

# Langzeitarchivierung in der NFDI

## Tendenzen und Bedarfe als Resultat einer Umfrage

Katharina Markus (ORCID:0000-0002-9316-8982), Kai Naumann (ORCID:0000-0002-2799-1030), Markus Schmalzl (ORCID:0009-0005-7574-2699), Juliane Watson, Dagmar Triebel (ORCID:0000-0003-1980-3148)

### Einleitung

Versäumnisse bei der Langzeitarchivierung könnten zur Achillesferse der Nationalen Forschungsdateninfrastruktur (NFDI) werden, denn digitale Forschungsdaten sind einer Vielzahl von Gefahren ausgesetzt. Hierzu zählen fehlende Nutzungsrechte<sup>1</sup>, fehlende organisatorische und finanzielle Sicherung von institutionellen Services<sup>2</sup>, fehlende Klärung von Verantwortung und Prozessen<sup>3</sup>, menschliche Fehler, Versagen von Hardware und Datenträgern<sup>4</sup>, Hardware-Obsoleszenz<sup>5</sup>, Software-Obsoleszenz<sup>6</sup>, nicht mehr ausführbare Forschungssoftware und weitere. Die Wiederherstellung der Daten ist in diesen Fällen nicht möglich oder äußerst ressourcen- und kostenintensiv und in einigen Fällen ist die Neuerhebung der verlorenen Daten auch nicht möglich<sup>7</sup>. Diese Risiken können schon innerhalb der von der Deutschen Forschungsgesellschaft (DFG) empfohlenen zehn Jahre

---

<sup>1</sup> “In the intervening 15 years, new issues and approaches to rights management in the digital environment – such as CC licenses – have arisen. These issues, along with content licenses, copyright and intellectual property laws, will inevitably continue to evolve; highly specific and legally binding terms may prove out of sync with long-term preservation goals.” (Sprout and Jordan 2018)

<sup>2</sup> “We also found databases whose search function had either disappeared or was nonfunctional or had lost other key functionality. We counted these as dead even if they still existed on the Web. [...] according to our criteria, a total of 203 (62%) were dead [...]. (Attwood, Agit, and Ellis 2015)

<sup>3</sup> “[...] leaving a very large portion (~3 TB) of legacy data in need of organization and integration. Routine data management tasks such as finding records, identifying data types and production dates,

sorting through multiple copies, culling corrupted and redundant files, and reorganizing data have been conducted manually, placing a significant burden on research staff. (Arora, Esteva, and Trelogan 2014)

<sup>4</sup> “Many of the tapes had suffered ‘stiction’ problems and were literally falling apart.[...] Several data sets were unrecoverable – some waterfowl data and land use data in the Eastern Townships.” (Bleakly 2002)

<sup>5</sup> (Digital Preservation Coalition (DPC) 2023b; Bleakly 2002)

<sup>6</sup> “Much of the software in which this data was created is now obsolete, and some of the data formats are challenging or impossible to read in modern software.” (Pinnick et al. 2018). “In the early 2000s, the data recovery company was unable to make sense of the recovered bit files -they were in a proprietary format and could not be accessed.” (Doig 2020)

<sup>7</sup> “Some geoscience data acquisition projects are too expensive to repeat, such as drilling deep boreholes to the depth of many kilometres costing tens of millions of pounds [...] Seismic data originating from earthquakes are another example of unique and unrepeatabe data [...]” (Pinnick 2017)

Aufbewahrung<sup>8</sup> auftreten und damit die Reproduzierbarkeit, weitere Verwendbarkeit (Reusability) und Nachvollziehbarkeit sowie die gute wissenschaftliche Praxis gefährden. Darüber hinaus ist nur mittels einer durchdachten Langzeitarchivierung gewährleistet, dass die Forschungsdaten, sofern erforderlich, über Jahrzehnte hinweg FAIR (Findable, Accessible, Interoperable, Reusable)<sup>9</sup> bleiben<sup>10</sup>.

Entsprechend hat auch der Rat für Informationsinfrastrukturen (RfII) erklärt, dass Langzeitarchivierung zu den „wichtigen Aufgabengebieten einer Nationalen Forschungsdateninfrastruktur“ gehört<sup>11</sup>. Eine initiale NFDI-interne Bedarfsermittlung in den Konsortien 2021 zeigte, dass bei diesen nur wenig Kenntnis über Langzeitarchivierung (Long-Term Archiving) herrschte<sup>12</sup>, ein Bedarf an einer Lösung schien aber zu bestehen. Langzeitarchivierung wurde daher auch ein Thema für einen potentiellen NFDI-Basisdienst und Teil des Base4NFDI-Antrags<sup>13</sup>. Gleichzeitig bildete sich eine Arbeitsgruppe Long-Term Archival (AG LTA)<sup>14</sup> in der Sektion Common Infrastructure<sup>15</sup>, die das Thema Langzeitarchivierung konsortien-übergreifend bearbeitet.

Um einen sinnvollen LZA-Basisdienst zu konzipieren, wurde der Bedarf in den NFDI-Konsortien in einer detaillierten Umfrage durch die AG LTA ermittelt. Die Ergebnisse der Umfrage werden im Folgenden diskutiert.

## Methodik und Grenzen der Auswertung

Die Umfrage zum Stand der Langzeitarchivierung in den Konsortien wurde erstellt, um bei (Fach-) Konsortien Bedarfe, Lücken und Kenntnisse im Kontext Langzeitarchivierung (LZA) zu identifizieren. Da die Konsortien selbst sehr divers aufgestellt sind, wurde die Umfrage von den Mitgliedern der AG LTA offen konzipiert, indem sie bei jeder Antwort Freitext erlaubte. Den Fragen steht ein einleitender Absatz voran, der Langzeitarchivierung entsprechend einschlägiger Standards (ISO 14721<sup>16</sup>) einführt und definiert. Darauf folgt ein Abschnitt zu den existierenden oder geplanten Angeboten des Konsortiums und ein Abschnitt zu den Bedarfen hinsichtlich eines neuen LZA-Dienstes oder am Ausbau eines vorhandenen LZA-Dienstes (siehe Anhang).

Mitglieder der AG LTA haben daraufhin Ansprechpartner:innen in allen Konsortien (26 zum Zeitpunkt der Umfrage, Liste entsprechend der NFDI<sup>17</sup>) ausfindig gemacht und kontaktiert und um die Ausfüllung der Umfrage gebeten. Die Mitglieder der AG konnten aufgrund ihrer Zugehörigkeit zu einem oder mehreren Konsortien zum Teil die Umfrage auch selber ausfüllen bzw. in den Konsortien ansprechen. In dem Zuge wurde keine

---

<sup>8</sup> (Deutsche Forschungsgemeinschaft. 2022b; 2022a)

<sup>9</sup> (Wilkinson et al. 2016)

<sup>10</sup> Auch die EOSC Long-Term Data Preservation Task Force verbindet Langzeitarchivierung und FAIR (Andreu et al. 2023).

<sup>11</sup> (Rat für Informationsinfrastrukturen 2016)

<sup>12</sup> Unveröffentlichte Ergebnisse. Genereller Bedarf wurde auch im Base4NFDI-Antrag erwähnt (Bernard et al. 2023)

<sup>13</sup> (Bernard et al. 2023)

<sup>14</sup> (Bach et al. 2022)

<sup>15</sup> (Diepenbroek, Schimmler, and Ebert 2021)

<sup>16</sup> ('ISO 14721:2012. Open Archival Information System (OAIS) Reference Model' 2012)

<sup>17</sup> (NFDI 2023)

Erlaubnis zur Veröffentlichung der Detailergebnisse eingeholt, damit die Antworten der einzelnen Konsortien praxisorientiert bleiben und möglichst wenig durch strategische Überlegungen zur Selbstdarstellung beeinflusst werden. Entsprechend enthält diese Publikation nur öffentlich zugängliche Informationen zu spezifischen Konsortien.

Die Antworten wurden von einem Team innerhalb der AG LTA ausgewertet. Durch die Freitext-Möglichkeiten ist eine quantitative Auswertung nur eingeschränkt möglich, sodass primär eine qualitative Auswertung durchgeführt wurde. Unklare Antworten (beispielsweise wurde eine Frage ähnlich "Gibt es im Konsortium [...]" nicht mit Ja oder Nein beantwortet, sondern z. B. mit der Beschreibung eines Forschungsdaten-Services) wurden für die Auswertung durch das Auswertungs-Team interpretiert (als Ja, Nein, oder ?). Wurde dagegen klar mit Ja oder Nein geantwortet und diese Antwort dann aber relativiert, wurde trotzdem das klare Ja oder Nein für die Auswertung genutzt. Das heißt aber auch, dass Aussagen wie "Ja, x gibt es bei dem einen Datenanbieter" eines Konsortiums und "Nein, x gibt es nur bei einem Datenanbieter" eines anderen Konsortiums trotz ähnlichen Inhalts in der Auswertung entsprechend gegensätzlich behandelt wurden. Diese Unschärfe wurde in Kauf genommen, um die Aussagen der Konsortien nicht zu verfälschen. Die beobachteten Tendenzen, die in dieser Publikation behandelt werden, blieben von dieser Unschärfe unbeeinflusst.

