

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

nestor Handbuch

Eine kleine Enzyklopädie
der digitalen Langzeitarchivierung

Version 2.3

Kapitel 7.3
Auswahlkriterien

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 7.3 „Auswahlkriterien“ (Version 2.3): [urn:nbn:de:0008-20100617148](http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:0008-20100617148)
<http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:0008-20100617148>



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

7.3 Auswahlkriterien

Jens Ludwig

Formate sind in unterschiedlichem Maße dem Risiko zu veralten ausgesetzt. Daher ist es naheliegend die langfristige Nutzbarkeit der digitalen Objekte durch die Verwendung eines geeigneten Formates zu unterstützen. Bevor man aber versucht zu beantworten, welches Format theoretisch am besten für die Langzeitarchivierung geeignet ist, muss man sich klarmachen, was die begrenzenden Faktoren der Formatwahl sind.

Die wichtigste und in gewissem Sinne triviale Einschränkung der Formatwahl ist, dass ein Format auch die benötigte Funktionalität aufweisen muss. Es gibt Formate mit identischen Funktionen, die leicht durcheinander ersetzt werden können, aber genauso Formate für Spezialzwecke, die man dann leider nicht mit für die Langzeitarchivierung besser geeigneten austauschen kann, weil diese Spezialfunktionen eben benötigt werden. Um ein Format auswählen zu können, muss man sich also bewusst sein, was für Funktionalitäten benötigt werden.

In diesem Zusammenhang gilt es auch die Position des „Langzeitarchiviers“ in der Verarbeitungskette zu berücksichtigen: Muss schon bei der Bearbeitung und Erstellung des digitalen Objekts das richtige Format ausgewählt werden, weil z.B. ein Dokument genauso wiederverwendet und bearbeitet werden soll? Dann muss man selbst der Ersteller sein oder hinreichenden Einfluss auf die Erstellung haben, sonst muss man hinnehmen, was man bekommt. Oder reicht ggf. eine statische Version, die nur den visuellen Eindruck erhält, und es ist deshalb möglich, das Objekt in ein neues, selbst ausgewähltes Format zu überführen?

Und selbst wenn die digitalen Objekte in den nach bestem Wissen und Gewissen ausgesuchten Formaten vorliegen, heißt das nicht, dass das Problem gelöst ist. Quasi jedes Format kann veralten, auch wenn es sich einmal als die beste Wahl dargestellt hat, Anforderungen können sich ändern und der technische Fortschritt kann neue Formate ermöglichen und erfordern. Aus all diesen Gründen kann man keine dauerhafte Festlegung auf ein bestimmtes Format treffen.

Kriterien

Trotz dieser Einschränkungen und Absicherungen lassen sich aber eine Reihe von allgemeinen Faktoren aufführen, was für Formate für digitale Objekte sinnvoll sind, die langfristig genutzt werden sollen. Und es gibt eine Vielzahl von Katalogen, die solche Faktoren aufführen: z.B. eher klassische Aufstellungen wie Lormant et al. (2005), Stanescu (2004) oder Arms, Fleischhauer (2005), deren Autoren auch die informative Seite der Library of Congress zum Thema Formate pflegen (Arms, Fleischhauer 2007), aber auch spezialisierte wie Barkstrom, Folk (2002), die besonders Erwägungen für naturwissenschaftliche Forschungsdaten berücksichtigen, oder Christensen et al. (2004), die Kriterien für Kontainerformate zur Internetarchivierung aufstellen. So interessant die unterschiedlichen Perspektiven und Kriterien im Detail sein mögen, auf einer abstrakten Ebenen lassen sich die Kriterien zusammenfassen. Angelehnt an Rog, van Wijk (2008) sind zu nennen:

