

H. Neuroth, A. Oßwald, R. Scheffel, S. Strathmann, M. Jehn (Hrsg.)

nestor Handbuch

Eine kleine Enzyklopädie
der digitalen Langzeitarchivierung

Version 2.0

Kapitel 14.2

Kosten

nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung
hg. v. H. Neuroth, A. Oßwald, R. Scheffel, S. Strathmann, M. Jehn
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

Die Herausgeber danken Anke Herr (Korrektur), Martina Kerzel (Bildbearbeitung) und
Jörn Tietgen (Layout und Formatierung des Gesamttextes) für ihre unverzichtbare
Unterstützung bei der Fertigstellung des Handbuchs.

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.

Die Inhalte dieses Buchs stehen auch als Onlineversion
(<http://nestor.sub.uni-goettingen.de/handbuch/>)
sowie über den Göttinger Universitätskatalog (<http://www.sub.uni-goettingen.de>) zur
Verfügung.

Die digitale Version 2.0 steht unter folgender Creative-Commons-Lizenz:
„Attribution-Noncommercial-Share Alike 3.0 Unported“
<http://creativecommons.org/licenses/by-nc-sa/3.0/>



Einfache Nutzungsrechte liegen beim Verlag Werner Hülsbusch, Boizenburg.
© Verlag Werner Hülsbusch, Boizenburg, 2009
www.vwh-verlag.de
In Kooperation mit dem Universitätsverlag Göttingen

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.

Druck und Bindung: Kunsthaus Schwanheide

Printed in Germany – Als Typoskript gedruckt –

ISBN: 978-3-940317-48-3

URL für Kapitel 14.2 „Kosten“ (Version 2.0): [urn:nbn:de:0008-20090811736](http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:0008-20090811736)
<http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:0008-20090811736>



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

14.2 Kosten

Thomas Wollschläger und Frank Dickmann

In diesem Kapitel werden Kostenfaktoren benannt, die für den Betrieb eines digitalen Langzeitarchivs von Bedeutung sind. Des Weiteren werden Ansätze vorgestellt, wie die individuellen Kosten der LZA in einer Institution ermittelt werden können.

Kostenfaktoren bei Einrichtung und Unterhaltung eines Langzeitarchivs

Abhängig vom konkreten Langzeitarchivierungskonzept der jeweiligen Einrichtung werden folgende Kostenfaktoren zu berücksichtigen sein:

Initiale Kosten

- Informationsbeschaffung über LZA-Systeme
- Erhebung von Bestand, Zugang und gewünschten Zugriffsoptionen für digitale Materialien im eigenen Haus
- Erhebung von vorhandenen Personal- und Technikressourcen im eigenen Haus
- Projektplanung, ggf. Consulting, Ausschreibung(en)

Beschaffungskosten

- Hardware: Speichersysteme und *sämtliche* infrastrukturellen Einrichtungskosten (Serveranbindungen, Datenleitungen, Mitarbeiterrechner usw.)
- Ggf. Lizenz(en) für Software-Systeme oder Beitrittskosten zu Konsortien
- Weitere Aufwendungen: z.B. Anpassungsentwicklungen von Open Source Software-Produkten, Entwicklung/Anpassung von Schnittstellen, Erstellung von physischen und digitalen Schutzmaßnahmen (auch solche aus rechtlichen Gründen)
- Ggf. Einstellung neuer Mitarbeiter und/oder Schulung vorhandener Mitarbeiter

Betriebskosten

- Dateningest des bisher vorhandenen Materials
- Dateningest des neu eingehenden Materials
- Laufende Storage-Kosten

- Sonstige Dauerbetriebskosten: z.B. Strom, Datenleitungskosten, *sämtliche* Sicherheitsmaßnahmen, Backups, regelmäßige Wartung(en) und Tests, Software-Upgrades
- Zukauf von weiteren Speichereinheiten
- Hard- und Software-Komplettersatz in Intervallen
- Ggf. laufende Lizenzkosten und/oder laufende Beitragszahlungen bei Konsortien

Die konkreten Kosten sind dabei jeweils abhängig von

- der Zahl und Komplexität der Workflows bei einer Institution
- der Menge, Heterogenität und Komplexität der zu archivierenden Objekte und ihrer Metadaten
- den gewünschten Zugriffsmöglichkeiten und Schnittstellen sowie ggf.
- den Anforderungen Dritter an die archivierende Institution bzw. Verpflichtungen der Institution gegenüber Dritten