## Ergebnisse und beobachtete Tendenzen

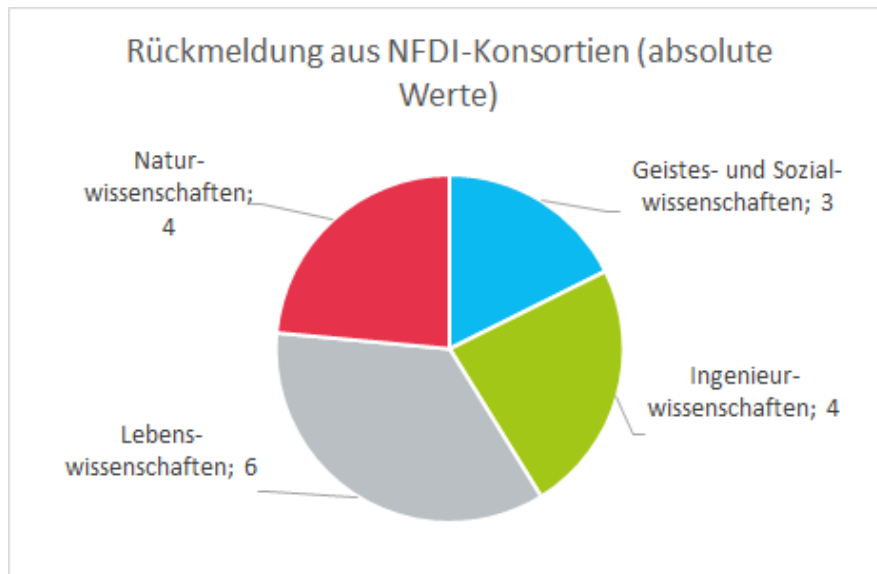
### Umfang der Rückmeldung

Obwohl alle Konsortien die Umfrage erhalten haben, konnte nicht von allen Konsortien eine Antwort eingeholt werden. Zum Zeitpunkt der Veröffentlichung haben 17 von 26 Konsortien (65%) Antworten übermittelt.

Diese teilen sich auf alle vier großen Wissenschaftsdisziplinen der NFDI auf: Geistes- und Sozialwissenschaften, Ingenieurwissenschaften, Lebenswissenschaften und Naturwissenschaften<sup>18</sup> (Abb. 1).

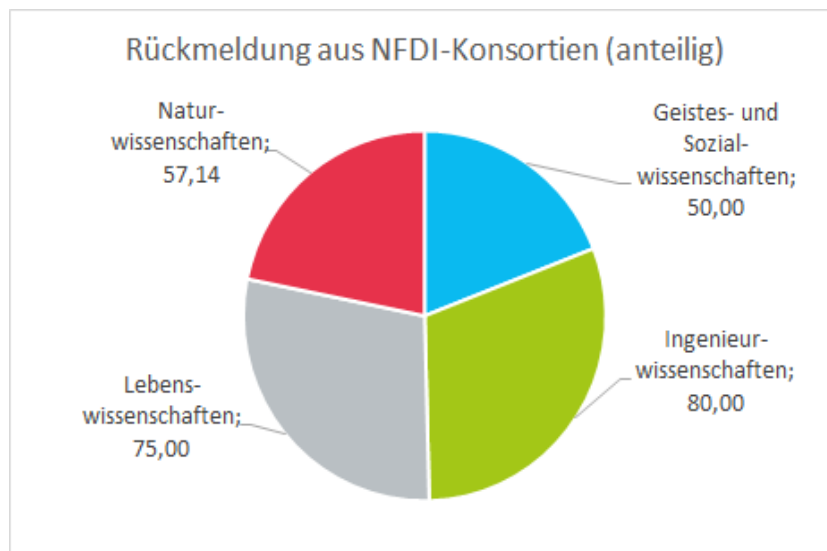
---

<sup>18</sup> (NFDI 2023)



**Abb. 1** Rückmeldung von Konsortien aufgeteilt nach Wissenschaftsdisziplinen  
Die Zuordnung zu Disziplinen wurde entsprechend der NFDI veröffentlichten Einteilung<sup>19</sup> vorgenommen.

Die meisten Rückmeldungen konnten aus den Lebenswissenschaften eingeholt werden (Antworten aus sechs Konsortien), die wenigsten aus den Geistes- und Sozialwissenschaften (Antworten aus drei Konsortien). Damit spiegelt die Umfrage vorwiegend Antworten aus den Lebens-, Natur- und Ingenieurwissenschaften wider. Da aber auch Geistes- und Sozialwissenschaften eher mit alten Daten (z. B. Nachlässen) arbeiten, ist in diesem Bereich dennoch ein überdurchschnittliches Bewusstsein für Langzeitarchivierung zu vermuten.



**Abb. 2** Rückmeldung von Konsortien anteilig an allen Konsortien einer Wissenschaftsdisziplin  
Die Angaben sind prozentual errechnet unter der Annahme, dass alle Konsortien einer Wissenschaftsdisziplin 100% ausmachen. Die Zuordnung zu Disziplinen wurde entsprechend der NFDI veröffentlichten Einteilung<sup>20</sup> vorgenommen.

<sup>19</sup> (NFDI 2023)

<sup>20</sup> (NFDI 2023)

Die absoluten Zahlen (Abb. 1) geben keinen Aufschluss über das Antwortverhalten der verschiedenen Wissenschaftsbereiche (Disziplinen). Entsprechend wurde der Rücklauf mit der Anzahl aller Konsortien in einer Disziplin verrechnet (Abb. 2). So kann man erkennen, dass viele Konsortien aus den Ingenieurwissenschaften geantwortet haben (80%), während nur 50% der Konsortien der Geistes- und Sozialwissenschaften die Umfrage ausgefüllt haben.

## Verständnis von Langzeitarchivierung, LZA-Terminologie und Standards

Langzeitarchivierung in dieser Veröffentlichung orientiert sich an der Definition des Open Archival Information System-Standards (OAIS): “Long Term Preservation: The act of maintaining information, Independently understandable by a Designated Community, and with evidence supporting its Authenticity, over the Long Term.”<sup>21</sup>. Im Folgenden werden die Begriffe Digitale Langzeitarchivierung, Langzeitarchivierung, Langzeitverfügbarkeit, Digital Preservation und Long Term Preservation synonym verwendet. Langzeitarchivierung setzt keine zeitliche Einschränkung voraus und kann z. B. zehn Jahre deutlich übersteigen. Damit unterscheidet sie sich von den DFG-Leitlinien zur Sicherung guter wissenschaftlicher Praxis, die unter “Archivierung” die Aufbewahrung der Daten in Einrichtungen “für einen Zeitraum von zehn Jahren” verstehen - obwohl entsprechend DFG auch kürzere oder längere Zeiträume angemessen sein können<sup>22</sup>.

Um den Umfang der Langzeitarchivierung und die Verantwortung für Erhaltungsmaßnahmen in Institutionen und Konsortien zu klären, wurden in der Umfrage und deren Auswertung die Object Levels of Preservation<sup>23</sup> verwendet. Diese unterteilen sich in

- Bitstream Preservation
- Logical Preservation (Logischer Erhalt) und
- Semantic Preservation (Semantischer Erhalt)

und bauen aufeinander auf.

Bitstream Preservation bedeutet den Erhalt von digitalen Objekte auf der Bitstream-Ebene. Bei Logical Preservation wird der Informationsverlust z. B. aufgrund obsoleter Formate, obsoleter Software oder Technologien verhindert - durch Erhaltungsmaßnahmen wie Migration auf aktuelle Dateiformate und Emulation. Semantic Preservation hat das Ziel, das Objekt auf konzeptioneller Ebene zu erhalten, z. B. durch Erhalt des Kontexts, der Dokumenten-Authentizität, der Struktur des Objekts, Beziehungen zwischen Objekten und ausreichender beschreibender Metadaten.

