten Systemen und Metadatenstandards, vgl. Abschnitt 3.4).

3.1.3. Datenübertragung und Eingangsprüfung

Die – gesammelte oder auch iterative – Übertragung der Daten und Metadaten an das FDZ kann grundsätzlich über verschiedene Wege erfolgen, z. B. auf einem Datenträger (wie USB-Stick oder externe Festplatte), über eine kollaborative Arbeitsplattform oder über ein speziell hierfür vorgesehenes digitales Tool.²⁶ Ein FDZ sollte die möglichen Übertragungswege klar an die Datengebenden kommunizieren. Unbedingt darauf zu achten ist, dass jederzeit der Schutz der Daten gewährleistet ist, etwa mithilfe von speziell gesicherten Diensten und Verschlüsselungstechnologien.²⁷

Nach Eingang der Daten und Metadaten im FDZ sollte geprüft werden, ob sie – orientiert an den vorab definierten, FDZ-spezifischen Mindestanforderungen an die Daten und Metadaten (vgl. Abschnitt 3.1.2) – alle Kriterien für eine Datenaufnahme erfüllen. Es wird dabei empfohlen, eine Checkliste anzulegen, anhand der alle relevanten Punkte nach und nach strukturiert abgeprüft und das Prüfergebnis schriftlich dokumentiert und nachvollzogen werden kann. Das genaue Vorgehen bei der Eingangsprüfung sowie die Granularität, in der sie vorgenommen wird, sind dabei ebenfalls FDZ-spezifisch festzulegen.

Beispiele:

- FDZ am IQB
 - technisch-formal: https://www.iqb.hu-berlin.de/fdz/Datenuebergabe/00_Checkliste_te.pdf
 - inhaltlich: https://www.iqb.hu-berlin.de/fdz/Datenuebergabe/Template_Ingest_.pdf
(in dieser Checkliste ist auch die Prüfung der Passung zum inhaltlichen Scope sowie die Prüfung des allgemeinen Nachnutzungspotenzials (vgl. Abschnitt 3.1.1) enthalten)
 - initiale Prüfung der Daten: (IQB – Institut zur Qualitätsentwicklung im Bildungswesen, 2021a): https://www.iqb.hu-berlin.de/fdz/Datenuebergabe/Vorlage_AbkStudi.pdf

²⁵ Strukturierte/standardisierte Metadaten beschreiben Daten mit einem einheitlichen Vokabular, beispielsweise Angaben zur Stichprobenziehung nach dem Metadatenstandard der *Data Documentation Initiative* (DDI) oder mit Messkonzepten, wie z. B. die *International Standard Classification of Education* (ISCED 2011) für Schultypen und Schulsysteme, (Netscher & Jensen, 2019a, S. 40; Wallace, 2001, S. 255). Im Gegensatz dazu sind semi-strukturierte oder unstrukturierte Metadaten fließtextbasiert und liefern detailliert den nötigen Kontext, z. B. in Form von Daten- und Methodenberichten oder Fragebögen. Die Nutzung beider Arten von Metadaten ist in den Sozialwissenschaften und angrenzenden Wissenschaften üblich.

²⁶ Die Übermittlung via unverschlüsselter E-Mail ist kein sicherer Übertragungsweg. Von externen Diensten wird – „speziell, wenn deren Nutzung mit Datenverarbeitungen außerhalb der Europäischen Union verbunden ist“ (z. B. bei Dropbox oder Onedrive) – abgeraten entsprechend RatSWD (2020, S. 23).

²⁷ Siehe hierzu beispielsweise auch <https://rdm-compas.org/articles/datenuernahme-2>.

- GESIS:
 - <https://www.gesis.org/datenservices/ueber-die-datenservices/standards-und-workflows-datenservices/datenuebergabe-ingest> bzw.
 - https://www.gesis.org/fileadmin/upload/Datenservices/Eingangskontrolle/Dateneingangskontrolle_n_GESIS.xlsx
- FDZ Bildung: <https://www.fdz-bildung.de/datenkuratierung>
- GESIS Archivierung BASIS: <https://rdm-compas.org/articles/tipps-checklisten-auswahlbewertung#accordion-beispiel-sowidatanet-checkliste-fur-die-dateneingangskontrolle-von-forschungsdaten3-%F0%9F%93%9D>

Oft werden in einem iterativen Prozess mit den Datengebenden die Daten korrigiert oder die Metadaten präzisiert oder erweitert. Ggf. sollte dabei auch vereinbart werden, ob und welche Überarbeitungen durch das FDZ übernommen werden (vgl. Kapitel 3.1.4).

3.1.4. Aufnahme und Weiterverarbeitung

Wenn die eingereichten Forschungsdaten und ihre Metadaten für eine Aufnahme in ein FDZ geeignet sind, müssen die entsprechenden Dateien als SIP schreibgeschützt in die Infrastruktur des FDZ überführt und technisch (z. B. via Checksummenprüfung²⁸) vor (auch unbeabsichtigter) Bearbeitung gesichert werden.²⁹ Die Speicherkapazität muss vom FDZ bereitgestellt werden. Wichtig ist auch, dass es im FDZ eine klare Ordner- und Dateistruktur sowie Dateibenennung, nach der die übergebenen Daten und Metadaten abgelegt werden, sowie eine funktionierende Versionskontrolle gibt.³⁰

Dazu werden zunächst relevante Elemente des SIP kopiert und bilden die Grundlage des AIP. Welche Elemente des SIP ins AIP gelangen, definiert das FDZ vorab. Die Zusammensetzung des AIP variiert dabei je nach Aufbereitungs-, Dokumentations- und Publikationsziel einer Studie (Mauer, 2012, S. 209). Zudem werden Daten und Metadaten ggf. noch editiert³¹, z. B. indem eine Anonymisierung³², Aggregation oder andere Änderungen an den Daten und Metadaten (z. B. Ergänzung des *Persistent Identifiers* zur Veröffentlichung) durchgeführt werden.³³ Ob das FDZ oder die Datengebenden diese Aufgaben übernehmen sollte im Zuge der Eingangsprüfung verhandelt und klar kommuniziert bzw. insbesondere bei rechtlich relevanten Aspekten (wie Verantwortung für eine Anonymisierung) auch in der Übergabvereinbarung festgehalten werden (vgl. Kapitel 3.2). In der Regel findet im AIP außerdem eine Anreicherung

²⁸ Die Checksumme einer Datei ist vergleichbar mit einem "digitaler Fingerabdruck" der Datei. Jede noch so kleine Änderung führt zu einer veränderten Checksumme. Siehe dazu beispielsweise <https://www.dpconline.org/handbook/technical-solutions-and-tools/fifty-and-checksums>

²⁹ Tipps hierzu auch unter: <https://rdm-compas.org/articles/datenubernahme-2>

³⁰ Tipps zur Dateioorganisation finden sich unter <https://rdm-compas.org/articles/fdm-hinweise-forschungsdaten-organisieren>, <https://www.forschungsdaten-bildung.de/dateien-benennen> sowie Recker und Brislinger (2019) (Abschnitt 5.3).

³¹ Siehe zur Datenaufbereitung und Dokumentation auch Brislinger und Moschner (2019) sowie <https://rdm-compas.org/articles/aufbereitung-qualitativer-und-quantitativer-daten>

³² Siehe zu Anonymisierung beispielsweise Ebel und Meyermann (2015) für quantitative Daten oder Meyermann und Porzelt (2014) für qualitative Daten.

³³ Zu initialen Prüfschritten und Datenaufbereitung siehe z. B. IQB – Institut zur Qualitätsentwicklung im Bildungswesen (2021a). Siehe zudem auch Beispiele unter 3.1.2 c.

der SIP-Dateien durch FDZ-Mitarbeitende statt – um die Übergabvereinbarung, andere rechtliche Dokumente (z. B. *Informed Consent* sowie dessen Bewertung), weitere Informationen (z. B. Dokumentation des *Appraisal*, Dokumentation der Kommunikation) sowie um strukturierte Metadaten, die für die Langzeitarchivierung wichtig sind (vgl. Kapitel 3.4). Alle Änderungen an den Daten und den Metadaten, die im AIP stattfinden, werden in einer Historiendatei dokumentiert und ebenfalls dem AIP beigefügt.

