

nestor Handbuch:
**Eine kleine Enzyklopädie
der digitalen Langzeitarchivierung**

4.4 Auswahlkriterien

Herausgeber

Heike Neuroth
Hans Liegmann †
Achim Oßwald
Regine Scheffel
Mathias Jehn
Stefan Strathmann

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Im Auftrag von

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

Tel. +49 (0) 55 1 39 38 66

Der Inhalt steht unter folgender Creative Commons Lizenz:
<http://creativecommons.org/licenses/by-nc-sa/2.0/de/>



4.4 Auswahlkriterien

Andrea Hänger, Karsten Huth und Heidrun Wiesenmüller

Allgemeines

Die Auswahl digitaler Objekte geschieht auf der Basis von definierten und auf die jeweilige Institution zugeschnittenen Kriterien – beispielsweise in Form von Sammelrichtlinien, Selektions- und Bewertungskriterien oder Kriterien für die Überlieferungsbildung. Im Bibliotheks- und Museumsbereich spricht man i.d.R. von Sammlungen, die aus den Sammelaktivitäten hervorgehen, im Archivbereich dagegen von Beständen, die das Resultat archivischer Bewertung darstellen. Der Begriff der Sammlung wird nur im Bereich des nicht-staatlichen Archivguts verwendet.

Bei digitalen Langzeitarchiven, die von öffentlichen Institutionen betrieben werden, sind die Auswahlkriterien i.d.R. aus dem Gesamtauftrag der Institution abzuleiten. In einigen Fällen gibt es auch gesetzliche Grundlagen – z.B. in den Archivgesetzen, die u.a. auch die formalen Zuständigkeiten staatlicher Archive regeln, oder den nationalen und regionalen Pflichtexemplargesetzen, welche Ablieferungspflichten an bestimmte Bibliotheken festlegen.

Festgelegte, dokumentierte und offen gelegte Auswahlkriterien sind in mehrfacher Hinsicht von zentraler Bedeutung für digitale Langzeitarchive: Als praktische Arbeitsanweisung für das eigene Personal unterstützen sie einen stringenten, von individuellen Vorlieben oder Abneigungen unabhängigen Aufbau der digitalen Sammlung bzw. der digitalen Bestände. Den Nutzern, aber auch den Produzenten bzw. Lieferanten der digitalen Objekte und der allgemeinen Öffentlichkeit machen sie das Profil der digitalen Sammlung bzw. der digitalen Bestände deutlich. Anhand der veröffentlichten Auswahlkriterien können beispielsweise Nutzer entscheiden, ob ein bestimmtes digitales Langzeitarchiv für ihre Zwecke die richtige Anlaufstelle ist oder nicht. Dasselbe gilt für Produzenten digitaler Objekte, soweit es keine gesetzlichen Ablieferungs- oder Anbietungspflichten gibt. Das Vorhandensein von Auswahlkriterien stellt deshalb auch einen wichtigen Aspekt von Vertrauenswürdigkeit dar.¹⁰ Gegenüber den Trägern wird anhand der Auswahlkriterien belegt, dass die Sammelaktivitäten dem Auftrag der Institution entsprechen. Und schließlich spielen die jeweiligen

10 Das Kriterium 1.1 im ‘Kriterienkatalog Vertrauenswürdige Archive’ lautet: „Das digitale Langzeitarchiv hat Kriterien für die Auswahl seiner digitalen Objekte entwickelt“. Vgl. nestor-Arbeitsgruppe Vertrauenswürdige Archive – Zertifizierung (2006), S. 7. Zur Vertrauenswürdigkeit digitaler Langzeitarchive allgemein s.u. Kap. 8.

Auswahlkriterien auch eine wichtige Rolle beim Aufbau von Netzwerken zur verteilten, kooperativen Langzeitarchivierung (beispielsweise im nationalen Rahmen).