In der Umfrage der AG LTA wurden die zwei Level Logical Preservation und Semantic Preservation zu “funktionale[r] Erhaltung” (Anhang, Umfrage (Deutsch)) zusammengefasst, sodass alle Konsortien, die angeben, funktionalen Erhalt anzubieten, sowohl Logical Preservation als auch Semantic Preservation abdecken können (siehe auch Abb. 3). Die Zusammenfassung dieser zwei Level nehmen auch die Autor:innen der NFDI4Culture

---

<sup>21</sup> (‘ISO 14721:2012. Open Archival Information System (OAIS) Reference Model’ 2012; The Consultative Committee for Space Data Systems (CCSDS) 2012)

<sup>22</sup> (Deutsche Forschungsgemeinschaft. 2022b; 2022a)

<sup>23</sup> (Lindlar et al. 2020)

Handreichung vor, in der sie "Content Preservation"<sup>24</sup> benutzen. Im Folgenden wird diese Definition von Content Preservation übernommen und synonym für funktionalen Erhalt benutzt (Abb. 3).

Die Konsortien beschäftigen sich derzeit kaum vertieft mit dem Thema der digitalen Langzeitarchivierung. Bei den 17 Antworten gaben elf Konsortien an, Bitstream Preservation zu betreiben und nur sechs Content Preservation. Vereinzelt wurde Bitstream Preservation bei der Angabe von Content Preservation nicht extra aufgezählt - da Content Preservation auf Bitstream Preservation aufbaut, wurde davon ausgegangen, dass in diesen Fällen beides gegeben ist.

Kein Konsortium hat explizit zurückgemeldet, keine Bitstream Preservation zu betreiben, aber bei sechs Rückmeldungen wurde diese Frage nicht beantwortet. Diese fehlende Antwort korrelierte immer mit einem fehlenden LZA-Dienst. Diese Konsortien verweisen dann zum Teil auf existierende Forschungsdaten-Services (Datenspeicherlösungen, Fachdatenbanken, fachliche Repositorien, generische Repositorien, Publikations- und Speicherangebote an Hochschulen, Forschungsdatenzentren, Archive, Software Directories, Workflow-Plattformen u. ä.), sehen sich nicht verantwortlich für Langzeitarchivierung oder scheinen keine Berührungspunkte mit dem Thema Langzeitarchivierung zu haben.

Die sechs Konsortien, die sich dagegen zur Content Preservation bekennen, geben oft einen bestimmten Forschungsdaten-Service an, der Daten über das Bitstream Level hinaus erhält. Das heißt nicht notwendigerweise, dass alle Daten des Fachbereiches in diesem Sinne langzeitarchiviert werden. Die Herausforderungen bei der Einbindung verschiedener Services und entsprechend verteilter LZA-Lösungen werden im nächsten Kapitel aufgegriffen.

Es ist zu vermuten, dass für die meisten Forschungsdaten-Services Datenverlust über das Bitstream Level hinaus relevant wird, selbst innerhalb der von der DFG empfohlenen zehn Jahre (vgl. entsprechende Meldungen der DPC 2023<sup>25</sup>). Forschungssoftware kann innerhalb weniger Jahre vom Markt genommen werden oder inkompatibel mit Formatversionen von vor 10 Jahren sein. Entsprechend besteht Bedarf, existierende Forschungsdaten-Services mit Langzeitarchivierungs-Lösungen zu verknüpfen oder auszubauen (inklusive personeller Ressourcen), um Verluste wichtiger Daten und Informationen zu vermeiden. Gleichzeitig müssen Forschungsförderer solche Entwicklungen finanziell unterstützen. Damit geht aus der Umfrage vor allem hervor, dass es der NFDI an einem gemeinsamen Verständnis von Langzeitarchivierung, an einer gemeinsamen Begriffsbasis, sowie an einer Umsetzung von Langzeitarchivierung fehlt.

---

<sup>24</sup> (Heseler, Büttner, and Arnold 2024)

<sup>25</sup> "Digits are born vulnerable. Every single byte of data depends on a global infrastructure of technology, process and people for its meaning and purpose to be realized. Much data serves the moment: it is quickly forgotten in a continuous flow of process and interaction."(Digital Preservation Coalition (DPC) 2023b)

Lindlar et al. 2020 / Thibodeau 2002 <sup>26</sup>	Projekt WissGrid 2011 <sup>27</sup>	Rfll 2016 <sup>28</sup>	EOSC TF Long Term Data Preservation 2023 <sup>29</sup>	NFDI4Culture 2023 <sup>30</sup>	AG LTA Umfrage 2023
“Bitstream Preservation”	“Bitstream Preservation”	-	-	“Bitstream Preservation”	“Bitstream Preservation”
“Logical Preservation”	“Content Preservation”	“Content Preservation”	“Logical-Technical Preservation”		
“Semantic Preservation”	“Data Curation”	“Datenkuration, Datenkuratierung”	“Conceptual preservation for understanding and reuse”	“Content Preservation”	“funktionale[r] Erhalt”

**Abb. 3** Bezeichnungen für Erhaltungslevel oder Kuratierungslevel in der Literatur, zugeordnet zu den Object Levels of Preservation. Die zitierten Bezeichnungen stammen aus den Quellen, die in der ersten Zeile referenziert werden. Die Zuordnung wurde von den Autor:innen dieser Publikation anhand der Definition in der jeweiligen Literatur-Quelle vorgenommen. Sie wurde nur von einzelnen dieser Quellen explizit durchgeführt. In einigen Fällen lässt sich Semantic Preservation und Content Preservation nicht klar voneinander abgrenzen, sodass diese Tabelle nur eine Annäherung ist.

<sup>26</sup> “These classes are the Physical, Logical, and the Conceptual Object, which themselves have unique properties [...] and bear distinct risks, e.g. lack of robust storage on the physical level, unsuitable file formats or dependencies on externally linked resources on the logical level, and incomplete accompanying descriptive metadata or semantic drift on the conceptual level.” (Lindlar et al. 2020). (Rudnik and Lindlar 2021; Thibodeau 2002)

<sup>27</sup> “Content Preservation For citations it is not sufficient that the bits of the object are still present. The used technology has to reproduce the content ensuring its authenticity even if the original technical environment might be no longer available. This is a shift of perspective from technical stability to technical reusability. [...]” (Aschenbrenner et al. 2011)

<sup>28</sup> “Je länger eine Aufbewahrung erfolgen soll, desto wahrscheinlicher sind Maßnahmen erforderlich, welche die Daten nicht nur technisch, sondern auch inhaltlich nutzbar erhalten (sogenannte Content Preservation). Notwendige Maßnahmen können z.B. Migration der Daten in neue Formate oder Anpassungen an neue Hard- und Softwareumgebungen sein. [...]” (Rat für Informationsinfrastrukturen 2016)

<sup>29</sup> “Logical-Technical Preservation. In addition to D and/or C above the repository takes long-term responsibility for ensuring that the data and metadata are updated over time to newer standards and formats in response to technical risks (e.g. file format or software obsolescence), the changing needs of the designated community (e.g. newer alternate formats become necessary for reuse).” (Andreu et al. 2023)

<sup>30</sup> “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.” (Heseler, Büttner, and Arnold 2024)

Die NFDI AG LTA empfiehlt daher:

- Ein öffentlich gut wahrnehmbares Glossar zu den wichtigsten Begriffen (z.B. auf [forschungsdaten.info](https://www.forschungsdaten.info)) zu schaffen.
- Dass sich die NFDI flankierend an den Bemühungen beteiligt, Begriffe auch im englischsprachigen Diskurs zu vereinheitlichen (vgl. die Publikation von 2020 von Lindlar und Kolleg:innen<sup>31</sup>). Hierfür eignen sich die Kanäle der EOSC.
- Dass den Konsortien Beratung und Aufklärung im Bereich LZA zur Verfügung gestellt wird.
- Sowie Finanzierungsmöglichkeiten für LZA-Lösungen zu schaffen<sup>32</sup>.

Auch bezüglich Standards zu Langzeitarchivierung (OAIS<sup>33</sup>, ISO 14721<sup>34</sup>, ISO/IEC 16363<sup>35</sup>, DIN 31644:2012-04<sup>36</sup>, DIN 31645:2011-11<sup>37</sup>, PREMIS<sup>38</sup>) herrschen unterschiedliche Wahrnehmungen bei den Antworten vor. Ein Standard wird immer dann besonders gut eingehalten, wenn er allen Beteiligten ersichtlich Vorteile verschafft oder hilft, Gefahren zu vermeiden. Standards in anderen Bereichen können die Tendenz haben, wirkungslos zu bleiben oder von neuen Gegebenheiten überholt zu werden. Für Videokassetten dauerte es von den 1970ern bis in die späten 1980er Jahre, bis sich der VHS-Standard durchsetzte. Auf der anderen Seite gelang es in wenigen Jahren, einheitliche Protokolle für das Internet zu schaffen. Im LZA-Kontext nehmen z. B. die Bemühungen, einen international akzeptierten Standard für die software-unabhängige Archivierung von Datenbankstrukturen und -inhalten zu schaffen, zu (siehe DILCIS Board<sup>39</sup>). Sie sind für verschiedene Forschungsgemeinschaften mit langfristiger Datenhaltung in umfangreichen Open Source-Datenbanken wichtig. Das SIARD Konzept und Format basiert dabei auf der Bildung von Information Packages (SIP) nach dem Unicode ISO Standard, XML und SQL:2008 sowie dem URI Internet Standard und dem Industry Standard ZIP<sup>40</sup>. SIARD ist einer der von DILCIS zur eArchivierung verwendeten Standards<sup>41</sup> und SIARD Tools sind in Fachrepositorien und Nationalarchiven im Einsatz. Aktuell wird eine SIARD-Pilotstudie im Kontext NFDI4Biodiversity durchgeführt<sup>42</sup>.