Beispielhaft nennt Mauer (2012, S. 209) für ein AIP bei GESIS:

- SIP: von Datengebenden übernommene Original-Objekte (zum Beispiel Datensätze, Dokumente) bzw. Kopien davon
- im Archiv aufbereitete Datensatzversionen
- Aufbereitungssyntax, die den Bezug zwischen Original- und Archivversion herstellen, sowie weitere Dokumentation der Aufbereitung
- im Archiv erzeugte Metadaten zur Beschreibung der Daten sowie weitere technische und administrative Metadaten

Beispielhaft beschreibt das FDZ am IQB (IQB – Institut zur Qualitätsentwicklung im Bildungswesen, 2021b) das AIP wie folgt:

- Originaldaten (=SIP), auch in langzeitverfügbaren Formaten
- bereitgestellte und zur Verfügung gestellte (Forschungs-)Daten (=DIP), auch in langzeitverfügbaren Formaten
- vorliegende Metadaten, auch in langzeitverfügbaren Formaten
- alle Dokumentations- und Aufbereitungsschritte (Aufbereitungssyntaxen, ...)
- weitere Dokumente (Bewertung der Daten, Verträge, Korrespondenz, Verträge, Zugriffskonzept)
- Checksummen pro digitalem Objekt

Ein noch konkreteres Beispiel der Ordnerstruktur eines AIP findet sich ebenfalls beim FDZ am IQB (IQB – Institut zur Qualitätsentwicklung im Bildungswesen, 2021b, 11f).

3.2. Ausreichende Kontrolle für die Langzeitarchivierung erlangen

Das FDZ sollte sicherstellen, dass mit den Datengebenden eine Übergabvereinbarung geschlossen wird, in der die rechtlichen, technischen und organisatorischen Aspekte der Datenaufnahme geregelt sind³⁴ und festgehalten ist, wie die SIP gestaltet sind, d. h. welche Dateien in welchen Formaten übergeben wird. Dies geschieht in der Regel über einen Datenaufnahmevertrag³⁵ teilweise aber auch in Form von Allgemeinen Geschäftsbedingungen oder Nutzungsbedingungen. Vor der rechtlich verbindlichen Übergabe der Daten müssen die Datengebenden klären, wer formal die Inhaber*innen der Urheber- und Nutzungsrechte sind. Nur die Partei, die die Nutzungsrechte hält, kann die Übergabe rechtsgültig durchführen und

³⁴ Siehe beispielsweise hierzu auch <https://rdm-compas.org/articles/rechtliches-datenubertragung-erstellung-und-empfang>

³⁵ auch Datenüberlassungsvertrag, Datenarchivierungsvertrag o. Ä. genannt.

damit ebendiese Rechte übertragen.³⁶ Ein harmonisierter Mustervertrag für eine Datenaufnahme in ein FDZ sowie ein Erläuterungsdokument zu verschiedenen rechtlichen Fragen wurden im Rahmen von KonsortSWD erarbeitet und veröffentlicht (Schallaböck, Kreuzer, Hoffstätter & Buck, 2023b).³⁷

Besonders zentral ist dabei, dass bei einer Archivierung von Forschungsdaten dem FDZ i. d. R. sogenannte einfache, also keine ausschließlichen, Nutzungsrechte übertragen werden. Sofern am übergebenen Material Urheberrechte bestehen, verbleiben diese bei den Urheber*innen (Ausnahme lediglich Tod der Urheber*innen und Erbfall) (§29 des Urheberrechtsgesetzes). Dem FDZ werden, im Detail abhängig von den Vertragsbedingungen, in der Regel lediglich die einfachen Nutzungsrechte, Daten und Metadaten zu speichern, zu vervielfältigen und an Dritte weiterzugeben sowie insbesondere das Recht, die digitalen Objekte (Daten, Metadaten etc.) zur Sicherung der langfristigen Nutzbarkeit in andere Formate zu transformieren, gewährt.³⁸ Im Fall von übertragenen einfachen Nutzungsrechten ist es den Datengebenden zudem möglich, den Vertrag zu kündigen. In dem Fall dürfte das FDZ die Daten nicht weiter halten und nutzen. Die grundsätzliche Entscheidungsmacht über die Nutzung der Daten verbleibt letztlich bei den Datengebenden. Zu betonen ist, dass nicht nur an den Forschungsdaten potenziell Nutzungsrechte bestehen können. Auch bei Erhebungsinstrumenten (z. B. bei Verwendung psychologischer Skalen bei Fragebögen) sind ggf. qua Nutzungsrechten (bzw. Lizenzen) bestehende Ansprüche Dritter zu prüfen (vgl. zu rechtlichen Aspekten insbesondere Abschnitt 3.1.2 a sowie die dortigen Referenzen).

Wenn die Daten an derselben Institution generiert wurden, an der auch das FDZ angesiedelt ist, so ist es meist rechtlich nicht möglich, einen Datenaufnahmevertrag abzuschließen. Dies muss jedoch im Einzelfall geprüft werden.³⁹ Dennoch sollte auch hier sichergestellt werden, dass eine Weitergabe und Veröffentlichung von Daten und Metadaten nicht zu rechtlichen Problemen für die Institution führen und/oder nicht im Sinn der Datengebenden sind. Ebenso sollten Verpflichtungen der Datengebenden gegenüber Dritten, etwa bzgl. Veröffentlichungsverpflichtungen und Fristen, erfasst werden, um diese möglichst in der Planung des FDZ zu berücksichtigen. Denkbar ist hierfür beispielsweise die Erstellung eines Dokuments, das von den Datengebenden unterzeichnet wird und dessen Inhalte vom FDZ berücksichtigt werden müssen. So können die jeweiligen Rollen und Verantwortungsbereiche festgelegt werden.

³⁶ siehe Lauber-Rönsberg (2021, S. 90–93).

³⁷ Basis waren die Verträge von zehn FDZ, die diese auf freiwilliger Basis für die Konzeption eines harmonisierten Vertrags zur Verfügung stellten, sowie darauf aufbauend die Expertise der Kanzlei iRights.Law. Im Erläuterungsdokument wird neben Erläuterungen zum Vertrag auch auf konkrete Fragen von Forschungsdatenzentren im Bereich Datenaufnahme eingegangen.

³⁸ Siehe auch <https://www.gesis.org/datenservices/ueber-die-datenservices/standards-und-workflows-datenservices/vorbereitung-datenuebergabe-pre-ingest> (Urheberrecht)

³⁹ Eine Einzelfallprüfung ist hier notwendig, da dies z. B. je nach Förderung, Arbeitsverhältnis etc. variieren kann. Siehe z. B. auch Wünsche, Soßna, Kreitlow und Voigt (2022), welche daher vor Projektbeginn eine Rechtklärung empfehlen.

3.3. Zielgruppe bestimmen

Es ist notwendig, dass ein FDZ seine vorgesehene Zielgruppe (*Designated Community*), also die erwarteten Endnutzer*innen, bestimmt, weil nur so auch das Grundwissen der Zielgruppe definiert und daran anknüpfend die Verstehbarkeit der bereitgestellten Informationen für diese sichergestellt werden kann (Keitel & Mitcham, 2023; OAIÜbersetzung/Terminologie, 2013, 16, 19, 30; vgl. dazu auch Kapitel 3.4).

Wichtig ist, dass bei der Definition der Zielgruppe auch mitbedacht wird, dass sich diese über die Zeit verändern kann (OAIÜbersetzung/Terminologie, 2013, S. 16). Beispielsweise wäre es denkbar, dass auch Forschende aus anderen Fachbereichen mit den Daten arbeiten möchten – etwa bei interdisziplinären Forschungsfragen und auch mit Blick auf die Verknüpfung verschiedener Datenbestände. Ebenso wäre es möglich, dass zunächst nur für ein nationales Publikum aufbereitete Daten auch für ein internationales Publikum interessant werden könnten. Darüber hinaus könnte es sein, dass bestimmte, zunächst nur für die reine Forschung gedachte Daten perspektivisch auch explizit für die Hochschullehre geöffnet werden. Ebenso könnte es auch eine Öffnung für Bereiche abseits von Forschung und Wissenschaft geben, beispielsweise für Journalismus, Wirtschaft oder die allgemeine Öffentlichkeit. Es sollte versucht werden, solche potenziellen Veränderungen direkt mitzuberücksichtigen, also die Zielgruppe idealerweise direkt etwas breiter zu fassen, da eine rückwirkende Sicherung der Verstehbarkeit für eine umfassendere Zielgruppe schwierig werden kann und Zusatzaufwand mit sich bringt (OAIÜbersetzung/Terminologie, 2013, S. 30–31).