Zumeist stellt die Aufnahme digitaler Objekte in die Sammlung bzw. die Bestände eine zusätzliche Aufgabe dar, die zu bestehenden Sammelaktivitäten bzw. Bewertungen für konventionelle Objekte hinzukommt. Viele Institutionen besitzen deshalb bereits Auswahlkriterien im analogen Bereich, die als Ausgangspunkt für entsprechende Richtlinien im digitalen Bereich dienen können. Mit Blick auf die Besonderheiten digitaler Objekte müssen diese freilich kritisch überprüft, abgeändert und erweitert werden. Dabei sind fünf Aspekte besonders zu beachten:

- *Spezielle Objekt- und Dokumenttypen:* Während sich für viele Arten von digitalen Objekten eine Entsprechung im konventionellen Bereich finden lässt, gibt es auch spezielle digitale Objekt- und Dokumenttypen, die in den Auswahlrichtlinien zu berücksichtigen sind. Beispielsweise besitzt eine E-Dissertation im PDF-Format ein analoges Pendant in der konventionellen, gedruckten Dissertation. Eine Entsprechung für originär digitale Objekte wie Websites oder Datenbanken lässt sich hingegen nicht in gleicher Weise finden. Deshalb ist eine Orientierung an vorhandenen konventionellen Auswahlkriterien hier nur bedingt möglich (nämlich nur für die inhaltlich-fachlichen Aspekte des Objektes).
- *Technische Anforderungen:* Anders als bei konventionellen Objekten spielen technische Anforderungen (z.B. das Dateiformat und die notwendige technische Umgebung zur Darstellung der Information) für die Abläufe im digitalen Langzeitarchiv eine wichtige Rolle. Sie sind deshalb in die Überlegungen mit einzubeziehen.
- *Veränderte Arbeitsabläufe:* Digitale Objekte sind unbeständiger als ihre papierenen Gegenstücke und weniger geduldig; sollen sie dauerhaft bewahrt werden, muss bereits bei ihrer Entstehung dafür gesorgt werden. Beispielsweise müssen Bibliotheken auf die Produzenten einwirken, damit diese ihre Publikationen in langzeitgeeigneter Form erstellen; ebenso müssen Archive bei den von ihnen zu betreuenden Behörden bereits bei der Einführung elektronischer Systeme präsent sein. Sollen Informationen aus Datenbanken oder Geoinformationssystemen archiviert werden, muss sichergestellt werden, dass vorhandene Daten bei Änderung nicht einfach überschrieben werden, sondern dass so genannte Historisierungen vorgenommen werden, die einen bestimmten Stand festhal-

ten.

- *Unterschiedliche Mengengerüste:* Die Zahl und der Umfang der theoretisch auswahlfähigen digitalen Objekte liegt häufig in deutlich höheren Größenordnungen als bei entsprechenden analogen Objekten. Beispielsweise sind Netzpublikationen sehr viel leichter zu realisieren als entsprechende Printpublikationen, so dass ihre Zahl die der gedruckten Publikationen bei weitem übersteigt. Ebenso werden zum Beispiel Statistikdaten in der Papierwelt nur in aggregierter, d.h. zusammengefasster Form als Quartals- oder Jahresberichte übernommen. In digitaler Form können jedoch auch die Einzeldaten übernommen und den Nutzern in auswertbarer Form zur Verfügung gestellt werden.
- *Schwer zu bemessender Arbeitsaufwand:* Der Umgang mit konventionellen Objekten erfolgt über etablierte Kanäle und Geschäftsgänge, so dass Aufwände gut zu messen und zu bewerten sind. Der Aufwand zur Beschaffung, Erschließung, Bereitstellung und Langzeitarchivierung digitaler Objekte ist dagegen wegen fehlender Erfahrungswerte schwer abzuschätzen.

