

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

nestor Handbuch

Eine kleine Enzyklopädie
der digitalen Langzeitarchivierung

Version 2.3

Kapitel 8.3

Migration

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 8.3 „Migration“ (Version 2.3): [urn:nbn:de:0008-20100617189](http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:0008-20100617189)
<http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:0008-20100617189>



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

8.3 Migration

Stefan E. Funk

Migration und Emulation

Wenn die Archivierung des Bitstreams sichergestellt ist (siehe Bitstreamerhaltung), kann man beginnen, sich über die Archivierung und vor allem über die Nutzung von digitalen Objekten Gedanken zu machen. Bei nicht digitalen Medien wie Büchern und Mikrofilmen hat man in den letzten Jahrzehnten und Jahrhunderten sehr viel Erfahrung mit deren Erhaltung gesammelt, das heißt, auf physikalischer Ebene konnten und können diese Medien sehr lange verfügbar gehalten werden. Ein Buch braucht als zu erhaltendes Objekt auch nur auf der physischen Ebene betrachtet zu werden, denn zum Benutzen eines Buches reicht es aus, das Buch selbst zu erhalten und so die Lesbarkeit zu gewährleisten.

Zwei Strategien, welche die Lesbarkeit der archivierten digitalen Dokumente über lange Zeit (Long Term) garantieren sollen, sind zum einen die Migration und zum anderen die Emulation. „Long term“ wird vom Consultative Committee for Space Data Systems (CCSDS) definiert als:

„Long Term is long enough to be concerned with the impacts of changing technologies, including support for new media and data formats, or with a changing user community. Long Term may extend indefinitely.“

Die Migration passt die digitalen Objekte selbst einem neuen Umfeld an, die Dokumente werden zum Beispiel von einem veralteten Dateiformat in ein aktuelles konvertiert. Mit der Emulation wird das originäre Umfeld der digitalen Objekte simuliert, das neue Umfeld also an die digitalen Objekte angepasst. Diese Strategien können alternativ genutzt werden; sie sind unabhängig voneinander.

Um ein digitales Dokument archivieren und später wieder darauf zugreifen zu können, sind möglichst umfassende Metadaten nötig, also Daten, die das digitale Objekt möglichst genau beschreiben. Dazu gehören in erster Linie die technischen Metadaten. Für die Migration sind weiterhin die Provenance Metadaten wichtig, die wie erläutert die Herkunft des Objekts beschreiben. Deskriptive Metadaten sind aus technischer Sicht nicht so interessant. Sie werden benötigt, um später einen schnellen und komfortablen Zugriff auf die Objekte zu ermöglichen. Rechtliche Metadaten können schließlich genutzt werden, um Einschränkungen für die Migration, die Emulation und den Zugriff auf die digitalen Objekte festzulegen.

Migration

Mit dem Stichwort Migration werden innerhalb der Langzeitarchivierungs-Community unterschiedliche Prozesse bezeichnet. Dies sind sowohl die Datenträgermigration als auch die Daten- oder Formatmigration.

Bei der Datenträgermigration werden Daten von einem Träger auf einen anderen kopiert, z.B. von Festplatte auf CD, von DVD auf Band etc. Diese Art der Migration ist die Grundlage der physischen Erhaltung der Daten, der Bitstream Preservation.

Bei einer Datenmigration (auch Formatmigration genannt) werden Daten von einem Datenformat in ein aktuelleres, möglichst standardisiertes und offen gelegtes Format überführt. Dies sollte geschehen, wenn die Gefahr besteht, dass archivierte Objekte aufgrund ihres Formates nicht mehr benutzt werden können. Das Objekt selbst wird so verändert, dass seine Inhalte und Konzepte erhalten bleiben, es jedoch auf aktuellen Rechnern angezeigt und benutzt werden kann. Problematisch ist bei einer Datenmigration der möglicherweise damit einhergehende Verlust an Informationen. So ist es zum Beispiel möglich, dass sich das äußere Erscheinungsbild der Daten ändert oder - noch gravierender - Teile der Daten verloren gehen.