Beispiele für Zielgruppen-Definitionen an FDZ aus dem Bereich der Sozial-, Verhaltens-, Bildungs- und Wirtschaftsdaten finden sich hier:

- FDZ am IQB: *„The FDZ at IQB's Designated Community consists of researchers from various disciplines in which educational processes are studied such as educational science, psychology, sociology, economics, and political science. Data access is granted to researchers from either universities or non-university research institutions, students being equally welcome.“* (https://www.iqb.hu-berlin.de/fdz/Grundlagen/CTS_with_Appendi.pdf, S. 4)
- FDZ-DZHW: *„Die zentrale Zielgruppe für die Datennutzung umfasst die nationale und internationale wissenschaftliche Gemeinschaft (Wissenschaftler*innen, Lehrende und Studierende) des Forschungsfeldes der Hochschul- und Wissenschaftsforschung sowie der angrenzenden Disziplinen. Darüber hinaus besteht eine Offenheit bezüglich weiterer Nutzer*innengruppen, sofern ihre Nutzungsintentionen den vorgesehenen Nutzungszwecken der Daten entsprechen.“* (<https://fdz.dzhw.eu/de/ueber-das-fdz>)
- AUSSDA: *„The primary beneficiaries of AUSSDA's services are social science researchers. Secondary user communities include students, educators, media representatives and the general public.“* (https://aussda.at/fileadmin/user_upload/p_aussda/Documents/CTS_Report_AUSSDA.pdf, S. 3)

- GESIS – Leibniz-Institut für Sozialwissenschaften (2024 [in Veröffentlichung]): „*GESIS’s services are “targeted primarily at scientists who work with methods of empirical social research, especially at universities and non-university research institutions in Germany and throughout the world” (“Target Groups”, <https://www.gesis.org/en/institute>). Accordingly, the Data Services are aimed at national and international researchers, academic teachers, and students in empirical social research with an emphasis on the areas of sociology and political science as well as social science in its entirety. Other target groups include those working in related political, social, and commercial social science environments. The services are offered to researchers (both in universities and non-university research institutions) and students.*

To be able to use the GESIS Data Services, users should bring the following competencies and skill sets:

- *Information literacy: ability to use online search tools (e.g. GESIS Search) to identify datasets that are suitable to meet the researcher’s needs*
- *Data literacy: “ability to read, understand, create, and communicate [quantitative social science] data as information” (https://en.wikipedia.org/wiki/Data_literacy).*
- *Statistical literacy: “The abilities to understand and reason with data, or arguments that use data [...]. [S]cientists also need to develop statistical literacy so that they can both produce rigorous and reproducible research and consume it.” (https://en.wikipedia.org/wiki/Statistical_literacy)*

This entails

- *the ability to read and understand documentation and metadata provided by GESIS Data Services to assess the usability and suitability of the data for the research problem;*
- *proficiency in standard quantitative methods for data collection and analysis in the social sciences at least on the level of a master’s degree to understand and work with the data.*
- *proficiency in using standard statistical software to render, process, and analyze the data, in particular STATA, SPSS.”*

3.4. Sicherstellen der unmittelbaren Verstehbarkeit von Information

Ein AIP enthält neben den Daten hauptsächlich Metadaten. Diese sind notwendig, um dauerhaft genügend Informationen über das AIP zu haben. Teilweise können die für eine Verstehbarkeit benötigten Metadaten bereits von den Datengebenden angefordert werden (vgl. Kapitel 3.1.2), teilweise müssen sie auch durch das FDZ ergänzt (insbesondere Metadaten für die Langzeitarchivierung) oder in ein strukturiertes Format überführt werden (vgl. Kapitel 3.1.4).

Die Erhaltungsmetadaten (*Preservation Metadata*) werden regelmäßig in die in Tabelle 1 dargestellten Metadatenklassen eingeteilt. Diese sind nicht immer trennscharf und es kann

somit zu inhaltlichen Überschneidungen kommen (Hein, 2012, S. 91; Jensen, Zenk-Möltgen & Wasner, 2019, S. 154; Keitel & Mitcham, 2023 angelehnt an Gartner, 2008):

*Tabelle 1: Metadatenklassen für die Erhaltung*⁴⁰

deskriptiv	Bibliographische Angaben, z. B. Titel, Autor*in sowie inhaltliche Angaben, z. B. Abstract, Erhebungsmethode
strukturell	Angaben zum Gesamtzusammenhang des archivierten Informationsobjektes, z. B. Beziehung von Dateien zueinander
administrativ	Informationen zum Archivierungsprozess, z. B. rechtlich (Lizenzen, Verträge, Zugangsrechte, Urheberrecht, ...), Provenienz (Entstehung, Dateihistorie, Bearbeiter*in, ...)
technisch	Großteilens automatisierte Metadaten, z. B. Dateiformate, Originalpfade, Dateigröße, Versionen, Größen, Checksummen

Deskriptive Metadaten geben Auskunft über den Inhalt und die Entstehung von digitalen Objekten und entscheiden somit auch über ihre Auffindbarkeit, Referenzierung und Nachnutzbarkeit. Einer der bekanntesten und zugleich vielfach verwendeten Metadatenstandards ist das *Dublin Core™ Metadata Element Set (DCMES)*⁴¹. Dieses beschreibt eine Web-Ressource mittels 15 optionaler Kernelemente (z. B. Urhebende, Titel, Beschreibung ...). Aber auch weitere Metadaten schemata und Vokabulare sollten z.B. für eine gute Auffindbarkeit der Daten⁴² genutzt werden. Insbesondere neuere Metadaten schemata wie *schema.org* oder disziplinspezifische wie der Standard der *Data Documentation Initiative (DDI)*⁴³, ein international anerkannter Standard für die Dokumentation von Daten, die durch Beobachtungsmethoden in den Sozial-, Verhaltens-, Wirtschafts- und Gesundheitswissenschaften erzeugt werden, sollten dabei genutzt werden (Limani et al., 2022).

Strukturelle Metadaten beschreiben die interne Struktur sowie Beziehungen zwischen den einzelnen Elementen. Dies können im einfachen Fall z. B. eine bestimmte Ordnerstruktur (z. B. einzelne Ordner für verschiedene Erhebungswellen) oder die Beziehung zwischen einzelnen Dokumenten zueinander sein. Für komplexere Dokumentstrukturen kann etwa das *METS (Metadata Encoding & Transmission Standard)*⁴⁴ angewendet werden.

Administrative Metadaten umfassen alle Informationen, die dem eigentlichen Archivierungsprozess, dem späteren Zugriff und der authentischen Wiedergabe der archivierten Ressourcen dienlich sind. Dies umfasst z. B. Urheberrechte und Lizenzinformationen (Wer darf unter welchen Bedingungen auf die Ressourcen Zugriff haben?). Zudem werden hier Informationen zur Herkunft und Archivierung, Prozesse, die die Ressource durchlaufen hat (Formatkonvertierung, ...) dokumentiert.

⁴⁰ Nach OAIS-Referenzmodell könnten hier die Erhaltungsmetadaten (*Preservation Description Information – PDI*) herangezogen werden:

https://wiki.dpconline.org/index.php?title=4.2.1.4.2_Preservation_Description_Information

⁴¹ <https://www.dublincore.org/specifications/dublin-core/dces/>

⁴² s. dazu auch Hoffstätter und Linne (2022).