Standards brauchen im Einführungsprozess Einrichtungen, die sie einhalten und auf andere Einrichtungen einwirken. Große Forschungsdaten-Einrichtungen (wie z. B. DESY)

---

<sup>31</sup> (Lindlar et al. 2020)

<sup>32</sup> Siehe auch Empfehlung 96 und 97 der EOSC Long Term Data Preservation Task Force (Andreu et al. 2023)

<sup>33</sup> (The Consultative Committee for Space Data Systems (CCSDS) 2012; 'ISO 14721:2012. Open Archival Information System (OAIS) Reference Model' 2012)

<sup>34</sup> ('ISO 14721:2012. Open Archival Information System (OAIS) Reference Model' 2012)

<sup>35</sup> ('ISO 16363:2012. Audit and Certification of Trustworthy Digital Repositories' 2012)

<sup>36</sup> ('DIN 31644:2012-04. Information Und Dokumentation - Kriterien Für Vertrauenswürdige Digitale Langzeitarchive' 2012)

<sup>37</sup> ('DIN 31645:2011-11. Information Und Dokumentation - Leitfaden Zur Informationsübernahme in Digitale Langzeitarchive' 2011)

<sup>38</sup> (PREMIS Editorial Committee 2015; Guenther, Dappert, and Peyrard 2016)

<sup>39</sup> (DILCIS Board 2024a)

<sup>40</sup> (DILCIS Board and Swiss Federal Archives 2021; SFA 2024)

<sup>41</sup> (DILCIS Board 2024b)

<sup>42</sup> (Weibulat et al. 2023)



haben deshalb auch bereits mit einem Projekt auf EU-Ebene zu Langzeitarchivierungssystemen zusammengearbeitet und dabei LZA-Standards beachtet<sup>43</sup>.

Insofern schlägt die AG LTA vor:

- Die NFDI bekennt sich zu den bestehenden LZA-Standards, zu ihrem Ausbau und Implementierung im Rahmen der FAIR-Prinzipien.
- Die sich an den FAIR-Prinzipien orientierende NFDI unterstützt den Aufbau von fachlicher Expertise im Bereich Langzeitarchivierung von Forschungsdaten und stärkt die Ausbildung an Einrichtungen mit LZA von Forschungsdaten als Kernaufgaben.
- Sie fördert das Engagement von Experten in nationalen und internationalen Initiativen wie z.B. in Research Data Alliance (RDA) Interest Gruppen<sup>44</sup>, Arbeitsgruppen des Digital Information LifeCycle Interoperability Standards Boards (DILCIS) und der EOSC Long Term Data Preservation Task Force (LTDP-TF) zur Konzepterstellung und Weiterentwicklung von Spezifikationen und Standards der Langzeitarchivierung (auf allen drei Leveln: Bitstream, logisch-technisch und semantisch).
- Die NFDI unterstützt Gremien wie CoreTrustSeal<sup>45</sup>, nestor AG Zertifizierung<sup>46</sup>, nestor AG OAIS Review<sup>47</sup>, PREMIS Maintenance Activity and Editorial Committee<sup>48</sup> bei der Dokumentation, konzeptionellen Anpassung und Ausbau von bestehenden Standards, um das Bewusstsein und die Akzeptanz für LZA in NFDI zu verbessern.

## Verteilte oder zentrale LZA-Lösungen in Konsortien

Die Frage, ob ein Konsortium Langzeitarchivierungsdienst(e) anbietet, hat bei der Umfrage zur Angabe einer Vielzahl von Forschungsdaten-Services (siehe auch mit NFDI kooperierende Institutionen<sup>49</sup>) geführt, die alle eigene Charakteristika haben. Sie unterscheiden sich zum Beispiel bezüglich Umfang der Langzeitarchivierung (Content Preservation) oder Existenz einer Preservation Policy. Als Konsequenz und um einen Überblick über tatsächliche und potentielle LZA-Dienste zu bekommen, wurde analysiert, ob in den jeweiligen Konsortien verteilte oder zentrale LZA verfolgt wird. Zentrale LZA heißt hier, dass das Konsortium einen zentralen LZA-Dienst für alle oder fast alle Daten, die im Fokus des Konsortiums sind, anbietet. Tendenziell ist das eine zentrale Organisationseinheit mit z. B. zentralen Standards und einem Geschäftsmodell, wobei das Hosting trotzdem an verschiedenen Institutionen stattfinden kann.

Fast alle Konsortien haben implizit eine verteilte LZA-Lösung angegeben, indem verschiedene existierende Forschungsdaten-Services bei der Frage nach einem

---

<sup>43</sup> (Fernandes et al. 2020; ARCHIVER Project 2019)

<sup>44</sup> Siehe beispielsweise auch (Taylor-Grant, Molloy, and Rhee 2015)

<sup>45</sup> (de Leeuw and Crabtree, n.d.)

<sup>46</sup> (Schoger and Keitel 2020; 'Normenausschuss 009-00-15 AA Records Management und Langzeitverfügbarkeit von Informationsobjekten', n.d.)

<sup>47</sup> ('Nestor - AG OAIS Review' 2022; 'ISO/TC 20/SC 13 - Space Data and Information Transfer Systems' 2021)

<sup>48</sup> ('PREMIS Maintenance Activity and Editorial Committee (Library of Congress)' 2022)

<sup>49</sup> (Miller, Rißler-Pipka, and Weimer 2022)

vorhandenen LZA-Dienst genannt wurden. Das ist im Einklang der NFDI-Bestrebungen, existierende Dienste und Lösungen auszubauen, nachzunutzen und zu vernetzen<sup>50</sup>. Unklar ist allerdings in vielen Fällen, ob diese Forschungsdaten-Services LZA im Sinne von Content Preservation anbieten. Wo ein ausdrückliches Bekenntnis dazu fehlte, war ein bewusster und geplanter Erhalt über das Bitstream Level hinaus nicht anzunehmen. Unklar ist auch oft, ob diese Forschungsdaten-Services Teil des Konsortiums und offizielle Partnerorganisationen sind oder nur logistisch einbezogen wurden, da sie z. B. bekannte Forschungsdaten-Services der Fachcommunity sind.

Eine Klärung der LZA-Abdeckung einer Fachdisziplin könnte durch die flächendeckende Zertifizierung von Forschungsdaten-Services mit dem CoreTrustSeal<sup>51</sup> oder/und dem nestor-Siegel<sup>52</sup> erreicht werden. Darüber hinaus sollten diese Dienste den Umfang der Erhaltungsmaßnahmen, die sie bei Bedarf an den eigenen Daten durchführen, klar benennen<sup>53</sup>. Die Beziehungen von verschiedenen Services zueinander sollten außerdem transparent sein, auch im Kontext Langzeitarchivierung<sup>54</sup>.

Die Nutzung von LZA-Tools zur Formatidentifizierung und Formatvalidierung (z. B. DROID, Siegfried, JHOVE) könnten zum Standard bei Forschungsdaten-Verarbeitung bzw. dem Publikationsprozess werden. Die Ergebnisse dieser Tool-basierten Analyse von Dateien könnten als LZA-Metadaten öffentlich zugänglich gemacht werden. Das gleiche gilt für Metadaten, die Emulation unterstützen.

Es bietet sich an, dass Konsortien in Fragen der Langzeitarchivierung in ihrem Fachbereich eine Koordinierungs- und Monitoring-Funktion übernehmen. Der Einfluss eines deutschen Konsortiums auf internationale Fachdatenbanken und Datenanbieter ist zwar begrenzt, aber auch nicht völlig zu vernachlässigen, gerade da Langzeitarchivierung im internationalen Raum zunehmend präsenter wird<sup>55</sup>.

Die AG LTA empfiehlt daher

- dass Konsortien eine koordinierende Funktion bezüglich der Umsetzung sowie Stand der Umsetzung von Langzeitarchivierung für Daten in ihrem Fokus übernehmen,
- dass Forschungsdaten-Services mit dem CoreTrustSeal oder/und dem nestor-Siegel zertifiziert werden und
- dass den Konsortien und beteiligten Einrichtungen Beratung und Finanzierung im Kontext LZA zur Verfügung gestellt wird.