Eine verlustfreie Migration ist dann möglich, wenn sowohl das Original-Format wie auch das Ziel-Format eindeutig spezifiziert sind, diese Spezifikationen bekannt sind UND eine Übersetzung von dem einen in das andere Format ohne Probleme möglich ist. Hier gilt: Je einfacher und übersichtlicher die Formate, desto größer ist die Wahrscheinlichkeit einer verlustfreien Migration. Bei der Migration komplexer Datei-Formate ist ein Verlust an Informationen wahrscheinlicher, da der Umfang einer komplexen Migration nicht unbedingt absehbar ist. Eine Migration eines Commodore-64 Computerspiels in ein heute spielbares Format für einen PC ist sicherlich möglich, jedoch ist es (a) sehr aufwändig, (b) schlecht bzw. gar nicht automatisierbar und (c) das Ergebnis (sehr wahrscheinlich) weit vom Original entfernt.

Beispiel: Alte und neue PCs

- Sie haben einen recht alten PC, auf dem Sie seit langem Ihre Texte schreiben, zum Beispiel mit einer älteren Version von Word 95 (Betriebssystem: Windows 95). Sie speichern Ihre Daten auf Diskette.
- Ihr neuer Rechner, den Sie sich angeschafft haben, läuft unter Windows XP mit Word 2003 und hat kein Diskettenlaufwerk mehr.
- Nun stehen Sie zunächst vor dem Problem, wie Sie Ihre Daten auf den neuen Rechner übertragen. Wenn Sie Glück haben, hat Ihr alter Rechner

schon USB, sodass Sie Ihre Daten mit einem USB-Stick übertragen können. Vielleicht haben Sie auch noch ein Diskettenlaufwerk, auf das Sie zurückgreifen können. Oder aber Ihr alter Rechner kann sich ins Internet einwählen und Ihre Daten können von dort mit dem neuen Rechner heruntergeladen werden. Hier ist unter Umständen ein wenig zu tun. Es gibt jedoch noch genügend Möglichkeiten, Ihre Daten zu übertragen.

- Nehmen wir an, Ihre Daten sind sicher und korrekt übertragen worden. Wenn Sie Glück haben, meldet sich Word 2003 und sagt, Ihre Dateien seien in einem alten .doc-Format gespeichert und müssen in das aktuelle Format konvertiert werden. Diese Konvertierung ist dann eine Migration in ein neues, aktuelleres .doc-Format. Wenn die Migration erfolgreich abläuft, sieht Ihr Dokument aus wie auf dem alten Rechner unter Word 95. Es besteht jedoch die Möglichkeit, dass Ihr Dokument sich verändert hat (Formatierung, Schriftart, Schriftgröße etc.).
- Sollten Sie Pech haben, erkennt Word das alte Format nicht und eine Migration ist nicht automatisch möglich. Dann bleibt noch die Möglichkeit, die alten Dateien mit einem Zwischenschritt über ein anderes Textformat, das beide Textprogramme beherrschen, zu konvertieren. Sicherlich können beide Programme einfache Textdateien verarbeiten (.txt), vielleicht auch Dateien im Rich-Text-Format (.rtf). Sie müssen nun Ihre Dokumente mit dem alten Word alle als Text- oder RTF-Datei neu speichern, diese erneut (wie oben beschrieben) auf den neuen Rechner übertragen und dann mit dem neuen Word (als Text- oder RTF-Datei) wieder öffnen. Sehr wahrscheinlich sind dann sehr viele Formatierungen (Inhaltsverzeichnisse, Überschriften, Schriftdicken, Schriftarten, etc.) verloren gegangen, da eine .txt-Datei keinerlei solcher Dinge speichern kann. Nur der Text entspricht dem originalen Dokument. Mit einer RTF-Datei haben Sie sicherlich weniger Informationsverlust. Sie führen also praktisch zwei Migrationen durch: .doc (Word 95) – .txt (bzw. .rtf) – .doc (Word 2003), siehe hierzu die Abbildungen 1 und 2.