⁴³ <https://ddialliance.org/>

⁴⁴ <http://www.loc.gov/standards/mets/mets-home.html>

Technische Metadaten (teilweise als Teilmenge von administrativen Metadaten) lassen sich auf Dateiebene inzwischen oft automatisiert ermitteln (z. B. Dateityp, Codierung, Erstellungsdatum, ...). Als Standard kann hier textMD⁴⁵ oder Elemente davon (z. B. Codierung, Sprache, ...) genutzt werden.

Weitere eher übergreifende Metadatenstandards (insbesondere für administrative und technische Metadaten) sind PREMIS („PREservation Metadata: Implementation Strategies“)⁴⁶ und LMER (Langzeitarchivierungsmetadaten für elektronische Ressourcen)⁴⁷. Eine Einordnung u. a. der Standards PREMIS, METS und LMER findet sich bei Dappert und Enders (2010). PREMIS ist vergleichsweise weit verbreitet.

Die Verwendung von Standards steigert im Allgemeinen die Interoperabilität. Setzen beispielsweise zwei Langzeitarchivsysteme auf denselben Metadatenstandard, lassen sich Informationen wie beispielsweise deskriptive Metadaten leicht austauschen. Die Einhaltung von Standards ermöglicht ebenso die Wiederverwendbarkeit der Informationen in anderen Zusammenhängen (z. B. bei der Auffindbarkeit bei Suchmaschinen) (Hein, 2012, S. 92). In der Community der FDZ in Deutschland wird aktuell (noch) kein Metadatenstandard für die Langzeitarchivierung einheitlich angewendet, einzelne FDZ haben jedoch das Vorhaben, PREMIS einzusetzen.⁴⁸

3.5. Bewährte Erhaltungspolicies und -abläufe einhalten

Für die langfristige Erhaltung eines AIP ist es notwendig, dass ein FDZ für sich entsprechende Erhaltungsabläufe definiert, die eigenen Grundsätze in *Preservation Policies* formuliert und dokumentiert und diese nicht zuletzt auch einhält (OAIS-Übersetzung/Terminologie, 2013, S. 32).

Institutionelle *Preservation Policies* sollten entsprechend neben technischen Strategien (*Technology Watch*) zur Langfristsicherung und -erhaltung auch Aspekte wie u. a. die Mission und Aufgabe des FDZ bzw. Archives, die Zielgruppen (*Community Watch*) (vgl. Kapitel 3.3), rechtliche Rahmenbedingungen (vgl. Abschnitt 3.1.2), die Zugriffskonzepte (z. B. sich verändernde Technik und mehr Zusatzwissen bzgl. Anonymität von Daten)⁴⁹ und das Sammlungsprofil (vgl. Kapitel 3.1.1) berücksichtigen sowie einen Notfallplan⁵⁰ und idealerweise

⁴⁵ <https://www.loc.gov/standards/textMD/>

⁴⁶ Einführung in PREMIS: <http://www.loc.gov/standards/premis/>; deutsch: https://www.loc.gov/standards/premis/understanding_premis_german.pdf.

PREMIS ist ein Standard und kein Tool oder eine Softwarelösung. Eine Liste mit Tools kann hier eingesehen werden: [https://coptr.digipres.org/index.php/PREMIS_\(Preservation_Metadata_Implementation_Strategies\)](https://coptr.digipres.org/index.php/PREMIS_(Preservation_Metadata_Implementation_Strategies))

⁴⁷ Die Langzeitarchivierungsmetadaten für elektronische Ressourcen (LMER) wurden von der Deutschen Bibliothek basierend auf Arbeiten der National Library of New Zealand entwickelt. Das Objektmodell basiert auf dem "Preservation Metadata: Metadata Implementation Schema" der Nationalbibliothek von Neuseeland (2003).

⁴⁸ Das Metadatenchema des FDZ am IQB wird beispielsweise hier transparent dokumentiert: <https://mdr.iqb.hu-berlin.de/#/catalog/c09f72c7-36cc-1580-b32f-605401c3c830>

⁴⁹ In einer Handreichung des RatSWD (2020, S. 19) zum Datenschutz wird etwas darauf hingewiesen, dass eine Anonymisierung in gewissen Zeitabständen reevaluiert werden sollte.

⁵⁰ Beispielsweise das *Disaster Planning* des ICPSR:

<https://www.icpsr.umich.edu/web/pages/datamanagement/disaster/index.html>

einen Nachfolgeplan bzw. Vorkehrungen für im Fall einer Auflösung des FDZ (nestor-Arbeitsgruppe Policy, 2014; VerbundFDB – AG Cessation, 2024 [in Veröffentlichung]) umfassen. Sie sollte in Rückkopplung mit den verschiedenen Stakeholdern abgestimmt werden. So werden Selbstvergewisserung und Selbstverpflichtung für die Einrichtung geschaffen. Zudem werden langfristig wirksame strategische und organisatorische Grundelemente eines digitalen Langzeitarchivs offengelegt und können so intern und extern zum allgemeinen Vertrauensgewinn beigetragen (nestor-Arbeitsgruppe Policy, 2014, S. 2).⁵¹ Eine *Preservation Policy* ist analog zum Prozess der Langzeitarchivierung insgesamt ein lebendes Dokument und wird immer wieder angepasst und weiterentwickelt.

Beispielhaft kann auf Konzepte der folgenden Institutionen verweisen werden:⁵²

- FDZ am IQB: (IQB – Institut zur Qualitätsentwicklung im Bildungswesen, 2021b):
<https://www.iqb.hu-berlin.de/fdz/Grundlagen/Langzeitverfuegb.pdf>
- GESIS:
https://www.gesis.org/fileadmin/upload/Datenservices/Preservation_Policy/Digital_Preservation_Policy_1.4.8.pdf
Preservation Policy ergänzt um einen *Preservation Plan* mit konkreten Maßnahmen
(https://www.gesis.org/fileadmin/upload/Datenservices/Preservation_Policy/Preservation_Plan_1-0-0.pdf)
- AUSSDA: (Bischof, 2022):
https://aussda.at/fileadmin/user_upload/p_aussda/Documents/Preservation_Plan_v1_0.pdf
- Deutsche Nationalbibliothek: (Schrimpf & Steinke, 2018):
<https://d-nb.info/1159789827/34>

Zur praktischen Umsetzung kann der schematische Überblick über die Handlungsebenen, Ziele und Maßnahmen für eine Langzeitarchivierung basierend auf Ideen von Thibodeau (2002), dargestellt in Tabelle 2, hilfreich sein. Im Folgenden wird das Konzept vorgestellt und um praktische Hinweise, insbesondere von Weber und Piesche (2021), ergänzt.

⁵¹ So kann eine *Preservation Policy* auch intern bei strategischen Entscheidungen herangezogen werden und auf die verschriftlichen Verpflichtungen verwiesen werden.

⁵² Ein ausführliche Schritt-für Schritt Leitfaden der *Digital Preservation Coalition* kann ebenfalls herangezogen werden: <https://www.dpconline.org/digipres/implement-digipres/policy-toolkit/policy-step>

Tabelle 2: Handlungsebenen, Ziele und Maßnahmen für die Langzeitarchivierung

Handlungsebene	Ziel	Maßnahme
physische Ebene (<i>Bitstream Preservation</i>)	Physischer Erhalt und Sicherung der Integrität der Daten	Regelmäßiger Austausch von Datenträgern, redundante Speicherung, Checksummenprüfung
logische Ebene (<i>Logical Preservation</i>)	Erhalt der technischen Interpretierbarkeit und Authentizität der Daten	Migration von Dateiformaten oder Emulation der ursprünglichen Systemumgebung
semantische Ebene (<i>Semantic Preservation</i>)	Erhalt der inhaltlichen Interpretierbarkeit und Wiederauffindbarkeit von Daten	Inhaltliche Beschreibung digitaler Ressourcen mittels Kontextinformationen

Basis und Voraussetzung für eine langfristige Erhaltung ist demnach die Sicherung der digitalen Dateien auf der physischen Ebene (*Bitstream Preservation*). Die Unversehrtheit wird bezogen auf die Datei selbst sowie deren Speichermedium sichergestellt. Die Integrität der Dateien kann durch die Nutzung von sog. *Fixity Checks* in Form einer Checksummen-Bildung geprüft werden (Weber & Piesche, 2021, S. 340).⁵³ Dabei wird aus den Daten mittels eines Algorithmus ein „Fingerabdruck“ generiert, welcher sich bei jeglicher Änderung an den Dateien verändert. Zudem muss eine regelmäßige Prüfung und ggf. ein Austausch der physischen Datenträger (Hard- und Softwarekomponenten), auf dem die Dateien liegen, durchgeführt werden. Darüber hinaus sollte eine redundante Sicherung⁵⁴ der Dateien implementiert werden. Bei der redundanten Sicherung auf unterschiedlichen Speicherbereichen dient die Checksumme zusätzlich dazu, die Gleichheit der Dateien zu gewährleisten, indem regelmäßig Checksummen der Dateien verglichen werden und bei Unregelmäßigkeiten eine gültige Kopie (Backup) zum Einsatz kommt (Weber & Piesche, 2021, S. 340). Zusätzlich zur redundanten Sicherung sollte eine Backup- und Disaster Recovery-Strategie aufgebaut werden, um ggf. Kopien von Dateien wiederherstellen zu können. Zu weiteren Aspekten zu Datensicherheit siehe auch Weber und Piesche (2021, 333f).