Die letzten beiden Punkte können u.U. dazu führen, dass Auswahlkriterien für digitale Objekte strenger gefasst werden müssen als für konventionelle Objekte, sofern nicht auf anderen Wegen – beispielsweise durch den Einsatz maschineller Methoden oder zusätzliches Personal – für Entlastung gesorgt werden kann. Die zusätzliche Berücksichtigung digitaler Objekte bei den Sammelaktivitäten bzw. bei der Bewertung kann außerdem Rückwirkungen auf die Auswahlkriterien für konventionelle Objekte derselben Institution haben, indem etwa die beiden Segmente in ihrer Bedeutung für die Institution neu gegeneinander austariert werden müssen.

Die zu erarbeitenden Auswahlkriterien¹¹ können sowohl inhaltlich-fachlicher als auch formal-technischer Natur sein. Darüber hinaus können beispielsweise auch finanzielle sowie lizenz- und urheberrechtliche Aspekte in die Auswahlkriterien mit eingehen; die folgende Liste erhebt keinen Anspruch auf Vollständigkeit.

Inhaltlich-fachliche Auswahlkriterien

Aus inhaltlich-fachlicher Sicht kommen typischerweise drei Kriterien in Betracht:

- *Verwaltungstechnische, institutionelle oder räumliche Zuständigkeit*, z.B. eines Unternehmensarchivs für die Unterlagen des Unternehmens; eines Museums für Digitalisate eigener Bestände; des Dokumentenservers einer

¹¹ Vgl. zum Folgenden auch die Ergebnisse einer Umfrage zu den in verschiedenen Institutionen angewendeten Auswahlkriterien, die im Rahmen der ersten Phase des nestor-Projektes durchgeführt wurde: Blochmann (2005), S. 9-31.

Universität für die dort entstandenen Hochschulschriften; einer Pflicht-exemplarbibliothek für die im zugeordneten geographischen Raum veröffentlichten Publikationen.

Leitfrage: Ist mein Archiv gemäß der institutionellen oder rechtlichen Vorgaben zur Übernahme des Objekts verpflichtet?

- *Inhaltliche Relevanz*, ggf. in Verbindung mit einer Qualitätsbeurteilung, z.B. thematisch in ein an einer Bibliothek gepflegtes Sondersammelgebiet fallend; zu einer Spezielsammlung an einem Museum passend; von historischem Wert für die zukünftige Forschung; von Bedeutung für die retrospektive Verwaltungskontrolle und für die Rechtssicherung der Bürger. Dazu gehört auch der Nachweis der Herkunft des Objekts aus seriöser und vertrauenswürdiger Quelle. Ggf. können weitere qualitative Kriterien angelegt werden, z.B. bei Prüfungsarbeiten die Empfehlung eines Hochschullehrers.

Leitfragen: Ist das Objekt durch sein enthaltenes Wissen bzw. seine Ästhetik, Aussagekraft o.ä. wichtig für meine Institution? Kann das Objekt bei der Beantwortung von Fragen hilfreich sein, die an meine Institution gestellt werden? Ist das Objekt aufgrund seiner Herkunft, seiner Provenienz von bleibendem (z.B. historischem) Wert?

- *Dokumentart*, z.B. spezifische Festlegungen für Akten, Seminararbeiten, Geschäftsberichte, Datenbanken, Websites etc.

Leitfragen: Besitzt mein Archiv schon Bestände der gleichen Dokumentart? Verfüge ich über das nötige Fachwissen und die nötigen Arbeitsmittel zur Erschließung und Verzeichnung der Dokumentart?

Formal-technische Auswahlkriterien

Aus formal-technischer Sicht steht auf der obersten Ebene das folgende Kriterium:

- *Lesbarkeit des Objekts im Archiv*, z.B. die Prüfung, ob ein Objekt mit den verfügbaren technischen Mitteln (Hardware/Software) des Langzeitarchivs dargestellt werden kann. Darstellen heißt, dass die vom Objekt transportierte Information vom menschlichen Auge erkannt, gelesen und interpretiert werden kann.