## Policies und Verantwortungen

Nicht zu unterschätzen im Kontext Langzeitarchivierung ist die Klärung von Verantwortlichkeiten, da eine fehlende oder unklare Aufgabenverteilung und Verantwortung ebenfalls zu Datenverlust führen kann.

---

<sup>50</sup> (Rat für Informationsinfrastrukturen 2016)

<sup>51</sup> (CoreTrustSeal Standards and Certification Board, n.d.)

<sup>52</sup> (Bähr et al. 2019)

<sup>53</sup> Siehe auch Empfehlung 19 der EOSC Long Term Data Preservation Task Force (Andreu et al. 2023)

<sup>54</sup> Siehe auch Empfehlung 40 der EOSC Long Term Data Preservation Task Force (Andreu et al. 2023)

<sup>55</sup> (Andreu et al. 2023)

Es bietet sich an, festzulegen, welche Daten eines Fachbereiches oder einer Fachgemeinschaft langfristigarchiviert werden sollen und diesen Beschluss in einer Policy festzuhalten. Nicht alle im Forschungsprozess generierten Daten müssen langfristigarchiviert werden, aber der Auswahlprozess sollte bewusst und transparent stattfinden. Der RfII nannte die Auswahl für die Langzeitarchivierung 2016 "Relevanzabschätzung" und sah diese Aufgabe und Expertise primär bei den Disziplinen<sup>56</sup>. Ob dies umgesetzt wurde, prüft die Umfrage mit der Frage, ob eine "Collection Policy" existiert. Dabei sollten die Kriterien der Auswahl für die Langzeitarchivierung (auch Bewertung, Appraisal genannt) möglichst transparent sein<sup>57</sup> und klar kommunizierbar sein. Beispiele wären die Zugehörigkeit der Daten zu einem Fachbereich, Copyright und Lizenzierung, Größe eines Datensatzes, Einzigartigkeit des Datensatzes usw.. Wurde bewusst entschieden, dass die Forschenden einer Forschungsorganisation oder eines Fachbereiches diese Entscheidung treffen können und sollen, sollte auch dies in einer Policy dokumentiert werden. Die meisten Konsortien gaben in der Umfrage allerdings an, keine Collection Policy zu haben. Hier besteht also noch immer ein Bedarf, Konsortien oder die Datenrepositorien der Konsortien bezüglich Auswahl für die Langzeitarchivierung zu beraten und zu unterstützen.

Der Umfang der Langzeitarchivierung, welche Erhaltungsmaßnahmen von Personen in welcher Funktion durchgeführt werden können oder auch für welche Zielgruppe die langfristigarchivierten Daten erhalten werden, wird generell in einer Preservation Policy festgehalten<sup>58</sup>. Die Konsortien können auf übergreifender Ebene klären, für welche Daten sie sich verantwortlich sehen, datenspeichernde Institutionen zur Zertifizierung (CoreTrustSeal und nestor Siegel) anhalten und evaluieren, ob die Daten des Fachbereiches ausreichend und angemessen erhalten werden. Das alles kann in einer Preservation Policy auf konsortialer Ebene dokumentiert und veröffentlicht werden<sup>59</sup>. Allerdings konnten die wenigsten Konsortien die Frage nach einer existierenden Preservation Policy (wenn auch nur an einzelnen Institutionen) mit Ja beantworten. Entsprechend gaben auch 37% aller Repositorien in einer europaweiten Umfrage an, keine Preservation Policy zu besitzen. In der Konsequenz wird fehlende Good Practices bei Langzeitarchivierung als eine der primären Herausforderungen genannt<sup>60</sup>.

Die AG LTA empfiehlt,

- klare Zuständigkeiten sowie Verantwortlichkeiten auch bei Erhaltungsmaßnahmen und Beobachtung von technischen Entwicklungen<sup>61</sup> zu definieren,
- einen Beratungsdienst einzurichten, der Konsortien in diesem Kontext unterstützt.

## Nicht abgedeckte Daten und Datentypen

Wie die Umfrage bei den NFDI-Konsortien ergab, existieren in den von den Konsortien repräsentierten Wissenschaftscommunities für zahlreiche Unterlagengruppen und Datentypen noch keine Lösungen für die Langzeitarchivierung im oben genannten Sinne.

---

<sup>56</sup> (Rat für Informationsinfrastrukturen 2016)

<sup>57</sup> Siehe auch Empfehlung 61 der EOSC Long Term Data Preservation Task Force (Andreu et al. 2023)

<sup>58</sup> (nestor-Arbeitsgruppe Policy 2014; nestor working group Policy 2014; Digital Preservation Coalition (DPC) 2023a)

<sup>59</sup> Siehe auch Empfehlung 114 der EOSC Long Term Data Preservation Task Force (Andreu et al. 2023)

<sup>60</sup> (Shearer et al. 2023)

<sup>61</sup> Siehe auch Empfehlungen 71-82 der EOSC Long Term Data Preservation Task Force (Andreu et al. 2023)

Entsprechend gaben elf von 17 Konsortien an, einen LZA-Dienst zu benötigen, zwei antworteten mit "nein" auf diese Frage und bei vier Konsortien war der Bedarf unklar. Bei dieser Auswertung wurde sowohl der generelle Bedarf als auch der spezifische Bedarf für eine LZA-Lösung für bestimmte Datentypen als Bedarf an einem LZA-Dienst interpretiert. Dieser Bedarf betrifft Daten, die aus der Hochleistungsforschung und dem Bereich Citizen Science resultieren, ebenso wie Informationen aus Ursprungssystemen, die in dieser oder ähnlicher Form bereits seit Jahrzehnten in der Forschung, in der Verwaltung oder in anderen Zusammenhängen erwachsen. Für diese können aber auf regionaler, nationaler und internationaler Ebene bereits Standards und best practices existieren, wie etwa die Archivierung von Textdokumenten oder Informationen aus Datenbanken.

Die für die folgenden Datentypen aus der Umfrage resultierenden Lücken weisen daher auch auf den geringen Stellenwert, die mangelnde Auseinandersetzung und die fehlenden Kenntnisse hin, die derzeit für den Bereich der Langzeitarchivierung in der NFDI bestehen. In der Umfrage wurden fehlende LZA-Lösungen für folgende Unterlagen- und Datengruppen bzw. Datenprodukt-Klassen definiert (entsprechend von den Konsortien in der Umfrage genannter Bedarf an einem LZA-Dienst, erweitert durch Autor:innen entsprechend Erfahrungswerten), z. B.:

- große bis sehr große Datensätze (Tera- und Petabyte-Skala) wie etwa aus der Datenmodellierung oder aus multidimensionalen Auswertungsansätzen integrierter Daten (Datacubes)
- Lösungen für die Softwarearchivierung (Research Software)
- Citizen-Science-Daten
- Datenbanken und Datentabellen
- Dokumente und Informationen aus der Projektverwaltung
- Messdaten und Messreihen
- Streaming-Daten
- Social-Media-Daten
- sog. "long tail", -Daten (vgl. auch eine Einschätzung dazu auf EU-Ebene<sup>62</sup>)

In den meisten Bereichen deckt sich dieser Bedarf mit Beobachtungen im internationalen Raum. Laut der Digital Preservation Coalition (DPC) sind sowohl Medien als auch vor allem die Datenformate und Infrastrukturen für Forschungsdaten sehr verwundbar. Kritisch gefährdet oder gefährdet sind der internationalen Umfrage der DPC<sup>63</sup> zufolge:

- obsolete optische und elektromagnetische Medien (z.B. Festplatten, USB-Sticks, DVD-ROMs), auch als Beilage zu gedruckten Forschungsergebnissen
- Forschungsdaten auf GitHub und in Wikis
- Handbücher, Anleitungen, informelle persönliche Korrespondenz zur Forschung (z.B. persönliche Netzlaufwerke, E-Mail-Konten, Nischen des Social Web)
- projektspezifische Datenportale im Internet, die dem technischen Wandel zum Opfer fallen, insbesondere in der Medizin, aber auch in anderen Bereichen der Wissenschaften, z.B. mit Citizen-Science-Beteiligung

---

<sup>62</sup> "A long-tail of data exists that has not been brought into a managed storage, curation, or preservation system." (Andreu et al. 2023)

<sup>63</sup> (Digital Preservation Coalition (DPC) 2023b)

- kommerzielle Software mit einer Rolle im Forschungsprozess, zu der die Firma erlischt, insbesondere Cloud-Dienste.
- Open Source Intelligence aus nichtstaatlichen Quellen und Daten von NGOs, auch Ergebnisse aus deren Analyse.
- audiovisuelle Aufzeichnungen, dreidimensionale Darstellungen aus Architektur und Ingenieurwesen, Datensätze aus dem Bereich der Virtuellen Realität.
- Objekte aller Art, deren ungeklärter Rechtsstatus (Immaterialgüterrecht) einer Erhaltung im Wege steht.

