

H. Neuroth, A. Oßwald, R. Scheffel, S. Strathmann, K. Huth (Hrsg.)

nestor Handbuch

Eine kleine Enzyklopädie
der digitalen Langzeitarchivierung

Version 2.3

Kapitel 8.2

Bitstream Preservation

nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung
hg. v. H. Neuroth, A. Oßwald, R. Scheffel, S. Strathmann, K. Huth
im Rahmen des Projektes: nestor – Kompetenznetzwerk Langzeitarchivierung und
Langzeitverfügbarkeit digitaler Ressourcen für Deutschland
nestor – Network of Expertise in Long-Term Storage of Digital Resources
<http://www.langzeitarchivierung.de/>

Kontakt: editors@langzeitarchivierung.de
c/o Niedersächsische Staats- und Universitätsbibliothek Göttingen,
Dr. Heike Neuroth, Forschung und Entwicklung, Papendiek 14, 37073 Göttingen

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen
Nationalbibliografie; detaillierte bibliografische Daten sind im Internet unter
<http://www.d-nb.de/> abrufbar.

Neben der Online Version 2.3 ist eine Printversion 2.0 beim Verlag Werner Hülsbusch,
Boizenburg erschienen.

Die digitale Version 2.3 steht unter folgender Creative-Commons-Lizenz:
„Namensnennung-Keine kommerzielle Nutzung-Weitergabe unter gleichen Bedingungen 3.0
Deutschland“
<http://creativecommons.org/licenses/by-nc-sa/3.0/de/>



Markenerklärung: Die in diesem Werk wiedergegebenen Gebrauchsnamen, Handelsnamen,
Warenzeichen usw. können auch ohne besondere Kennzeichnung geschützte Marken sein und
als solche den gesetzlichen Bestimmungen unterliegen.

URL für Kapitel 8.2 „Bitstream Preservation“ (Version 2.3): [urn:nbn:de:0008-20100305123](http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:0008-20100305123)
<http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:0008-20100305123>



Gewidmet der Erinnerung an Hans Liegmann (†), der als Mitinitiator und früherer Herausgeber des Handbuchs ganz wesentlich an dessen Entstehung beteiligt war.

8.2 Bitstream Preservation

Dagmar Ulbrich

Grundlage aller Archivierungsaktivitäten ist der physische Erhalt der Datenobjekte, die Bitstream¹ Preservation. Es wird eine Speicherstrategie vorgeschlagen, die auf einer redundanten Datenhaltung auf mindestens zwei unterschiedlichen, marktüblichen und standardisierten Speichertechniken basiert. Die eingesetzten Speichermedien sollten regelmäßig durch aktuelle ersetzt werden, um sowohl dem physischen Verfall der Speichermedien als auch dem Veralten der eingesetzten Techniken vorzubeugen. Es werden vier Arten von Migrationsprozessen vorgestellt. Das sind: Refreshment, Replication, Repackaging und Transformation. Als Medienmigration im engeren Sinne werden nur die beiden ersten, Refreshment und Replication, angesehen. Sie bezeichnen das Auswechseln einzelner Datenträger (refreshing) oder eine Änderung eingesetzter Speicherverfahren (replication). Durch die kurzen Lebenszyklen digitaler Speichermedien erfolgt ein Erneuern der Trägermedien oft im Rahmen der Aktualisierung der eingesetzten Speichertechnik.

Physischer Erhalt der Datenobjekte

Um digitale Daten langfristig verfügbar zu halten, muss an zwei Stellen angesetzt werden. Zum einen muss der physische Erhalt des gespeicherten Datenobjekts (Bitstreams) auf einem entsprechenden Speichermedium gesichert werden. Zum anderen muss dafür Sorge getragen werden, dass dieser Bitstream auch interpretierbar bleibt, d.h. dass eine entsprechende Hard- und Software-Umgebung verfügbar ist, in der die Daten für einen menschlichen Betrachter lesbar gemacht werden können. Ohne den unbeschädigten Bitstream sind diese weiterführenden Archivierungsaktivitäten sinnlos. Der physische Erhalt der Datenobjekte wird auch als „Bitstream Preservation“ bezeichnet. Für den physischen Erhalt des Bitstreams ist eine zuverlässige Speicherstrategie erforderlich.