Leitfrage: Verfügt mein Archiv über die nötigen Kenntnisse, Geräte und Software, um das Objekt den Nutzern authentisch präsentieren zu können?

Aus diesem obersten formal-technischen Zielkriterium lassen sich weitere Unterkriterien ableiten:

- *Vorhandensein der notwendigen Hardware*, z.B. die Feststellung, ob ein einzelner Rechner oder ein ganzes Netzwerk benötigt wird; ob die Nutzung

des Objekts an ein ganz spezielles Gerät gebunden ist usw. Außerdem muss geprüft werden, ob das Objekt mit den vorhandenen Geräten gespeichert und gelagert werden kann.

Leitfragen: Verfügt mein Archiv über ein Gerät, mit dem ich das Objekt in authentischer Form darstellen und nutzen kann? Verfügt mein Archiv über Geräte, die das Objekt in geeigneter Form speichern können?

- *Vorhandensein der notwendigen Software*, z.B. die Feststellung, ob die Nutzung eines Objekts von einem bestimmten Betriebssystem, einem bestimmten Anzeigeprogramm oder sonstigen Einstellungen abhängig ist. Außerdem muss das Archiv über Software verfügen, die das Speichern und Auffinden des Objektes steuert und unterstützt.

Leitfragen: Verfügt mein Archiv über alle Programme, mit denen ich das Objekt in authentischer Form darstellen und nutzen kann? Verfügt mein Archiv über Programme, die das Objekt in geeigneter Form speichern und wiederfinden können?

- *Vorliegen in geeigneten Formaten*, bevorzugt solchen, die normiert und standardisiert sind, und deren technische Spezifikationen veröffentlicht sind. Dateiformate sollten nicht von einem einzigen bestimmten Programm abhängig, sondern idealerweise weltweit verbreitet sein und von vielen genutzt werden. Je weniger Formate in einem Archiv zulässig sind, desto leichter kann auch das Vorhandensein der notwendigen Hard- und Software geprüft werden.

Leitfragen: Hat mein Archiv schon Objekte dieses Formats im Bestand? Sind die notwendigen Mittel und Kenntnisse zur Nutzung und Speicherung des Formats offen zugänglich und leicht verfügbar?

- *Vorhandensein geeigneten Personals*, z.B. die Feststellung, ob die Mitarbeiterinnen und Mitarbeiter über das technische Fachwissen verfügen, das zur Nutzung und Speicherung des Objekts notwendig ist.

Leitfragen: Habe ich Personal, dem ich aus technischer Sicht die Verantwortung für das Objekt anvertrauen kann? Verfüge ich über die Mittel, um Personal mit den entsprechenden Kenntnissen einzustellen oder um Dienstleister mit der Aufgabe zu betrauen?

Auswahlkriterien für Netzpublikationen

Eine für Bibliotheken besonders wichtige Gattung digitaler Objekte sind die so genannten *Netzpublikationen*, auch als „Medienwerke in unkörperlicher Form“ bezeichnet und als „Darstellungen in öffentlichen Netzen“¹² definiert. Auch für diese gelten die oben dargestellten allgemeinen Auswahlkriterien, doch sollen im Folgenden noch einige spezielle Hinweise aus bibliothekarischer Sicht gegeben werden¹³. Dabei ist es nützlich, die Vielfalt der Netzpublikationen in zwei Basistypen zu unterteilen: In die Netzpublikationen mit Entsprechung in der Printwelt einerseits und die sog. Web-spezifischen Netzpublikationen andererseits.¹⁴

Bei den *Netzpublikationen mit Entsprechung in der Printwelt* lassen sich wiederum zwei Typen unterscheiden:

- *Druckbildähnliche Netzpublikationen*, welche ein weitgehend genaues elektronisches Abbild einer gedruckten Publikation darstellen, d.h. 'look and feel' des gedruckten Vorbilds möglichst exakt nachahmen wollen und diesem bis hin zum äußeren Erscheinungsbild entsprechen (z.B. Titelblatt, festes Layout mit definierten Schriftarten und -größen, feste Zeilen- und Seitenumbrüche etc.).
- *Netzpublikationen mit verwandtem Publikationstyp in der Printwelt*, welche zwar keine Druckbildähnlichkeit aufweisen, jedoch einem aus der Printwelt bekannten Publikationstyp zugeordnet werden können, z.B. ein Lexikon im HTML-Format.