## Konsequenz und Empfehlungen

Langzeitarchivierung ermöglicht den langfristigen Erhalt von wertvollen Forschungsdaten für zehn Jahre und darüber hinaus und wurde auch schon vom RfII als notwendig und wünschenswert angesehen<sup>64</sup>. Allerdings zeigt die Umfrage zu Langzeitarchivierung in den Konsortien der NFDI, dass sich bisher nur wenige Konsortien mit dem Thema im Sinne von Erhalt über das Bitstream Level hinaus auseinandergesetzt haben, geschweige denn entsprechende Lösungen identifiziert haben. Der Anspruch der NFDI, Daten FAIR verfügbar zu machen, wird daher derzeit und soweit absehbar auch in näherer Zukunft nicht vollständig erfüllt werden können und in den meisten Konsortien bestand ein Bedarf für einen LZA-Dienst.

Die AG LTA sieht daher einen Bedarf bei der Aufklärung und Beratung von Konsortien bezüglich der Themenbereiche:

- Was ist LZA und welchen Nutzen hat LZA?
- Welche LZA-Standards gibt es und wie können diese umgesetzt werden?
- Wie lautet die gängige, international anerkannte LZA-Terminologie?
- Welche LZA-Lösungen (strukturelle, technische, organisatorische, personelle) gibt es, wie können diese angebunden werden?
- Welche Verantwortungen sollten geklärt werden?
- Welche Aspekte müssen beim Forschungsdatenmanagement für die angestrebten Erhaltungslevel der LZA berücksichtigt werden?

Bei vielen Konsortien und deren genannten Forschungsdaten-Services besteht Bedarf, Bewusstsein und Funktionalitäten zur LZA zu etablieren oder auszubauen.

Hier braucht es einen Beratungsdienst innerhalb oder in enger Zusammenarbeit mit der NFDI. Das Kompetenznetzwerk nestor<sup>65</sup> muss in diesem Kontext einbezogen werden, da dort schon entsprechende Expertise existiert.

Darüber hinaus empfiehlt die AG LTA:

- Eine finanzielle Unterstützung von LZA-Vorhaben in Institutionen und Konsortien.
- Forschungsförderer müssen Geld für Langzeitarchivierung in Forschungsprojekten bereitstellen.

---

<sup>64</sup> (Rat für Informationsinfrastrukturen 2016)

<sup>65</sup> (nestor, n.d.)

- Forschungsdaten-Services mit dem CoreTrustSeal oder/und dem nestor Siegel zu zertifizieren.
- LZA-Standards mitzugestalten.
- Den Konsortien den Stand der Langzeitarchivierung in ihren Fachbereichen zu beobachten, Lücken aufzuzeigen und Lösungen zusammenzubringen.

## Literaturverzeichnis

- Andreu, Toni, Lluís Anglada, David Antos, Thomas Bähr, Maciej Brzeźniak, Pierre-Yves Burgi, Cécile Cavet, et al. 2023. 'Recommendations Consultation. EOSC-A Long Term Data Preservation Task Force', October. <https://doi.org/10.5281/zenodo.10014698>.
- ARCHIVER Project. 2019. 'PETRA III / EuXFEL DATA ARCHIVING'. ARCHIVER. 19 July 2019. [https://web.archive.org/web/20230308132824/https://www.archiver-project.eu/sites/default/files/Technical%20Summary%20%20-%20PETRAIII\\_EuXFEL%20Data%20Archiving.pdf](https://web.archive.org/web/20230308132824/https://www.archiver-project.eu/sites/default/files/Technical%20Summary%20%20-%20PETRAIII_EuXFEL%20Data%20Archiving.pdf).
- Arora, Ritu, Maria Esteva, and Jessica Trelogan. 2014. 'Leveraging High Performance Computing for Managing Large and Evolving Data Collections'. *International Journal of Digital Curation* 9 (2): 17–27. <https://doi.org/10.2218/ijdc.v9i2.331>.
- Aschenbrenner, Andreas, Harry Enke, Thomas Fischer, and Jens Ludwig. 2011. 'Diversity and Interoperability of Repositories in a Grid Curation Environment'. *Journal of Digital Information* 12 (2). <https://jodi-ojs-tdl.tdl.org/jodi/article/view/1896>.
- Attwood, Teresa K., Bora Agit, and Lynda B. M. Ellis. 2015. 'Longevity of Biological Databases'. *EMBNET.Journal* 21 (0): 803. <https://doi.org/10.14806/ej.21.0.803>.
- Bach, Felix, Andreas Degkwitz, Wolfram Horstmann, Peter Leinen, Michael Puchta, and Thomas Stäcker. 2022. 'Concept for Setting up an LTA Working Group in the NFDI Section "Common Infrastructures"', April. <https://doi.org/10.5281/zenodo.6451456>.
- Bähr, Thomas, Thomas Gerdes, Henk Harmsen, Christian Keitel, Christoph Schmidt, Astrid Schoger, Sabine Schrimpf, et al. 2019. *Erläuterungen Zum Nestor-Siegel Für Vertrauenswürdige Digitale Langzeitarchive, Version 2.1*. Nestor-Arbeitsgruppe Zertifizierung. Frankfurt am Main: nestor c/o Deutsche Nationalbibliothek. <http://nbn-resolving.de/urn:nbn:de:0008-2019062507>.
- Bernard, Lars, Reinhard Altenhöner, Franziska Böhm, Michael Diepenbroek, Barbara Ebert, Juliane Fluck, Sonja Herres-Pawlis, et al. 2023. 'Base4NFDI - Basic Services for NFDI', December. <https://doi.org/10.5281/zenodo.10245518>.
- Bleakly, Denise R. 2002. 'Long-Term Spatial Data Preservation and Archiving: What Are the Issues?' Report SAND2002-0107. Sandia National Laboratories. <https://doi.org/10.2172/793225>.
- CoreTrustSeal Standards and Certification Board. n.d. 'CoreTrustSeal Requirements 2023-2025'. Accessed 29 October 2023. <https://doi.org/10.5281/zenodo.7051012>.
- Deutsche Forschungsgemeinschaft. 2022a. 'Guidelines for Safeguarding Good Research Practice. Code of Conduct. V. 1.1'. <https://doi.org/10.5281/zenodo.6472827>.
- . 2022b. 'Leitlinien Zur Sicherung Guter Wissenschaftlicher Praxis. Kodex. V. 1.1'. <https://doi.org/10.5281/zenodo.6472827>.
- Diepenbroek, Michael, Sonja Schimmler, and Barbara Ebert. 2021. 'Sektionskonzept Common Infrastructures Zur Einrichtung Einer Sektion Im Verein Nationale Forschungsdateninfrastruktur (NFDI) e.V.' *Zenodo*. <https://doi.org/10.5281/zenodo.5607490>.
- Digital Preservation Coalition (DPC). 2023a. 'Preservation Policy Toolkit - Digital Preservation Coalition'. Digital Preservation Policy Toolkit. Version 2. 2023. <https://web.archive.org/web/20231031005940/https://www.dpconline.org/digipres/implement-digipres/policy-toolkit>.
- . 2023b. 'The Bit List 2023. The Global List of Endangered Digital Species. Fourth Edition Revised November 2023'. Digital Preservation Coalition (DPC). <http://doi.org/10.7207/dpcbitlist-23>.
- DILCIS Board. 2024a. 'DILCIS Board Home'. 2024. <https://web.archive.org/web/20240219163250/https://dilcis.eu/>.
- . 2024b. 'Used Standards'. 2024.