Um die logische Ebene (*Logical Preservation*) im Sinne einer technischen Interpretierbarkeit der Dateien (z. B. Software, um die Dateien zu öffnen und lesen) sicherzustellen, müssen die Dateiformate und entsprechende Software nachverfolgt werden. Bei drohender oder akuter Obsoleszenz von Dateiformaten müssen Maßnahmen getroffen werden – in der Regel Formatmigration oder Emulation. Bei der möglichst verlustfreien Formatmigration ist eine Standardisierung der zu migrierenden Formate und die Kenntnis ihres Aufbaus zentral. Darum sollte bei der langfristigen Aufbewahrung von Forschungsdaten auf die Verwendung offener, einfacher und standardisierter Formate geachtet werden. In den Sozialwissenschaften und angrenzenden Wissenschaften wird auch in der Archivierung oft mit proprietären

⁵³ Siehe hierzu auch <https://rdm-compas.org/articles/datenerhalt-erhaltungsmaßnahmen>

⁵⁴ Als redundante Sicherung kann die erweiterte 3-2-1-1-0-Regel angewendet werden: 3 verschiedene Kopien aller Dateien des Unternehmens, 2 Kopien lagern auf unterschiedlichen Speichermedien, 1 Kopie der Speichermedien befindet sich fern ab des Unternehmenssitzes, 1 Kopie ist offline oder unzerstörbar sowie 0 Fehler im Backup (durch regelmäßige Restores zu verifizieren) (<https://www.storage-insider.de/data-protection-day-2022-die-3-2-1-1-0-regel-fuer-backups-a-1091301/>).

Dateiformaten gearbeitet, beispielsweise Formate der Software Stata und SPSS. Hier muss besonders auf die Integrität der Dateien und der Erhalt der logischen Ebene geachtet werden⁵⁵, da ein offenes, nachvollziehbares Speicherformat nicht bzw. nur eingeschränkt vorhanden ist. Bestrebungen wie ein offenes (Austausch-)Format⁵⁶, die perspektivisch ggf. auch als Speicherformat weiterentwickelt werden können, sind zu begrüßen. Eine weitere Option ist die Emulation der ursprünglichen Systemumgebung, in der die Dateiformate noch geöffnet werden können. Die ist insbesondere für Dateiformate relevant, die nicht oder nur mit hohem Aufwand migriert werden können.⁵⁷ Erhaltungsmetadaten (s. Abschnitt 3.4) können beispielsweise bei der Nachverfolgung der Dateiformate unterstützen (Hein, 2012, S. 104ff).

Sind die Dateien erhalten und technisch interpretierbar, bedarf es noch ausreichend Informationen in Form von beschreibenden Metadaten, um die Daten (z. B. einzelne Werte in einer Datentabelle) zu verstehen (*Semantic Preservation*) (vgl. Kapitel 3.4).

Alle Ebenen müssen berücksichtigt werden, um Dateien langfristig zu sichern und zu erhalten. Der Prozess und die Strukturen müssen zudem regelmäßig geprüft, angepasst und weiterentwickelt werden, um den technologischen und kulturellen Wandel sowie Veränderungen der eigenen Zielgruppe (vgl. Kapitel 3.3) zu berücksichtigen. Treten bei der (regelmäßigen) Prüfung dieser Aspekte Änderungen auf, kann ein Re-Appraisal anstehen, also eine erneute Prüfung der Datenaufnahme bzw. Prüfung der relevanten Aspekte.

Die konkrete technische Umsetzung der Anforderungen zur langfristigen Erhaltung kann sehr unterschiedlich gehandhabt werden. Von einer vorstrukturierten digitalen Ordnerstruktur mit Checklisten, Prozessdefinitionen und Automatisierungen bis zu einer umfangreichen Archivsoftware. Ein Überblick über sogenannte OAIS-konforme Software findet sich bei Weber und Piesche (2021, S. 349–352). Oft kooperieren Institutionen hier auch mit anderen Organisationen oder Fachbereichen, z. B. Rechenzentren, um Synergien zu nutzen. Soll die Langzeitarchivierung bzw. Teile davon ausgelagert werden, können verschiedene Services in Betracht gezogen werden. Wichtig ist hier, die Arbeitsteilung und Verantwortungsbereiche der Akteure genau festzulegen und zu dokumentieren. In welchem Umfang Dienste ausgelagert werden, sollte jedes FDZ selbst evaluieren und festlegen. Auswirkungen einer Auslagerung für eine eventuelle Zertifizierung beispielsweise nach CoreTrustSeal (s. auch Kapitel 4) sollten dabei bedacht werden (Hoffstätter & Beck, 2023, S. 5–6).

3.6. Verfügbarmachung und Bereitstellung

Die sechste Mindestanforderung eines Archivs nach dem OAIS-Modell ist die Verfügbarmachung und Bereitstellung der Dateien. Dies wurde bereits in der Handreichung von Hoffstätter und Linne (2022) beschrieben.

⁵⁵ Die Aufbewahrung der Originaldateien (SIP) inklusive aller Migrationsschritte ist hier besonders wichtig.

⁵⁶ <https://www.konsortswd.de/angebote/open-data-format/>

⁵⁷ vgl. dazu auch Weber und Piesche (2021, S. 341–342).

4. Zusammenfassung/Ausblick

In der vorliegenden Handreichung wurden FDZ im Bereich der Sozial-, Verhaltens-, Bildungs- und Wirtschaftsdaten Hilfestellungen gegeben, die zur Erfüllung der besonders zentralen Anforderungen bei der Datenarchivierung – von der Datenaufnahme bis zur Langzeitarchivierung – im Sinne eines OAIS beitragen können.

Dabei wurden fünf Anforderungen an ein FDZ näher in den Blick genommen:

- Bzgl. der Anforderung „Verhandeln und Annehmen von Information“ wurde darauf eingegangen, welche Schritte beim *Pre-Ingest* und *Ingest* vom FDZ vorzunehmen sind bzw. was dabei zu beachten ist.
- Bzgl. der Anforderung „Ausreichende Kontrolle für die Langzeitarchivierung erlangen“ wurden Informationen zu den relevanten Inhalten einer Übergabvereinbarung zwischen Datengebenden und FDZ gegeben.
- Bzgl. der Anforderung „Vorgesehene Zielgruppe bestimmen“ wurde erläutert, was bei der Definition der Zielgruppe des FDZ zu beachten ist.
- Bzgl. der Anforderung „Sicherstellen der unmittelbaren Verstehbarkeit von Information“ wurde dargestellt, wie eine langfristige Verständlichkeit der Daten geschaffen werden kann
- Bzgl. der Anforderung „Bewährte Erhaltungspolicies und -abläufe einhalten“ wurde aufgezeigt, welche Ebenen der Erhaltung es gibt und wie die Erhaltungsmaßnahmen umgesetzt und dokumentiert werden können