1 Der Begriff „Bitstream“ wird hier als selbsterklärend angesehen. Eine Erläuterung des Begriffs findet sich in: Rothenberg, Jeff (1999): *Ensuring the Longevity of Digital Information*. <http://www.clir.org/pubs/archives/ensuring.pdf>

Bei diesem Text handelt es sich um eine ausführlichere Fassung eines gleichnamigen Artikels, der 1995 in der Zeitschrift „Scientific American“, Band 272, Nummer 1, Seiten 42-47 erschienen ist.

Alle hier aufgeführten URLs wurden im Mai 2010 auf Erreichbarkeit geprüft .

Verfahrensvorschläge für eine Bitstream Preservation

Die nachstehenden vier Verfahrensvorschläge können als Grundlage für eine zuverlässige Speicherstrategie zur Sicherstellung des physischen Erhalts der archivierten Datenobjekte verwendet werden:²

1. *Redundante Datenhaltung*: Die Daten sollten in mehrfacher Kopie vorliegen. Zur Sicherung gegen äußere Einflüsse empfiehlt sich auch eine räumlich getrennte Aufbewahrung der unterschiedlichen Kopien.
2. *Diversität eingesetzter Speichertechnik*: Die Daten sollten auf mindestens zwei unterschiedlichen Datenträgertypen gesichert werden.
3. *Standards*: Die verwendeten Speichermedien sollten internationalen Standards entsprechen und auf dem Markt eine weite Verbreitung aufweisen.
4. *Regelmäßige Medienmigration*: Die verwendeten Speichertechniken bzw. Datenträger müssen regelmäßig durch neue ersetzt werden.

Redundanz, Speichertechniken und Standards

Eine mehrfach *redundante Datenhaltung* ist in vielen Bereichen der Datensicherung üblich. Bei wertvollen, insbesondere bei nicht reproduzierbaren Daten wird man sich nicht auf eine einzige Kopie verlassen wollen. Um das Risiko äußerer Einflüsse wie Wasser- oder Brandschäden zu verringern, bietet sich die räumlich getrennte Aufbewahrung der Kopien an. Um auch die Gefahr eines Datenverlusts durch menschliches Versagen oder Vorsatz einzuschränken, kann eine Aufbewahrung bei zwei unabhängigen organisatorischen Einheiten in das Redundanzszenario mit einbezogen werden. Zusätzliche Sicherheit lässt sich gewinnen, indem die jeweiligen Kopien *auf unterschiedlichen Speichertechniken* gehalten werden. Dies mindert das Risiko eines Datenverlusts durch Veralterung einer der eingesetzten Techniken. Sofern vorhanden, sollten Fehlererkennungs- und Korrekturmechanismen zur Sicherung der Datenintegrität eingesetzt werden. Weiter sollte die Funktionstüchtigkeit der Speichermedien und Lesegeräte anhand von Fehlerstatistiken überwacht werden. Die sachgerechte Handhabung von Datenträgern und Lesegeräten ist in jedem Fall vorauszusetzen. Alle

2 Die Auflistung erhebt keinen Anspruch auf Vollständigkeit. Ähnliche Aufstellungen finden sich z.B. in: Rathje, Ulf (2002): Technisches Konzept für die Datenarchivierung im Bundesarchiv. In: Der Archivar, H. 2, Jahrgang 55, S.117-120, <http://www.bundesarchiv.de/imperia/md/content/abteilungen/abtb/1.pdf> und: o.V. (o.J.) Digital preservation. Calimera Guidelines. S.3. http://www.calimera.org/Lists/Guidelines%20PDF/Digital_preservation.pdf

verwendeten Speichertechniken bzw. -medien sollten auf internationalen *Standards* basieren und über eine möglichst breite Nutzergruppe verfügen.