- <https://web.archive.org/web/20240219163555/https://dilcis.eu/about/used-standards>.
- DILCIS Board, and Swiss Federal Archives. 2021. 'SIARD 2.2'. SIARD (Software Independent Archiving of Relational Databases). 2021. <https://web.archive.org/web/20240219162844/https://dilcis.eu/content-types/siard>.
- 'DIN 31644:2012-04. Information Und Dokumentation - Kriterien Für Vertrauenswürdige Digitale Langzeitarchive'. 2012. DIN-Norm. Beuth. <https://doi.org/10.31030/1854419>.
- 'DIN 31645:2011-11. Information Und Dokumentation - Leitfaden Zur Informationsübernahme in Digitale Langzeitarchive'. 2011. DIN-Norm. Beuth. <https://doi.org/10.31030/1818277>.
- Doig, James. 2020. 'Preserving the Bits: A Salutory Tale from the National Archives of Australia'. *Digital Preservation Coalition Blog* (blog). 6 November 2020. <https://web.archive.org/web/20240103082849/https://www.dpconline.org/blog/james-doig-preserving-the-bits>.
- Fernandes, João, Bob Jones, Marion Devouassoux, Dominique Buyse, Ruben Van Calenberg, Vaggelis Motsenitalis, and Jakub Urban. 2020. 'ARCHIVER D2.1- State of the Art, Community Requirements and OMC Results'. Zenodo. <https://doi.org/10.5281/zenodo.3974087>.
- Guenther, Rebecca Squire, Angela Dappert, and Sébastien Peyrard. 2016. 'An Introduction to the PREMIS Data Dictionary for Digital Preservation Metadata'. In *Digital Preservation Metadata for Practitioners: Implementing PREMIS*, edited by Angela Dappert, Rebecca Squire Guenther, and Sébastien Peyrard, 23–36. Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-319-43763-7\\_3](https://doi.org/10.1007/978-3-319-43763-7_3).
- Heseler, Jörg, Alexandra Büttner, and Matthias Arnold. 2024. *Grundlagen Der Digitalen Langzeitarchivierung: Eine Handreichung Zur Digitalen Langzeitarchivierung Aus Perspektive Der NFDI4Culture Community. Aufgabenbereich Datenpublikation Und Langzeitarchivierung (TA4)*. Arbeitsgruppe Digitale Publikationen und Daten (NFDI4Culture). <https://doi.org/10.5281/zenodo.10423858>.
- 'ISO 14721:2012. Open Archival Information System (OAIS) Reference Model'. 2012. ISO-Standard. <https://web.archive.org/web/20231211011858/http://www.iso.org/standard/57284.html>.
- 'ISO 16363:2012. Audit and Certification of Trustworthy Digital Repositories'. 2012. ISO-Standard. <https://web.archive.org/web/20231223120511/https://www.iso.org/standard/56510.html>.
- 'ISO/TC 20/SC 13 - Space Data and Information Transfer Systems'. 2021. ISO. 27 August 2021. <https://web.archive.org/web/20231215060030/https://www.iso.org/committee/46612.html>.
- Leeuw, Lisa de, and Jonathan Crabtree. n.d. 'CoreTrustSeal Maintenance WG'. Research Data Alliance (RDA). Accessed 20 December 2023. <https://web.archive.org/web/20230610213245/https://www.rd-alliance.org/groups/repository-audit-and-certification-dsa%E2%80%93partnership-wg.html>.
- Lindlar, Michelle, Pia Rudnik, Sarah Jones, and Laurence Horton. 2020. "'You Say Potato, I Say Potato" Mapping Digital Preservation and Research Data Management Concepts towards Collective Curation and Preservation Strategies'. *International Journal of Digital Curation* 15 (1): 26. <https://doi.org/10.2218/ijdc.v15i1.728>.
- Miller, Bernhard, Nanette Rißler-Pipka, and Lukas Weimer. 2022. 'Infrastructure Networks within NFDI: Survey among Institutions Partnering with NFDI Consortia'. Zenodo.



- <https://doi.org/10.5281/zenodo.6474644>.
- nestor. n.d. 'Nestor - Startseite'. Accessed 21 December 2023.  
[https://web.archive.org/web/20231221110257/https://www.langzeitarchivierung.de/Webs/nestor/DE/Home/home\\_node.html](https://web.archive.org/web/20231221110257/https://www.langzeitarchivierung.de/Webs/nestor/DE/Home/home_node.html).
- 'Nestor - AG OAIS Review'. 2022. Nestor – Kompetenznetzwerk Langzeitarchivierung Und Langzeitverfügbarkeit Digitaler Ressourcen in Deutschland e.V. 3 January 2022.  
[https://web.archive.org/web/20230602143603/https://www.langzeitarchivierung.de/Webs/nestor/DE/Arbeitsgruppen/AG\\_OAIS\\_Review/ag\\_oais\\_review\\_node.html](https://web.archive.org/web/20230602143603/https://www.langzeitarchivierung.de/Webs/nestor/DE/Arbeitsgruppen/AG_OAIS_Review/ag_oais_review_node.html).
- nestor working group Policy. 2014. 'Guidelines for the Creation of an Institutional Policy on Digital Preservation'. Frankfurt am Main : nestor – Network of Expertise in Long-Term Storage c/o Deutsche Nationalbibliothek. <https://nbn-resolving.de/urn:nbn:de:0008-2014111006>.
- nestor-Arbeitsgruppe Policy. 2014. 'Leitfaden Zur Erstellung Einer Institutionellen Policy Zur Digitalen Langzeitarchivierung'. Frankfurt am Main : nestor - Kompetenznetzwerk Langzeitarchivierung c/o Deutsche Nationalbibliothek. <https://nbn-resolving.de/urn:nbn:de:0008-2014052004>.
- NFDI. 2023. 'Konsortien | NFDI'. Verschiedene Disziplinen – Ein Ziel. 5 October 2023.  
<https://web.archive.org/web/20231005183731/https://www.nfdi.de/konsortien/>.
- 'Normenausschuss 009-00-15 AA Records Management und Langzeitverfügbarkeit von Informationsobjekten'. n.d. din-de. Accessed 21 December 2023.  
<https://web.archive.org/web/20231221153546/https://www.din.de/de/mitwirken/normenausschuesse/nid/nationale-gremien/wdc-grem:din21:54774796>.
- Pinnick, Jaana. 2017. 'Exploring Digital Preservation Requirements: A Case Study from the National Geoscience Data Centre (NGDC)'. *Records Management Journal* 27 (2): 175–91. <https://doi.org/10.1108/RMJ-04-2017-0009>.
- Pinnick, Jaana, Andrew Riddick, Robert McLaverty, and Garry Baker. 2018. 'A Case Study: Management and Exploitation of the Nuclear Decommissioning Agency Geoscience Data Archive'. *New Review of Information Networking* 23 (1–2): 99–110.  
<https://doi.org/10.1080/13614576.2018.1544089>.
- PREMIS Editorial Committee. 2015. 'PREMIS'.  
<https://web.archive.org/web/20231209121615/https://loc.gov/standards/premis/>.
- 'PREMIS Maintenance Activity and Editorial Committee (Library of Congress)'. 2022. 1 February 2022.  
<https://web.archive.org/web/20231227174834/https://www.loc.gov/standards/premis/ma.html>.
- Rat für Informationsinfrastrukturen. 2016. 'Leistung Aus Vielfalt. Empfehlungen Zu Strukturen, Prozessen Und Finanzierung Des Forschungsdatenmanagements in Deutschland'. Göttingen. urn:nbn:de:101:1-201606229098.
- Rudnik, Pia, and Micky Lindlar. 2021. 'Foliensatz: Crashkurs Digitale Langzeitarchivierung - Ebenen-Modelle der LZA (3-Ebenen-Modell von K. Thibodeau & Levels of Digital Preservation der NDSA)'. February 26. <https://doi.org/10.5281/zenodo.4761975>.
- Schoger, Astrid, and Christian Keitel. 2020. 'Nestor - AG Zertifizierung'. Nestor – Kompetenznetzwerk Langzeitarchivierung Und Langzeitverfügbarkeit Digitaler Ressourcen in Deutschland e.V. 12 February 2020.  
[https://web.archive.org/web/20231220125132/https://www.langzeitarchivierung.de/Webs/nestor/DE/Arbeitsgruppen/AG\\_Zertifizierung/ag\\_zertifizierung\\_node.html](https://web.archive.org/web/20231220125132/https://www.langzeitarchivierung.de/Webs/nestor/DE/Arbeitsgruppen/AG_Zertifizierung/ag_zertifizierung_node.html).
- SFA, Swiss Federal Archives. 2024. 'SIARD Suite'. 3 January 2024.  
<https://web.archive.org/web/20240219162542/https://www.bar.admin.ch/bar/en/home/archivierung/tools---hilfsmittel/siard-suite.html>.
- Shearer, Kathleen, Silvia Mirlene Nakano Koga, Eloy Rodrigues, Natalia Manola, martine Pronk, and Vanessa Proudman. 2023. 'Current State and Future Directions for Open Repositories in Europe'. Zenodo. <https://doi.org/10.5281/zenodo.10255559>.