Abbildung 1: Ein Word-Dokument mit Grafiken, Formatierungen, Link, etc.

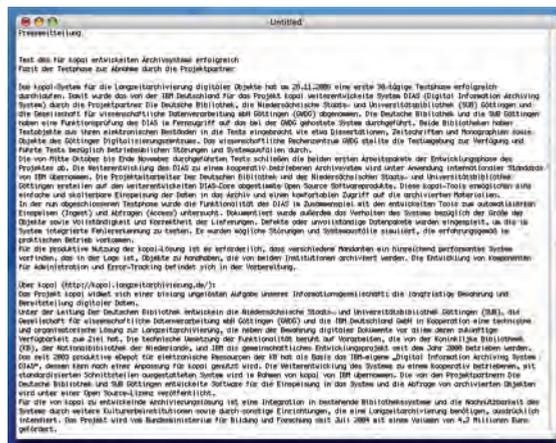


Abbildung 2: Das selbe Dokument im .txt-Format ohne Formatierungen

Beispiel: Zeichenkodierungen

- Eine Organisation, die in den 80er Jahren ihre Daten mit IBM Mainframes bearbeitet hat, möchte diese Daten auch auf späteren Systemen nutzen können. Die IBM Mainframes nutzten einen Zeichenstandard namens EBCDIC.⁷
- In den 90er Jahren installierte Rechner nutzten den ASCII Zeichencode (American National Standard Code for Information Interchange), welcher nicht alle Zeichen des EBCDIC abdeckte. Die Organisation musste sich nun entscheiden, ob sie alle Dokumente nach ASCII konvertierten (und einen permanenten Verlust von Daten hinnahmen) oder sie nur bei Bedarf in ASCII umwandelten und die Originaldaten in EBCDIC beließen. So hatte man den gleichen Verlust beim Umwandeln, jedoch für spätere Zeit die Originaldaten erhalten.
- Zum Jahrtausendwechsel begann UNICODE⁸ die Welt zu erobern und tatsächlich enthält UNICODE alle Zeichen des EBCDIC, sodass nun alle Dokumente 1:1 von EBCDIC in UNICODE konvertiert werden konnten (sofern die Originaldateien noch existierten!). Bei einer sofortigen Konvertierung in ASCII wären tatsächlich Daten verloren gegangen.

Zusammenfassung: Vor- und Nachteile von Migration*Vorteile von Migration*

- Migration ist technisch (verglichen mit Emulation) gut zu realisieren.
- Migration kann in vielen Fällen automatisiert werden.
- Die migrierten Dokumente sind unabhängig von weiteren Komponenten (abgesehen von der aktuellen Darstellungssoftware).
- Die originalen Objekte können aufbewahrt werden, um evtl. später darauf zurückgreifen zu können.

Nachteile von Migration

- Jedes Objekt muss einzeln migriert werden.
- Die Wahrscheinlichkeit von Datenverlust bzw. Datenveränderung ist (besonders über mehrere Migrationsschritte) sehr hoch.
- Jede Version (Migration) eines Objekts inklusive des Original-Dokuments sollte gespeichert werden. Damit ist unter Umständen ein hoher Speicherplatzbedarf verbunden.

7 Extended Binary Coded Decimal Interchange Code: <http://www.natural-innovations.com/computing/asciiebcdic.html>

8 <http://www.unicode.org>

- Für jedes Format und für jeden Migrations-Schritt muss es ein Migrations-Werkzeug geben.
- Migration ist nicht für alle Formate realisierbar.

Literatur

Consultative Committee for Space Data Systems (2001): *Reference Model for an Open Archival Information System (OAIS)*, CCSDS 650.0-B-1, BLUE BOOK, <http://public.ccsds.org/publications/archive/650x0b1.pdf>

Jenkins, Clare (2002): *Cedars Guide to Digital Preservation Strategies*, <http://www.webarchive.org.uk/wayback/archive/20050409230000/http://www.leeds.ac.uk/cedars/guideto/dpstrategies/dpstrategies.html>