Regelmäßige Medienmigration

Als Medienmigration kann jeder Vorgang betrachtet werden, bei dem das physische Trägermedium eines Datenobjekts innerhalb eines Archivs geändert und der Vorgang mit der Absicht durchgeführt wird, das Datenobjekt zu erhalten, indem die alte Instanz durch die neue ersetzt wird. Eine entsprechende Definition von „Digital Migration“ findet sich im OAIS-Referenzmodell³:

Digital Migration is defined to be the transfer of digital information, while intending to preserve it, within the OAIS. It is distinguished from transfers in general by three attributes:

- *a focus on the Preservation of the full information content*
- *a perspective that the new archival implementation of the information is a replacement for the old; and*
- *full control and responsibility over all aspects of the transfer resides with the OAIS.*

Im OAIS-Referenzmodell werden vier Arten der Migration genannt: Refreshment, Replication, Repackaging und Transformation.⁴

Refreshment: Als Refreshment werden Migrationsprozesse bezeichnet, bei denen einzelne Datenträger gegen neue, gleichartige Datenträger ausgetauscht werden. Die Daten auf einem Datenträger werden direkt auf einen neuen Datenträger gleichen Typs kopiert, der anschließend den Platz des alten in der Speicherinfrastruktur des Archivs einnimmt. Weder an den Daten noch an der Speicherinfrastruktur werden also Änderungen vorgenommen, es wird lediglich ein Datenträger gegen einen gleichartigen anderen ausgetauscht.

Replication: Eine Replication ist ein Migrationsprozess, bei dem ebenfalls Daten von einem Datenträger auf einen neuen kopiert werden. Bei der Replica-

3 Consultative Committee for Space Data Systems (CCSDS) (2002): *Reference Model for an Open Archival Information System (OAIS)*. Blue Book. Washington DC. Seite 5-1. vgl. auch Kapitel 4.
<http://public.ccsds.org/publications/archive/650x0b1.pdf>

4 Consultative Committee for Space Data Systems (CCSDS) (2002): *Reference Model for an Open Archival Information System (OAIS)*. A.a.O. Seite 5-4.

tion jedoch kann es sich bei dem neuen Datenträger auch um einen andersartigen, z.B. aktuelleren, handeln. Andersartige Datenträger erfordern eine entsprechende Anpassung der Speicherinfrastruktur. Der neue Datenträger kann in der Regel nicht unmittelbar den Platz des alten einnehmen. Der wesentliche Unterschied zum Refreshment liegt daher in den mit dem Kopierprozess einhergehenden Änderungen der verwendeten Speicherinfrastruktur.

Repackaging: Ein Repackaging ist ein Migrationsprozess, bei dem ein sogenanntes Archivpaket verändert wird. Diese Änderung betrifft nicht die eigentlichen Inhaltsdaten, sondern die Struktur des Archivpakets.

Transformation: eine Transformation ist ein Migrationsprozess, bei dem auch die Inhaltsdaten des Archivpakets verändert werden.

Refreshment und Replication können als Medienmigrationen im engeren Sinne angesehen werden. Der Umkopierprozess erfolgt in beiden Fällen mit der Absicht, das Trägermedium zu ersetzen, unabhängig davon, welche Inhalte auf ihm abgelegt sind. Die Replication wird im Folgenden im Sinne eines Technologiewechsels interpretiert.⁵ Ein Refreshment beschränkt sich dagegen auf den Wechsel einzelner Datenträger innerhalb einer Speichertechnik, z.B. einer Magnetbandgeneration. Bei Repackaging und Transformation dagegen werden auch die Datenobjekte selbst umgeschrieben. Ein Beispiel für ein Repackaging ist die Änderung des Packformats von ZIP zu TAR. Eine Formatmigration, z.B. von JPG zu TIFF, ist dagegen eine Transformation, da die Inhalte des Archivpakets verändert werden. Die Unterscheidung dieser vier Arten von Migrationen erleichtert die begriffliche Abgrenzung einer Medienmigration von einer Formatmigration. Eine Formatmigration umfasst letztlich immer auch eine Medienmigration, da ein neues Datenobjekt erstellt und auf einem eigenen Trägermedium abgelegt wird. Die Formatmigration erfolgt aber mit Blick auf die künftige Interpretierbarkeit des Bitstreams, die Medienmigration im engeren Sinne hingegen dient dessen Erhalt. Für die Bitstream Preservation sind nur die beiden ersten, Refreshment und Replication, wesentlich, da die beiden anderen den Bitstream verändern. Ein Refreshment ist in der Regel weniger aufwändig als eine Replication, da nicht das Speicherungsverfahren, sondern nur einzelne Datenträger erneuert werden.