- **Offenheit:** Ist die Spezifikation des Formates frei verfügbar oder ist sie ein Betriebsgeheimnis eines Herstellers? Ist das Format durch Normungsinstitutionen standardisiert worden? Mit der Spezifikation besteht die Möglichkeit, das Format zu verstehen und ggf. selbst Nutzungssoftware zu entwickeln, auch wenn es keinen Anbieter mehr gibt.
- **Verbreitung:** Wie verbreitet ist das Format? Wie häufig wird es genutzt, wieviel unabhängige Anbieter von Nutzungssoftware gibt es? Eine hohe Verbreitung ist ein Indiz dafür, dass das Format noch lange und von vieler Software unterstützt wird, da ein großer Markt dafür vorhanden ist.
- **Komplexität:** Wie kompliziert ist das Format? Technische Komplexität erschwert die fehlerfreie Entschlüsselung bzw. Nutzung. Je mehr Wissen zum Verständnis eines Formates notwendig ist, desto eher kann ein Teil des notwendigen Wissens verloren gehen.
- **Schutzmechanismen:** Kopierschütze und Verschlüsselungen mögen für bestimmte Anwendungen sinnvoll sein, für die Langzeitarchivierung sind sie es nicht. Die langfristige Erhaltung setzt das Kopieren der digitalen Objekte voraus und eine Verschlüsselung erfordert als Minimum die zusätzliche Kenntnis des Schlüssels und Verschlüsselungsverfahrens.
- **Selbstdokumentation:** Wenn ein Format die Integration von Metadaten ermöglicht, dann erleichtert das voraussichtlich das Verständnis des digitalen Objekts und verringert die Abhängigkeit von externen Metadatenquellen.
- **Robustheit:** Je robuster ein Format ist, desto weniger wirken sich Veränderungen aus. Wie stark wirken sich Fehler einzelner Bits auf die Nutz-

barkeit des gesamten Objekts aus? Gibt es nur einen kleinen, vernachlässigbaren Darstellungsfehler oder lässt es sich ggf. überhaupt nicht mehr nutzen? Wie kompatibel sind verschiedene Versionen bzw. Weiterentwicklungen des Formats untereinander?

- Abhängigkeiten: Formate, die weniger von spezieller Hard- oder Software oder anderen Ressourcen (z.B. Internetzugang) abhängig sind als andere, sind zu bevorzugen.

Wie bereits erwähnt wurde, sind über diese Kriterien hinaus die spezifisch benötigten Funktionalitäten zu erwägen. Diese selbst nur für bestimmte Medientypen auszuführen, würden den Umfang dieses Kapitels sprengen. Gute weiterführende Quellen für bestimmte Medientypen sind neben dem Kapitel „Vorgehensweise für ausgewählte Objekttypen“ dieses Handbuchs auch Arms, Fleischhauer (2007) und AHDS (2006).

Literatur

- AHDS (arts and humanities data service): Preservation Handbooks. 2006. <http://ahds.ac.uk/preservation/ahds-preservation-documents.htm>
- Arms, Caroline/ Fleischhauer, Carl: *Digital Formats: Factors for Sustainability, Functionality, and Quality*. 2005. Paper for IS&T Archiving 2005 Conference, Washington, D.C. http://memory.loc.gov/ammem/techdocs/digform/Formats_IST05_paper.pdf
- Arms, Caroline/ Fleischhauer, Carl (Hrsg.): Sustainability of Digital Formats. Planning for Library of Congress Collections. 2007. <http://www.digitalpreservation.gov/formats/index.shtml>
- Barkstrom, Bruce R./ Folk, Mike: *Attributes of File Formats for Long Term Preservation of Scientific and Engineering Data in Digital Libraries*. 2002. http://www.ncsa.uiuc.edu/NARA/Sci_Formats_and_Archiving.doc
- Christensen, Steen S. et al.: *Archival Data Format Requirements*. 2004. http://netarkivet.dk/publikationer/Archival_format_requirements-2004.pdf
- Lormant, Nicolas et al.: *How to Evaluate the Ability of a File Format to Ensure Long-Term Preservation for Digital Information?* 2005. Paper for PV 2005, The Royal Society, Edinburgh. <http://www.ukoln.ac.uk/events/pv-2005/pv-2005-final-papers/003.pdf>
- Rog, Judith/ van Wijk, Caroline: *Evaluating File Formats for Long-term Preservation*. 2008. http://www.kb.nl/hrd/dd/dd_links_en_publicaties/publicaties/KB_file_format_evaluation_method_27022008.pdf
- Stanescu, Andreas: *Assessing the Durability of Formats in a Digital Preservation Environment*. In: D-Lib Magazine, November 2004, Volume 10 Number 11. doi:10.1045/november2004-stanescu