Die Ermittlung von Kosten für die Langzeitarchivierung

Die tatsächliche Ermittlung der Kosten, die auf eine Einrichtung für die Langzeitarchivierung ihrer digitalen Dokumente zukommen, gestaltet sich in der Praxis noch relativ schwierig. Viele LZA-Unternehmungen befinden sich derzeit noch im Projektstatus oder haben noch nicht lange mit dem produktiven Betrieb begonnen. Daher liegen noch wenige Erfahrungswerte vor, wie sich insbesondere der laufende Betrieb eines solchen Archivs kostenmäßig erfassen lässt. Außerdem befindet sich nach wie vor die zunehmende Menge und Varianz insbesondere der Internet-Publikationen in einem Wettlauf mit den technischen Möglichkeiten, die von Gedächtnisorganisationen zur Einsammlung und Archivierung eingesetzt werden können.

Einen begrenzten Anhaltspunkt können die angesprochenen Unternehmungen zumindest in der Hinsicht liefern, was die Ersteinrichtung eines digitalen Langzeitarchivs betrifft. BMBF und DFG haben eine ganze Reihe von solchen Projekten gefördert und verschiedene Institutionen haben Projekte aus eigenen Mitteln finanziert.¹ Das bisher am umfangreichsten geförderte LZA-Vorhaben in Deutschland war das Projekt kopal mit einem Fördervolumen von 4,2 Mio. Euro.² Diese Kosten schließen die vollständige Entwicklung eines Ar-

1 Siehe dazu die Projektübersicht in der nestor-Informationsdatenbank http://nestor.sub.uni-goettingen.de/nestor_on/browse.php?zeitg=10

Alle hier aufgeführten URLs wurden im April 2009 auf Erreichbarkeit geprüft .

2 Vgl. Wollschläger (2007), S. 247.

chivsystems einschließlich Objektmodell, Aufbau von Hard- und Softwareumgebungen in mehreren Einrichtungen und mehrjährige Forschungsarbeiten ein. Zum Projektende hat kopal allerdings in einem Servicemodell konkrete Kosten für den Erwerb eines vollständigen Archivs zum Eigenbetrieb vorgelegt. Wenn das kopal-Archivsystem unter Zukauf von Beratung und ggf. Entwicklung eigenständig betrieben wird, soll ein Erstaufwand für Hard- und Software eines Systems mittlerer Größe von ca. 750.000 € anfallen. Hiervon entfielen 40% auf Softwarelizenzen und 60% auf Systembereitstellung und -betrieb.³ Wiewohl solche Angaben nur exemplarisch sein können, kann dennoch davon ausgegangen werden, dass die Kosten für die Ersteinrichtung eines LZA-Systems in einer Einrichtung einen gewissen Schwellenwert nicht unterschreiten werden.

Die Zahl der Ansätze, die bisher versucht haben, Modelle für die Betriebskostenermittlung digitaler LZA zu entwickeln, ist begrenzt. Nennenswert ist hierbei der Ansatz des LIFE-Projekts aus Großbritannien. „The LIFE Project“ war ein einjähriges Projekt (2005/2006) der British Library (BL) in Zusammenarbeit mit dem University College London (UCL) mit dem Hauptziel, ein Kostenmanagement für die Langzeiterhaltung elektronischer Ressourcen zu entwickeln. Es wurde eine Formel zur Ermittlung der Archivierungskosten entwickelt. Manche Fragen mussten noch offen bleiben, so war es z.B. bislang nicht adäquat möglich, im Rahmen des Projektes die Kosten der Langzeiterhaltung von gedruckten und elektronischen Veröffentlichungen zu vergleichen. Die Formel lautet: $L_T = Aq + I_T + M_T + Ac_T + S_T + P_T$. Dabei stehen die Werte für folgende Parameter.⁴

- L = complete lifecycle cost over time 0 to T.
- Aq = Acquisition
- I = Ingest
- M = Metadata
- Ac = Access
- S = Storage
- P = Preservation

Jeder der Parameter kann weiter in praktische Kategorien und Elemente aufgeteilt werden. Alle Schritte können entweder, wenn der Prozess direkt kalkulierbar ist, als Kostenfaktor berechnet werden oder, wenn nötig, jeweils auch noch