Die Ausführungen verdeutlichen, dass viele Weichen für eine funktionierende Langzeitarchivierung bereits bei der Datenaufnahme gestellt werden. Insgesamt ist es notwendig, alle Anforderungen klar zu definieren, gleichzeitig aber auch die „Arbeitsprozesse und Strukturen ständig aktiv zu evaluieren und immer wieder neu zu gestalten, da alle bestehenden Lösungen immer nur über einen begrenzten Zeitraum hinweg einsetzbar sind“ (Nestor-Arbeitsgruppe Policy, 2014, S. 2).⁵⁸

Empfohlen wird, diese Thematik auch in Zusammenhang mit externen Zertifizierungen zu sehen. Während die Akkreditierung durch den RatSWD⁵⁹ den Schwerpunkt auf den Zugang zu sensiblen Forschungsdaten nach bestimmten Mindestkriterien legt, ist das CoreTrustSeal⁶⁰ (CTS) eine Zertifizierungsmöglichkeit, um die Vertrauenswürdigkeit⁶¹ digitaler Langzeitarchive

⁵⁸ Auch die Prozesse und Mechanismen einer Langzeitarchivierung können unterschiedlich umfangreich ausgestaltet sein und sich über die Zeit weiterentwickeln. Die *Levels of Digital Preservation* (<https://ndsa.org/publications/levels-of-digital-preservation/>) oder der Vorschlag zu *Curation & Preservation Levels* des CoreTrustSeal Standards And Certification Board (2023) (<https://zenodo.org/records/8083359>) können hier als Grundlage für eine Langzeitarchivierung und auch für Möglichkeiten zur Weiterentwicklung genutzt werden.

⁵⁹ <https://www.konsortswd.de/angebote/datenzentren/qualitaetssicherung-zertifizierung/akkreditierung/>

⁶⁰ <https://www.coretrustseal.org>

⁶¹ Die TRUST-Prinzipien (Lin et al., 2020b, 2020a) stehen mit *Transparency* (Transparenz), *Responsibility* (Verantwortung), *User focus* (Nutzerfokussierung), *Sustainability* (Nachhaltigkeit) und *Technology* (Technologie) als eine Zusammenstellung übergeordneter Leitlinien für vertrauenswürdige Archive. Damit haben sie eine

zu zeigen. Dabei muss die Institution die internen Prozesse über ein Self-Assessment-Verfahren anhand von 16 Kriterien darstellen. Diese werden von einem Board geprüft und für drei Jahre vergeben.⁶² Mit diesen Zertifizierungen kann gegenüber verschiedenen Stakeholdern (Datengebende, Datennutzende, Forschungsförderer), aber auch innerhalb der eigenen Organisation die Vertrauenswürdigkeit, Legitimierung und Qualität nachgewiesen und transparent gemacht sowie die Sichtbarkeit erhöht werden (Recker, Helbig & Neumann, 2020, S. 101). Zudem kann eine Zertifizierung dazu beitragen, dass die eigenen Prozesse, die organisationsinterne Kommunikation sowie die Dokumentation hinterfragt, standardisiert und optimiert werden. Idealerweise sollten die CTS-Anforderungen direkt bei der Konzeption und Planung eines Repositoriums als Checkliste hinzugezogen werden. Dokumente von bereits zertifizierten Institutionen können als wertvolle Informationsquelle dienen.

Diese Ausführungen hier sind als Startpunkt zu sehen. Die nähere Beschäftigung mit diesen Themen sollte unbedingt im Dialog mit anderen Institutionen der Community, Nachbarcommunities sowie Verbänden und Netzwerken geführt werden, um Synergien zu nutzen.⁶³ Für FDZ im Bereich der Sozial-, Verhaltens-, Bildungs- und Wirtschaftsdaten kann es hilfreich sein, zu prüfen, ob sie sich diesen bestehenden Verbänden und Netzwerken anschließen können, um so von bereits existierenden Lösungen und Strukturen zu profitieren und gleichzeitig eigene Entwicklungen in der Runde zu teilen und zu diskutieren. In diesem Zusammenhang sollte auch ein einheitliches Verständnis von Begriffen und Prozessen vorangetrieben werden (Naumann, 2023). Zudem könnte es auch prüfenswert sein, ob sich bestimmte Prozesse, die momentan noch sehr FDZ-spezifisch verlaufen, ggf. zwischen verschiedenen (ähnlichen) FDZ auch noch weiter harmonisieren und standardisieren sowie auch technisch unterstützen und teilautomatisieren lassen, wie etwa die Definition von Mindestanforderungen und die Eingangsprüfung. Auch derartige übergreifende Prozessoptimierung sollte als zentrale Aufgabe eines FDZ bei der Datenarchivierung gesehen werden. Gleichzeitig sollte aber auch die Notwendigkeit einer gewissen Pluralität an Zielen und Methoden unterschiedlicher Institutionen akzeptiert werden (Naumann, 2023).

gewisse Überschneidungsmenge mit Konzepten wie dem OAIS und dem CoreTrustSeal, welche aber den klaren Fokus auf den Aspekt der Langzeitarchivierung legen. Wie genau diese und weitere Konzepte wie die FAIR- und CARE-Prinzipien zusammenhängen wird aktuell in einer Arbeitsgruppe erarbeitet (<https://www.rd-alliance.org/group/rdawds-trust-principles-outreach-and-adoption-working-group/case-statement/rdawds-trust>).

⁶² Neben dieser Basiszertifizierung gibt es weitere Zertifizierungen wie das nestor-Siegel, welches mit einer erweiterten Selbstevaluierung für digitale Langzeitarchive auf der Grundlage der DIN-Norm 31644 arbeitet. Dieses Siegel ist dabei umfangreicher und detaillierter als das CoreTrustSeal. Darüber hinaus gibt es die Zertifizierung nach dem Standard ISO 16363:20128, die die Qualität über ein Auditverfahren mit Begehung sicherstellt.

⁶³ Fokus, Struktur und Aufgaben der Verbände können dabei unterschiedlich sein, sich aber auch überschneiden. Der Verbund Forschungsdaten Bildung (VerbundFDB) etwa ist ein thematisch orientiertes Kooperationsvorhaben von Forschungsdatenzentren aus dem Bereich der Bildungsforschung. Eng mit dem VerbundFDB zusammen arbeitet QualiBi als eine Plattform für Forschungsdaten der qualitativen Bildungsforschung. Das im Aufbau befindliche QualidataNet setzt auf einen methodischen Schwerpunkt als Netzwerk von Forschungsdatenzentren, Archiven und Repositorien, die Forschungsmaterialien der qualitativen Sozialforschung archivieren und für die Nachnutzung bereitstellen.