- Sprout, Bronwen, and Mark Jordan. 2018. 'Distributed Digital Preservation: Preserving Open Journal Systems Content in the PKP PN'. *Digital Library Perspectives* 34 (4): 246–61. <https://doi.org/10.1108/DLP-11-2017-0043>.
- Taylor-Grant, Rebecca, Laura Molloy, and Hea Lim Rhee. 2015. 'Archives and Records Professionals for Research Data IG'. RDA. 26 November 2015. <https://web.archive.org/web/20240221074127/https://www.rd-alliance.org/groups/archives-records-professionals-for-research-data.html>.
- The Consultative Committee for Space Data Systems (CCSDS). 2012. 'Reference Model for an Open Archival Information System (OAIS). Recommended Practice CCSDS 650.0-M-2'. <https://web.archive.org/web/20231230090114/https://public.ccsds.org/Pubs/650x0m2.pdf>.
- Thibodeau, Kenneth. 2002. 'Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years'. In *The State of Digital Preservation: An International Perspective*, 4–31. Washington, D.C.: Council on Library and Information Resources. <https://web.archive.org/web/20230822111705/http://www.clir.org/wp-content/uploads/sites/6/pub107.pdf>.
- Weibulat, Tanja, Anna Lisa Schwartz, Stefan Seifert, Markus Schmalzl, and Dagmar Triebel. 2023. 'Was Haben Mehлтаupilze Mit Digitaler Langzeitarchivierung Zu Tun?' July 25. <https://doi.org/10.5281/zenodo.8268840>.
- Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. 'The FAIR Guiding Principles for Scientific Data Management and Stewardship'. *Scientific Data* 3 (1): 160018. <https://doi.org/10.1038/sdata.2016.18>.

## Anhang

### Umfrage (Deutsch)

Die Langzeitarchivierung (und –verfügbarkeit) von Forschungsdaten bedarf einer kontinuierlichen Aufmerksamkeit und Weiterentwicklung und gewinnt mit einem ständig überproportionalen Wachstum der Datenmenge sowie der Ausweitung auf immer neue Datentypen an Dringlichkeit, sei es durch Anforderungen von Förderern - die DFG verlangt 10 Jahre Aufbewahrung - oder der Aufgabe digitale Kultur- und Wissensobjekte auf unbestimmte Zeit zu speichern und verfügbar zu halten. Dabei ist zu unterscheiden zwischen dem Erhalt auf der Ebene des Bitstream und der funktionalen Erhaltung. Letzteres umfasst die Archivierung der digitalen (Daten-)Objekte in ihrem lesbaren, funktionalen, i.e. logischen, semantischen und relationalen Kontext, was je nach Datentyp und -format einen weitaus größeren Aufwand bedeutet als die reine Erhaltung des Bitstream. Mögliche Erhaltungsmaßnahmen sind z. B. Migration von Datenfiles in ein neues Dateiformat, Migration von Systemen oder Emulation der ursprünglichen Systemumgebung über 10 und mehr Jahre hinweg. Langzeitarchivierung strebt Erhalt sowohl auf Bitstream als auch auf funktionalem Level an.

Innerhalb der Sektion Common Infrastructure der NFDI hat sich die AG LTA formiert, die sich u.a. zum Ziel gesetzt hat:

- Bedarfe und Anforderungen der Konsortien an LZA-Services zu erheben
- Bestehende Service-Angebote und entwickelte Prozesse zur LZA zu dokumentieren
- Angebote zu vernetzen und Bedarfe und Angebote wo immer möglich zusammenzubringen (Broker)
- Übergreifende Services bedarfsorientiert zu entwickeln
- Geschäftsmodelle zu entwickeln und zu etablieren

#### Fragen zu **Angeboten** des Konsortiums

1. Bietet das Konsortium Langzeitarchivierungsdienst(e) im obigen Sinne an?

Wenn ja:

- a. An wen richtet sich dieser Dienst, von wem kann/darf er genutzt werden?
- b. Gibt es Auswahlverfahren für Daten zur Langzeitarchivierung (collection policy)?
- c. Welche Archivierungsansätze (bitstream, funktionale Ebene) werden angeboten?
- d. Welche digitalen Objekte (digital entities) sind zur Archivierung vorgesehen ?
- e. Gibt es zu diesem Dienst Geschäftsmodelle (Kosten, Verträge, ...)?
- f. Hat der Dienst oder die anbietende Institution eine Preservation Policy?
- g. Wurde der Dienst zertifiziert (nestor-Siegel, CoreTrustSeal bzw. Data Seal of Approval)?
- h. Gibt es zur Nutzung des Archivs abgestimmte Prozesse und Rollenmodelle?
- i. Welche LZA Standards und Protokolle werden dabei berücksichtigt bzw. verwendet (OAIS, PREMIS, ggf. Community spezifische LZA-Standards, ...)?
- j. Welche technischen Rahmenbedingungen gibt es zu berücksichtigen?
  - i. Gibt es Spezialisierungen auf bestimmte Datentypen (Texte, Videos, Messdaten,...)?
  - ii. Gibt es Begrenzungen auf z. B. spezifische Dateiformate?
  - iii. Gibt es Begrenzungen der Datenmenge?

Wenn nein:

- k. Ist ein solcher Dienst geplant?
- l. Nutzen Sie Angebote außerhalb des Konsortiums?
- m. Planen Sie, ein vorhandenes Repositorium zu einem LZA-Dienst auszubauen?  
Wenn ja, welches?

#### Fragen zu **Bedarfen** des Konsortiums

1. Besteht im Konsortium ein Bedarf nach einem LZA Dienst bzw. Ausbau eines vorhandenen LZA-Dienstes?

Wenn ja:

- a. Welche technischen Rahmenbedingungen gibt es zu berücksichtigen?
  - i. Datentypen
  - ii. Dateiformate
  - iii. Datenmenge
- b. Welche anderen Rahmenbedingungen wie z.B. Zugangsbeschränkungen und Sicherheitsanforderungen sind zu beachten?

## Umfrage (Englisch)

Long Term Archiving (LTA) of research data needs continuous attention and development. Developing an approach and solutions regarding LTA becomes also increasingly urgent due to fast growth of data volume and use of new data types but also due to requirements of funders – DFG requires retention of data for 10 years – or due to the responsibility of storing and making cultural objects accessible for an indeterminate timespan. Preservation of data differs between preservation on the bitstream level and preservation on the functional level. The functional level contains archiving of data objects in their readable, functional, i. e. logical, semantic and relational context, which requires more effort than preservation on the bitstream level. Possible preservation measures are for example migration to new file formats, migration to new systems or emulation of system environments for 10 years or more. Long Term Archiving aims to preserve data on bitstream level as well as on the functional level.

Within the Section Common Infrastructure of NFDI the Working Group LTA was formed to work on:

- Collecting needs and requirements of consortia regarding a LTA-service
- Documenting existing services and developing processes for LTA
- Connecting services and needs where possible (Broker)
- Developing and establishing business models

Questions about **services** of the consortium:

1. Does the consortium offer a LTA-service (as described above)?

If yes

- a. Who are the intended users of the service, who can/is allowed to use the service?
- b. Are data selected for long term preservation (collection policy)?
- c. Which preservation levels are offered (bitstream, functional level)?
- d. Which digital objects (digital entities) are intended for archiving?
- e. Does the service have a business model (costs, contracts, ...)?

- f. Does the service or the institution where the service is located have a preservation policy?
- g. Is the service certified (nestor seal, CoreTrustSeal, Data Seal of Approval)?
- h. Are there processes and role models for use of the LTA-service?
- i. Which LTA-Standards and protocols are used (OAIS, PREMIS, possibly community-specific LTA-standards, ...)?
- j. Are there technical factors that need to be taken into consideration?
  - i. Specialization on specific data types (text, videos, measurement data,...)?
  - ii. Limitation regarding e. g. specific file formats?
  - iii. Limitations regarding data volume?

If no

- k. Is the consortium planning a LZA-service?
- l. Is the consortium using LTA-services outside the consortium?
- m. Is the consortium planning to develop an existing repository towards a LTA-service? If yes, which repository?

Questions about the **needs** of the consortium

1. Is there a need in the consortium for a LTA-Service or further development and expansion of an existing LTA-Service?

If yes:

- a. Which technical factors need to be taken into consideration?
  - i. File types
  - ii. File formats
  - iii. Data amounts
- b. Which other factors, like e. g. access restrictions and security demands need to be taken into account?

## Impressum

### Zitiervorschlag

Markus, Katharina; Naumann, Kai; Schmalzl, Markus; Watson, Juliane & Triebel, Dagmar (2024). Langzeitarchivierung in der NFDI. Tendenzen und Bedarfe als Resultat einer Umfrage. 10.5281/zenodo.10822613

### Finanzierung

Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) als Teil des Projektes 'Aufbau einer Nationalen Forschungsdateninfrastruktur' - NFDI4Microbiota - Projektnummer

460129525, NFDI4Biodiversity - Projektnummer 442032008 und durch EU COST Action CA17106 (<https://www.cost.eu/actions/CA17106/>, WG4).

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 \(CC BY 4.0\)](#). Für die enthaltenen Materialien gelten, sofern angegeben, andere Lizenzen bzw. sind alle Rechte vorbehalten.