3 Siehe kopal (2007), S. 2.

4 Vgl. McLeod/Wheatley/Ayris (2006), S. 6. Das tiefergestellte (T) in der Formel bedeutet als Attribut der Parameter „over time“.

in beliebig viele Unterpunkte untergliedert werden. So kann die Berechnung für die jeweilige Institution individuell angepasst werden. Innerhalb des LIFE-Projekts wurden zum einen beispielhafte Berechnungen der LZA-Kosten des Projektmaterials vorgenommen und dabei Kosten für „the first year of a digital asset’s existence“ und „the tenth year of the same digital assets’ existence“ vergleichbar ermittelt⁵ und exemplarisch auch die Kosten pro Speichermenge. Zum anderen hat das Projekt die entwickelten Formelwerke zur Verfügung gestellt, so dass interessierte Institutionen selbst Berechnungen anhand der Individualparameter vornehmen können.

In Anbetracht der aktuellen Erkenntnisse fokussieren Kostenschätzungen für die LZA hauptsächlich auf die Kosten pro Speichermenge, wie z.B. in LIFE. Bei eingehender Betrachtung der Prozesse, die im Rahmen eines implementierten LZA-Systems anfallen, sind jedoch die Speicherkosten nicht der Hauptkostenfaktor. Vielmehr lässt sich aus den Prozessen Acquisition, Metadata und Preservation eine hohe Personalintensität ableiten, insbesondere aus dem Grund, da diese Prozesse nur sehr eingeschränkt automatisierbar sind. Daher sind es vielmehr die Personalkosten, die langfristig den höchsten Anteil an den LZA-Kosten haben werden.⁶

Ebenso von Bedeutung ist die Anzahl unterstützter Formate, da diese Anzahl und der Personaleinsatz eng miteinander verknüpft sind. Jedes zusätzliche Format erfordert zusätzlichen Aufwand durch qualifiziertes Personal. Demgegenüber steht aber der Nutzen, den das Angebot eines Formats liefert. Hierzu wurde festgestellt, dass weniger häufig verwendete Formate ein unterdurchschnittliches Kosten-Nutzen-Verhältnis aufweisen. Beispielsweise haben in LIFE die Formate PDF, TXT und HTML ca. 85% aller Dokumente abgedeckt, allerdings nur 7% der Kosten verantwortet, während die 12 am wenigsten verwendeten Formate 0,1% der Dateien abgedeckt, aber ca. 41% der Kosten hervorgerufen haben. Grundlage dabei sind die Gesamtkosten über einen Zeitraum von 20 Jahren.⁷ Aus diesem Grund ist eine Einschränkung der Formatvielfalt ein empfehlenswertes Kostensteuerungsinstrument für die LZA.

Eine bedeutende Frage für die Festlegung der Archivierungsstrategie – nämlich für das eigentliche „Preservation Planning“, die Erhaltungsmaßnahmen über die Lebenszeit eines digitalen Objekts – einer Institution ist, ob auf Dauer Migrationen oder Emulationen kostengünstiger sind. Hierzu sind noch keine abschließenden Aussagen möglich. Generell verbreitet ist die Auffassung, dass Migration der kostengünstigere Weg sei. Innerhalb von LIFE wurden dazu An-

5 Vgl. ebenda, S. 3.

6 Vgl. Ashley (1999), S. 123.

7 Vgl. Björk, B.-C. (2007), S. 23.

sätze formuliert, die jedoch hauptsächlich sehr exemplarische Migrationen behandeln und noch nicht repräsentativ sind.⁸ Andere Studien kommen dagegen zu dem Schluss, dass Emulationen auf längere Sicht kostengünstiger seien:

While migration applies to all objects in the collection repetitively, emulation applies to the entire collection as a whole. This makes emulation most cost-effective in cases of large collections, despite the relatively high initial costs for developing an emulation device. When considering the fact that only small fragments of digital archives need to be rendered in the long run, it may turn out that from a financial perspective emulation techniques will be more appropriate for maintaining larger archives.⁹

Da die bestehenden Langzeitarchive gerade erst dabei sind, die ersten „echten“ Maßnahmen von Preservation Planning umzusetzen, wird hier auf Erfahrungswerte zu warten sein, die entsprechende Ergebnisse unterstützen können.