5 Eine Replication muss nach der zitierten Definition nicht notwendig von einem veralteten Medium auf ein aktuelleres erfolgen, sondern ggf. auch auf ein gleichartiges. In der Praxis wird das aber eher selten der Fall sein.

Refreshment und Replication

Ein Erneuern (refreshing) einzelner Datenträger kann aufgrund von Fehlerraten oder auf der Basis bestimmter Kriterien wie Zugriffshäufigkeit oder Alter erfolgen. Der Aufwand solcher Maßnahmen ist gegen die Wahrscheinlichkeit eines Datenverlusts durch einen fehlerhaften Datenträger abzuwägen. Auf der einen Seite können zusätzliche Kontrollverfahren eine sehr hohe Systemlast erzeugen, die den aktiven Zugriff auf die Daten beträchtlich einschränken kann. Zudem sind die Beurteilungskriterien wie Zugriffshäufigkeit, Alter und ggf. die tolerierbare Fehlerrate oft strittig und zum Teil nur mit teurer Spezialsoftware oder auch gar nicht feststellbar. Nicht selten können sie auch im Einzelfall durch Unterschiede in Produktionsablauf oder Handhabung zwischen Datenträgern desselben Typs stark variieren. Auf der anderen Seite wird die Haltbarkeit von Trägermedien aufgrund des raschen Technologiewandels meist gar nicht ausgereizt. Die Wahrscheinlichkeit schadhafter Datenträger durch altersbedingten Verfall ist daher eher gering. Um diesen Zusammenhang deutlich zu machen, kann die durchschnittliche Lebensdauer eines Datenträgers von seiner durchschnittlichen Verfallszeit unterschieden werden.⁶

„*Medium Expected Lifetime (MEL)*: The estimated amount of time the media will be supported and will be operational within the electronic deposit system.”

“*Medium Decay Time (MDT)*: The estimated amount of time the medium should operate without substantial read and write errors.”

Die Definition der durchschnittlichen Lebensdauer enthält zwei durch „und“ verbundene Zeitangaben. Die eine bezieht sich auf die Dauer der Unterstützung eines Speichermediums durch den Hersteller, die andere auf die Dauer des Einsatzes eines Speichermediums im digitalen Archiv. Diese beiden Zeitspannen können durchaus differieren. Nicht selten zwingt die wegfallende Unterstützung durch den Hersteller zur Migration, auch wenn die vorhandenen Systeme voll funktionsfähig sind und noch weiter betrieben werden könnten. Für Speichertechniken, die vom Hersteller nicht mehr unterstützt werden, können Ersatzteile oder technische Betreuung nicht mehr garantiert werden. Ihr Weiterbetrieb ist daher nicht ratsam. Der Begriff der durchschnittlichen Lebensdauer wird aus diesen Gründen hier als die durchschnittlich zu erwartende Hersteller-Unterstützung interpretiert. Solange diese durchschnittliche Lebensdauer unter der durchschnittlichen Verfallszeit liegt, ist ein Ausfall einzelner

6 Van Diessen, Raymond J. und van Rijnssoever, Ben J. (2002): *Managing Media Migration in a Deposit System. IBM/KB Long-Term Preservation Study Report Series Nr. 5*. Amsterdam: IBM Niederlande. S.4.
<http://www-5.ibm.com/nl/dias/resource/migration.pdf>

Datenträgern selten zu erwarten. Statt aufwändiger Kontrollen der Datenträger kann es in diesem Fall einfacher sein, auf eine redundante Datenhaltung zu vertrauen, im konkreten Fehlerfall einzelne Datenträger oder Laufwerke zu ersetzen und den gesamten Bestand im Rahmen eines Technologiewechsels (Replication) komplett auszutauschen.