Bei der Erarbeitung von Auswahlkriterien für diese beiden Typen ist i.d.R. eine Orientierung an bereits vorhandenen Sammelrichtlinien für konventionelle Materialien möglich. Besondere Beachtung verdient dabei der durchaus nicht seltene Fall, dass zur jeweiligen Netzpublikation eine gedruckte Parallelausgabe vorliegt. Unter Abwägung des zusätzlichen Aufwandes einerseits und des möglichen Mehrwerts des digitalen Objekts andererseits ist festzulegen, ob in einem solchen Fall nur die konventionelle oder nur die digitale Version in das Archiv aufgenommen wird, oder ob beide Versionen gesammelt werden.

Zu den *Web-spezifischen Netzpublikationen* zählen beispielsweise Websites oder Blogs. Sie können keinem aus der Printwelt bekannten Publikationstyp zugeordnet werden, so dass eine Orientierung an bestehenden Sammelrichtlinien

12 Gesetz über die Deutsche Nationalbibliothek (2006), § 3, Abs. 3.

13 Auf andere Arten von Gedächtnisorganisationen ist die folgende Darstellung nicht zwingend übertragbar.

14 Für die folgenden Ausführungen vgl. Wiesenmüller et al. (2004), S. 1423-1437. Unbenommen bleibt, dass die im Folgenden genannten Typen von Netzpublikationen auch in Offline-Versionen vorkommen können.

nur sehr bedingt möglich ist. Für diese Publikationstypen müssen daher neue Auswahlkriterien entwickelt werden.¹⁵

Der Umgang mit *Websites* wird dadurch erschwert, dass unterhalb der Website-Ebene häufig weitere Netzpublikationen - mit oder ohne Entsprechung in der Printwelt - liegen, die getrennt gesammelt, erschlossen und bereitgestellt werden können (z.B. ein Mitteilungsblatt auf der Website einer Institution). In den Auswahlkriterien muss also auch festgelegt sein, unter welchen Umständen (nur) die Website als Ganzes gesammelt wird, oder zusätzlich bzw. stattdessen auch darin integrierte Netzpublikationen in das Archiv aufgenommen werden sollen. Bei Websites, die immer wieder ergänzt, aktualisiert oder geändert werden und deshalb in Zeitschnitten zu sammeln sind, muss jeweils auch das Speicherintervall festgelegt werden.

Bei der Erarbeitung von Auswahlkriterien für Websites sollte unterschieden werden zwischen solchen, welche Personen oder Körperschaften (inkl. Gebietskörperschaften, Ausstellungen, Messen etc.) repräsentieren, und solchen, die sich einem bestimmten Thema widmen – wobei freilich auch Mischformen möglich sind.

Bei *repräsentierenden Websites* setzen die Auswahlkriterien in erster Linie beim Urheber an: Ist die repräsentierte Person oder Körperschaft für mein Archiv relevant? Welche Arten von Personen und Körperschaften sollen schwerpunktmäßig gesammelt, welche ausgeschlossen werden? Ein zusätzliches Kriterium können die auf der Website gebotenen Informationen sein, was sich am besten am Vorhandensein und an der Gewichtung typischer Elemente festmachen lässt: Beispielsweise könnten Websites, die umfangreiche Informationen zur repräsentierten Person oder Körperschaft, einen redaktionellen Teil und/oder integrierte Netzpublikationen bieten, mit höherer Priorität gesammelt werden als solche, die im wesentlichen nur Service- und Shop-Angebote beinhalten.