Literaturverzeichnis

- Altenhöner, R. & Oellers, C. (Hrsg.). (2012). *Langzeitarchivierung von Forschungsdaten. Standards und disziplinspezifische Lösungen*. Berlin: Scivero.
- Bischof, C. (2022). *Preservation Plan (1.0)*. Vienna: The Austrian Social Science Data Archive. Verfügbar unter https://aussda.at/fileadmin/user_upload/p_aussda/Documents/Preservation_Plan_v1_0.pdf
- Brislinger, E. & Moschner, M. (2019). Datenaufbereitung und Dokumentation. In U. Jensen, S. Netscher & K. Weller (Hrsg.), *Forschungsdatenmanagement sozialwissenschaftlicher Umfragedaten. Grundlagen und praktische Lösungen für den Umgang mit quantitativen Forschungsdaten* (S. 97–114). Berlin: Verlag Barbara Budrich. Zugriff am 05.02.2019.
- Butzlaff, I. (2022). *Data Deposit Guideline (Public version) (v2.0)*. Vienna: The Austrian Social Science Data (AUSSDA). Verfügbar unter https://aussda.at/fileadmin/user_upload/p_aussda/Documents/Data-Deposit-Guideline_SUF_v2_0.pdf
- The Consultative Committee For Space Data Systems. (2012). *Recommendation for Space Data System Practices. Reference Model For An Open Archival Information System (OAIS)*. Verfügbar unter <https://public.ccsds.org/Pubs/651x0m1.pdf>
- CoreTrustSeal Standards And Certification Board. 2023, 1. January. *Curation & Preservation Levels: CoreTrustSeal Discussion Paper*. doi:10.5281/ZENODO.8083359
- Dappert, A. & Enders, M. (2010). Digital Preservation Metadata Standards. *Information Standards Quarterly*, 22(2), 4. Verfügbar unter https://www.loc.gov/standards/premis/FE_Dappert_Enders_MetadataStds_isqv22no2.pdf
- Ebel, T. & Meyermann, A. (2015). *Hinweise zur Anonymisierung von quantitativen Daten* (Forschungsdatenzentrum (FDZ) Bildung am DIPF, Hrsg.) (forschungsdaten bildung informiert Nr. 3). Frankfurt am Main: Verbund Forschungsdaten Bildung. Verfügbar unter https://www.forschungsdaten-bildung.de/get_files.php?action=get_file&file=fdb-informiert_nr-7.pdf
- Gartner, R. (2008). Metadata for digital libraries: state of the art and future directions. *JISC Technology and Standards Watch*. Verfügbar unter https://www.webarchive.org.uk/wayback/archive/20140613220103/http://www.jisc.ac.uk/media/documents/techwatch/tsw_0801pdf.pdf
- GESIS – Leibniz-Institut für Sozialwissenschaften. (2024 [in Veröffentlichung]). *GESIS Data Services. CoreTrustSeal Requirements 2023–2025*.
- Hein, S. (2012). Metadaten für die Langzeitarchivierung. In R. Altenhöner & C. Oellers (Hrsg.), *Langzeitarchivierung von Forschungsdaten. Standards und disziplinspezifische Lösungen* (S. 87–110). Berlin: Scivero.
- Hoffstätter, U. & Beck, K. (2023). *Workshopdokumentation: CoreTrustSeal: Erstveranstaltung*. doi:10.5281/ZENODO.10213806

- Hoffstätter, U. & Linne, M. (2022). *Datenzugang. Einführung in das Thema Zugang zu Daten der Sozial-, Verhaltens-, Bildungs- und Wirtschaftswissenschaften in Forschungsdatenzentren*. doi:10.5281/ZENODO.7347063
- Forschungsdatenzentrum am Institut zur Qualitätsentwicklung im Bildungswesen. (2021a). *Aufbereitung der Datensätze am FDZ am IQB / Steps of Data Preparation* (IQB – Institut zur Qualitätsentwicklung im Bildungswesen, Hrsg.). Berlin. Verfügbar unter https://www.iqb.hu-berlin.de/fdz/Datenuebergabe/Vorlage_AbkStudi.pdf
- Forschungsdatenzentrum am Institut zur Qualitätsentwicklung im Bildungswesen. (2021b). *Konzept zur Langzeitverfügbarkeit digitaler Datensätze des FDZ am IQB*. (IQB – Institut zur Qualitätsentwicklung im Bildungswesen, Hrsg.). Berlin.
- Jensen, U., Zenk-Möltgen, W. & Wasner, C. (2019). Metadatenstandards im Kontext sozialwissenschaftlicher Daten. In U. Jensen, S. Netscher & K. Weller (Hrsg.), *Forschungsdatenmanagement sozialwissenschaftlicher Umfragedaten. Grundlagen und praktische Lösungen für den Umgang mit quantitativen Forschungsdaten* (S. 151–178). Berlin: Verlag Barbara Budrich. Zugriff am 05.02.2019.
- Keitel, C. & Mitcham, J. (2023). *Defining the Designated Community*. doi:10.7207/twgn23-01
- Lauber-Rönsberg, A. (2021). Rechtliche Aspekte des Forschungsdatenmanagements. In M. Putnings, H. Neuroth & J. Neumann (Hrsg.), *Praxishandbuch Forschungsdatenmanagement* (S. 89–114). Berlin: De Gruyter Saur. doi:10.1515/9783110657807-005
- Lauber-Rönsberg, A., Krahn, P. & Baumann, P. (2018). *Gutachten zu den rechtlichen Rahmenbedingungen des Forschungsdatenmanagements*. Verfügbar unter https://tu-dresden.de/gsw/phil/irget/jfbimd13/ressourcen/dateien/dateien/DataJus/DataJus_Kurzfassung_Gutachten_12-07-18.pdf?lang=de&set_language=de
- Lavoie, B. (2014). *The Open Archival Information System (OAIS) Reference Model: Introductory Guide (2nd Edition)*. doi:10.7207/twr14-02
- Liegmann, H. & Neuroth, H. (2010). Einführung. In H. Neuroth, A. Oßwald, R. Scheffel, S. Strathmann & K. Huth (Hrsg.), *nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*. Göttingen: Hülsbusch / Univ.-Verl. Göttingen.
- Limani, F., Younes, Y., Bach, J. S., Hiseni, V., Mutschke, P. & Mathiak, B. 2022, 1. January. *KonsortSWD Measure 5.2: Enhancing data findability Milestones 1, 2, and 3 report*. doi:10.5281/ZENODO.7224672
- Lin, D., Crabtree, J., Dillo, I., Downs, R. R., Edmunds, R., Giaretta, D. et al. (2020a). The TRUST Principles for digital repositories. *Scientific data*, 7(1), 144. doi:10.1038/s41597-020-0486-7
- Lin, D., Crabtree, J., Dillo, I., Downs, R. R., Edmunds, R., Giaretta, D. et al. (2020b). Die TRUST-Prinzipien für digitale Repositorien. doi:10.5281/ZENODO.6256222
- Mauer, R. (2012). Das GESIS Datenarchiv für Sozialwissenschaften. In R. Altenhöner & C. Oellers (Hrsg.), *Langzeitarchivierung von Forschungsdaten. Standards und disziplinspezifische Lösungen* (S. 197–215). Berlin: Scivero.

- Mauer, R. (2016). *Auswahl und Bewertung von Forschungsdaten für die Archivierung und Nachnutzung*. GESIS – Leibniz-Institut für Sozialwissenschaften. Verfügbar unter http://www.nestor.sub.uni-goettingen.de/school_2016/slides/PERICLES_WP7_T7-3_UGOE_nestor_PERICLES_School_Presentation_03.pdf
- Meyermann, A. & Porzelt, M. (2014). *Hinweise zur Anonymisierung von qualitativen Daten* (forschungsdaten bildung informiert Nr. 1). Frankfurt am Main: Forschungsdatenzentrum (FDZ) Bildung am DIPF; Deutsches Institut für Internationale Pädagogische Forschung.
- Meyermann, A. & Porzelt, M. (2019). *Datenschutzrechtliche Anforderungen in der empirischen Bildungsforschung. Eine Handreichung*: DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation : Frankfurt am Main. doi:10.25656/01:21990
- Naumann, K. (2023). *Langzeitarchivierung im Landesarchiv Baden-Württemberg*. doi:10.5281/ZENODO.10276163
- Nestor-Arbeitsgruppe Policy. (2014). *Leitfaden zur Erstellung einer institutionellen Policy zur digitalen Langzeitarchivierung* (nestor-Arbeitsgruppe Policy, Hrsg.) (nestor-materialien Nr. 18). Verfügbar unter <https://d-nb.info/1051731216/34>
- Netscher, S. & Jensen, U. (2019a). Forschungsdatenmanagement systematisch planen und umsetzen. In U. Jensen, S. Netscher & K. Weller (Hrsg.), *Forschungsdatenmanagement sozialwissenschaftlicher Umfragedaten. Grundlagen und praktische Lösungen für den Umgang mit quantitativen Forschungsdaten* (S. 37–55). Berlin: Verlag Barbara Budrich. doi:10.3224/84742233.04
- Netscher, S. & Jensen, U. (2019b). Forschungsdatenmanagement systematisch planen und umsetzen. In U. Jensen, S. Netscher & K. Weller (Hrsg.), *Forschungsdatenmanagement sozialwissenschaftlicher Umfragedaten. Grundlagen und praktische Lösungen für den Umgang mit quantitativen Forschungsdaten* (S. 37–55). Berlin: Verlag Barbara Budrich. Zugriff am 05.02.2019.
- OAIS-Übersetzung/Terminologie, nestor Arbeitsgruppe. (2012). *Referenzmodell für ein Offenes Archiv-Informationssystem. Deutsche Übersetzung*: nestor – Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit Digitaler Ressourcen für Deutschland.
- OAIS-Übersetzung/Terminologie, nestor Arbeitsgruppe. (2013). *Referenzmodell für ein Offenes Archiv-Informationssystem. Deutsche Übersetzung 2.0*: nestor – Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit Digitaler Ressourcen für Deutschland. doi:10.18452/1532
- Rat für Sozial- und Wirtschaftsdaten. (2022). *Tätigkeitsbericht 2020 der vom RatSWD akkreditierten Forschungsdatenzentren (FDZ)*. doi:10.17620/02671.65
- RatSWD. (2020). Handreichung Datenschutz. 2. vollständig überarbeitete Auflage. *RatSWD Output*, 8(6). doi:10.17620/02671.50
- Recker, J. & Brislinger, E. (2019). Dateiorganisation in empirischen Forschungsprojekten. In U. Jensen, S. Netscher & K. Weller (Hrsg.), *Forschungsdatenmanagement*