Eine Replication im Sinne eines Technologiewechsels umfasst Änderungen in der bestehenden Speicherinfrastruktur. Erforderliche Technologiewechsel können sehr unterschiedlich ausfallen. Sie können von einer Magnetbandgeneration zur nächsten reichen oder einen vollständigen Wechsel z.B. von Magnetbändern zu optischen Medien bedeuten. Im ersten Schritt muss die neue Speichertechnik in die bestehende Infrastruktur integriert werden. Anschließend müssen die betroffenen Datenbestände von der alten Technik auf die neue umkopiert werden. Bei großen Datenmengen mit ggf. hohen Sicherheits- oder Verfügbarkeitsansprüchen können diese Umkopierprozesse aufwändig und langwierig sein. Die Lesegeschwindigkeit der älteren Speichermedien wird in der Regel langsamer sein als die Schreibgeschwindigkeit der neuen. Beide müssen für einen Kopierprozess koordiniert werden, ggf. über Zwischenspeicher. Der Übertragungsvorgang muss abgeschlossen sein, bevor die alte Speichertechnik unbrauchbar wird. An diesem Punkt sei auf die oben ausgeführte Interpretation von „Medium Expected Lifetime“ hingewiesen. Dass der Migrationsprozess abgeschlossen sein muss, bevor eine Speichertechnik nicht mehr auf dem Markt ist, wäre ein sehr hoher Anspruch, da viele Speichermedien nur drei bis fünf Jahre lang angeboten werden. Unter Umständen kann ein solcher Anspruch je nach Wert der betroffenen Daten gerechtfertigt sein. Häufig bieten Hersteller die Unterstützung von Speichermedien einige Jahre länger an, als diese Technik aktiv vertrieben wird. Dies verlängert die zuverlässige Einsatzdauer von Speichertechniken. Eine zusätzliche Sicherheit kann in diesem Kontext auch der Verfahrensvorschlag unterschiedliche Speichertechniken einzusetzen bieten.

Zusammenfassung

Ein Langzeitarchiv muss über zuverlässige Speicherstrategien verfügen, die nicht nur ein „Refreshment“ eingesetzter Datenträger innerhalb einer Speichertechnik ermöglichen, sondern darüber hinaus auch die Erneuerung ganzer Speichertechniken. Solche Strategien müssen sicherstellen, dass zu keinem Zeitpunkt Datenbestände unzugänglich werden, weil ihre Trägermedien nicht mehr lesbar sind.

Literatur

- Rothenberg, Jeff (1999), *Ensuring the Longevity of Digital Information*. <http://www.clir.org/pubs/archives/ensuring.pdf>
Bei diesem Text handelt es sich um eine ausführlichere Fassung eines gleichnamigen Artikels, der 1995 in der Zeitschrift „Scientific American“, Band 272, Nummer 1, Seiten 42-47 erschienen ist.
- Rathje, Ulf (2002): *Technisches Konzept für die Datenarchivierung im Bundesarchiv*. In: *Der Archivar*, H. 2, Jahrgang 55, S.117-120.
http://www.archive.nrw.de/archivar/hefte/2002/Archivar_2002-2.pdf
- o.V. (o.J.) *Digital preservation*. Calimera Guidelines. http://www.calimera.org/Lists/Guidelines%20PDF/Digital_preservation.pdf
- Consultative Committee for Space Data Systems (CCSDS) (2002): *Reference Model for an Open Archival Information System (OAIS)*. Blue Book. Washington DC. Seite 5-1. <http://public.ccsds.org/publications/archive/650x0b1.pdf>
- Van Diessen, Raymond J. und van Rijnsoever, Ben J. (2002): *Managing Media Migration in a Deposit System*. IBM/KB Long-Term Preservation Study Report Series Nr. 5. Amsterdam: IBM Niederlande.
<http://www-5.ibm.com/nl/dias/resource/migration.pdf>