Konsequenzen für die Gedächtnisorganisationen

Angesichts der zu erwartenden nicht unerheblichen Kosten für die *Ersteinrichtung* eines LZA-Systems dürften kleinere Einrichtungen nicht umhin kommen, zwecks Einrichtung eines solchen Systems mit anderen Institutionen zu kooperieren bzw. sich einem bestehenden System anzuschließen und/oder sich den Zugang dazu über Lizenzen zu sichern. Selbst größere Institutionen werden für die Einrichtung eines LZA-Systems oft kooperative Formen wählen, um hohe Ersteinrichtungskosten aufzuteilen, die sich sonst nicht auf mehrere Schultern verteilen lassen. Ebenso könnte angesichts der noch bestehenden Unsicherheit, wie sich künftig die Kosten für den Dauerbetrieb des Langzeitarchivs und das Preservation Planning entwickeln werden, die Entscheidung zugunsten der Variante ausfallen, sich in bestehende Systeme einzukaufen oder über kostenpflichtige Lizenzen Teilnehmer an einem kommerziell ausgerichteten System zu werden. Letzteres macht in der Regel Zugeständnisse an die gewünschte Preservation Policy notwendig, so dass eine Gedächtnisorganisation abwägen muss, welche Kosten – Lizenzen für ein kommerzielles System oder eigene Entwicklungskosten, z.B. für die Anpassung von Open Source Software – die jeweils lohnendere Investition ist.

Die Teilnahme an kooperativen Formen der Langzeitarchivierung ist unter Kostenaspekten in jedem Fall empfehlenswert. Hierbei können Institutionen über z.B. gemeinsame Speichernutzung bzw. gegenseitiges Backup, gegenseitige Nutzung von Entwicklungsergebnissen, gemeinsame Adressierung

8 Vgl. Ebenda, S. 10.

9 Zitiert nach Oltmans/Kol (2005), #5 – Conclusion.

übergreifender Herausforderungen oder kooperative Verwaltung von Open Source Software Synergien schaffen und erhebliche Ressourceneinsparungen ermöglichen.

Quellen und Literatur

- Ashley, K. (1999): *Digital Archive Costs: Facts and Fallacies*, in: *Proceedings of the DLM-Forum on Electronic Records (DLM '99)*, DLM-Forum, Brussels, 1999, S. 123, http://ec.europa.eu/archives/ISPO/dlm/fulltext/full_ashl_en.htm
- Björk, B.-C. (2007): *Economic evaluation of LIFE methodology*, LIFE Project, London, UK, 2007, URL: <http://eprints.ucl.ac.uk/7684/1/7684.pdf>
- Kopal (2007): *kopal: Ein Service für die Langzeitarchivierung digitaler Informationen*. Projekt kopal (Kooperativer Aufbau eines Langzeitarchivs digitaler Informationen), 2007, http://kopal.langzeitarchivierung.de/downloads/kopal_Services_2007.pdf
- McLeod, Rory; Wheatley, Paul; Ayris, Paul (2006): *Lifecycle Information for E-literature* : A summary from the LIFE project ; Report Produced for the LIFE conference 20 April 2006. LIFE Project, London (via <http://www.ucl.ac.uk/life/lifeproject/> or directly under <http://eprints.ucl.ac.uk/archive/00001855/01/LifeProjSummary.pdf>)
- Oltmans, Erik; Kol, Nanda (2005): *A Comparison Between Migration and Emulation in Terms of Costs*. In: RLG DigiNews, Volume 9, Number 2, 15.04.2005 (<http://worldcat.org/arcviewer/1/OCC/2007/07/10/0000068902/viewer/file1.html#article0>)
- Wollschläger, Thomas (2007): *kopal – ein digitales Archiv zur dauerhaften Erhaltung unserer kulturellen Überlieferung*. In: *Geschichte im Netz : Praxis, Chancen, Visionen ; Beiträge der Tagung .hist2006*, Berlin: Clio-online und Humboldt-Universität zu Berlin, 2007, S. 244 – 257 (Historisches Forum 10 (2007), Teilband I).
- Siehe außerdem die Einträge in der nestor-Informationsdatenbank zum Thema „Kosten“ unter http://nestor.sub.uni-goettingen.de/nestor_on/browse.php?show=8