- sozialwissenschaftlicher Umfragedaten. Grundlagen und praktische Lösungen für den Umgang mit quantitativen Forschungsdaten* (S. 81–95). Berlin: Verlag Barbara Budrich. Zugriff am 05.02.2019.
- Recker, J., Helbig, K. & Neumann, J. (2020). Zertifizierung von Forschungsdatenrepositorien: Wege, Praxiserfahrungen und Perspektiven. doi:10.17192/BFDM.2020.2.8280
- Schallaböck, J., Hoffstätter, U., Buck, D. & Linne, M. (2023). *Mustervertrag Datennutzung KonsortSWD*. doi:10.5281/ZENODO.8186162
- Schallaböck, J., Kreutzer, T., Hoffstätter, U. & Buck, D. (2023a). *Mustervertrag Datenaufnahme KonsortSWD*. doi:10.5281/zenodo.7648897
- Schallaböck, J., Kreutzer, T., Hoffstätter, U. & Buck, D. (2023b). *Mustervertrag Datenaufnahme KonsortSWD*. doi:10.5281/ZENODO.7648898
- Schrimpf, S. (2012). Überblick über das OAIS-Referenzmodell. In R. Altenhöner & C. Oellers (Hrsg.), *Langzeitarchivierung von Forschungsdaten. Standards und disziplinspezifische Lösungen* (S. 51–68). Berlin: Scivero.
- Schrimpf, S. & Steinke, T. (2018). *Langzeitarchivierungs-Policy der Deutschen Nationalbibliothek* (Version 1.2 Stand: 4.5.2018). Leipzig: Deutsche Nationalbibliothek.
- Schumann, N. (2012). Einführung in die digitale Langzeitarchivierung. In R. Altenhöner & C. Oellers (Hrsg.), *Langzeitarchivierung von Forschungsdaten. Standards und disziplinspezifische Lösungen* (S. 39–48). Berlin: Scivero.
- Schwens, U. & Liegmann, H. (2011). D 9 Langzeitarchivierung digitaler Ressourcen. In R. Kuhlen, T. Seeger & D. Strauch (Hrsg.), *Grundlagen der praktischen Information und Dokumentation* (5., völlig neu gefasste Ausgabe, S. 567–570). München: Saur; De Gruyter. doi:10.1515/9783110964110.567
- Thibodeau, K. (2002). Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years. In Council on Library and Information Resources (Ed.), *The state of digital preservation. An international perspective; conference proceedings, Documentation Abstracts, Inc. Institutes for Information Science, Washington, D.C., April 24-25, 2002* (pp. 4–31). Washington, D.C.: CLIR. Verfügbar unter <https://www.clir.org/wp-content/uploads/sites/6/pub107.pdf>
- Uwe Jensen, G. (2012). *Metadaten für Forschungsdaten: Welche Standards gibt es?* 101. Deutscher Bibliothekartag, Hamburg. Zugriff am 18.04.2016.
- Verbund Forschungsdaten Bildung. (2018). *Formulierungsbeispiele für „informierte Einwilligungen“* (Version 2.1): DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation. doi:10.25656/01:22301
- VerbundFDB – AG Cessation. (2024 [in Veröffentlichung]). *Template zur Erfassung von relevanten Informationen für eine Nachfolgeregelung für den Fall der Auflösung eines Datenzentrums* (Version 1.0).

- Wallace, D. A. (2001). Archiving metadata forum: Report from the Recordkeeping Metadata Working Meeting, June 2000. *Archival Science*, 1(3), 253–269. doi:10.1007/BF02437690
- Watteler, O. (2010). *Erstellung von Methodenberichten für die Archivierung von Forschungsdaten* (GESIS – Leibniz-Institut für Sozialwissenschaften, Hrsg.). Köln: GESIS – Leibniz-Institut für Sozialwissenschaften.
- Watteler, O. & Ebel, T. (2019). Datenschutz im Forschungsdatenmanagement. In U. Jensen, S. Netscher & K. Weller (Hrsg.), *Forschungsdatenmanagement sozialwissenschaftlicher Umfragedaten. Grundlagen und praktische Lösungen für den Umgang mit quantitativen Forschungsdaten* (S. 57–80). Berlin: Verlag Barbara Budrich. Zugriff am 05.02.2019.
- Weber, A. & Piesche, C. (2021). 4.2 Datenspeicherung, -kuration und Langzeitverfügbarkeit. In M. Putnings, H. Neuroth & J. Neumann (Hrsg.), *Praxishandbuch Forschungsdatenmanagement* (S. 327–356). Berlin: De Gruyter Saur. doi:10.1515/9783110657807-019
- Whyte, A. & Wilson, A. (2010). *How to Appraise & Select Research Data for Curation*. Digital Curation Center (DCC); Australian National Data Service (ANDS).
- Wünsche, S., Soßna, V., Kreitlow, V. & Voigt, P. (2022). Urheberrechte an Forschungsdaten – Typische Unsicherheiten und wie man sie vermindern könnte. doi:10.17192/bfdm.2022.1.8369

Impressum

Kontakt:

Ute Hoffstätter

Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW)

Lange Laube 12

30159 Hannover

Tel.: +49 511 450670-404

hoffstaetter@dzhw.eu

April 2024

KonsortSWD Working Paper:

KonsortSWD baut als Teil der Nationalen Forschungsdateninfrastruktur Angebote zur Unterstützung von Forschung mit Daten in den Sozial-, Verhaltens-, Bildungs- und Wirtschaftswissenschaften aus. Unsere Mission ist es, die Forschungsdateninfrastruktur zur Beforschung der Gesellschaft zu stärken, zu erweitern und zu vertiefen. Sie soll nutzungsorientiert ausgestaltet sein und die Bedürfnisse der Forschungscommunities berücksichtigen. Wichtiger Grundstein ist dabei das seit über zwei Jahrzehnten durch den Rat für Sozial- und Wirtschaftsdaten (RatSWD) aufgebaute Netzwerk von Forschungsdatenzentren. In dieser Reihe erscheinen Beiträge rund um das Forschungsdatenmanagement, die im Kontext von KonsortSWD entstehen. Beiträge, die extern und doppelblind begutachtet wurden, sind entsprechend gekennzeichnet.

KonsortSWD wird im Rahmen der NFDI durch die Deutsche Forschungsgemeinschaft (DFG) gefördert – Projektnummer: 442494171.



Diese Veröffentlichung ist unter der Creative-Commons-Lizenz (CC BY 4.0) lizenziert:

<https://creativecommons.org/licenses/by/4.0/>

DOI: 10.5281/zenodo.10418834

Zitationsvorschlag:

Hoffstätter, U., & Weber, A. (2024). *Langzeitarchivierung von Forschungsdaten. Einführung in das Thema zu Daten der Sozial-, Verhaltens-, Bildungs- und Wirtschaftswissenschaften in Forschungsdatenzentren*. KonsortSWD Working Paper Nr.9/2024. Konsortium für die Sozial-, Verhaltens-, Bildungs- und Wirtschaftswissenschaften (KonsortSWD).

<https://doi.org/10.5281/zenodo.10418834>