Bei *thematischen Websites* kommt neben der inhaltlichen Relevanz auch die Qualität als Auswahlkriterium in Frage. Zwar kann i.d.R. keine Prüfung auf Richtigkeit oder Vollständigkeit der gebotenen Information geleistet werden, doch

15 Auch Online-Datenbanken sind am ehesten den Web-spezifischen Netzpublikationen zuzuordnen, weil es in der Printwelt keinen Publikationstyp gibt, der in Funktionalität und Zugriffsmöglichkeiten mit ihnen vergleichbar ist. Ein grundsätzlicher Unterschied zu einem gedruckten Medium ist z.B., dass dessen gesamter Inhalt sequentiell gelesen werden kann, während bei einer Datenbank gemäß der jeweiligen Abfrage nur eine Teilmenge des Inhalts in lesbarer Form generiert wird. Was jedoch den in Online-Datenbanken präsentierten *Inhalt* angeht, so kann es natürlich durchaus Entsprechungen zu Produkten aus der Printwelt geben (z.B. sind in vielen Fällen gedruckte Bibliographien durch bibliographische Datenbanken abgelöst worden).

können als Auswahlkriterien u.a. der Umfang, die Professionalität der Darbietung und die Pflege der Website herangezogen werden, außerdem natürlich der Urheber (z.B. Forschungsinstitut vs. Privatperson).

Detaillierte Sammelrichtlinien für Netzpublikationen, die als Anregung dienen können, sind beispielsweise im Rahmen des PANDORA-Projekts von der Australischen Nationalbibliothek erarbeitet und veröffentlicht worden.¹⁶

Quellenangaben

Blochmann, Andrea (2005): Langzeitarchivierung digitaler Ressourcen in Deutschland: Sammelaktivitäten und Auswahlkriterien (nestor – Kompetenznetzwerk Langzeitarchivierung, AP 8.2). Version 1.0. Frankfurt am Main: nestor.

http://www.langzeitarchivierung.de/downloads/nestor_ap82.pdf (08.10.2007).

Gesetz über die Deutsche Nationalbibliothek (2006): vom 22. Juni 2006.

<http://www.d-nb.de/wir/pdf/dnbg.pdf> (08.10.2007).

National Library of Australia (2005): Online Australian publications: selection guidelines for archiving and preservation by the National Library of Australia. Rev. August 2005. Canberra: National Library of Australia.

<http://pandora.nla.gov.au/selectionguidelines.html> (14.10.2007).

nestor-Arbeitsgruppe Vertrauenswürdige Archive – Zertifizierung (Hrsg.) (2006): Kriterienkatalog vertrauenswürdige digitale Langzeiarchive. Version 1 (Entwurf zur öffentlichen Kommentierung). (nestor-Materialien 8). Frankfurt am Main: nestor.

<http://edoc.hu-berlin.de/series/nestor-materialien/2006-8/PDF/8.pdf> (08.10.2007). urn:nbn:de:0008-2006060710.

Arbeitskreis Archivische Bewertung im VdA – Verband deutscher Archivarinnen und Archivare (Hrsg.) (2004): Positionen des Arbeitskreises Archivische Bewertung im VdA – Verband deutscher Archivarinnen und Archivare zur archivischen Überlieferungsbildung: vom 15. Oktober 2004.

http://www.vda.archiv.net/texte/ak_bew_positionen2004.doc (12.10.2007)

Wiesenmüller, Heidrun et al. (2004): Auswahlkriterien für das Sammeln von Netzpublikationen im Rahmen des elektronischen Pflichtexemplars: Empfehlungen der Arbeitsgemeinschaft der Regionalbibliotheken. In: Bibliotheksdienst 11. 2004 (Jg. 38), S. 1423-1444.

http://www.zlb.de/aktivitaeten/bd_neu/heftinhalte/heft9-1204/digitalebib1104.pdf (08.10.2007).

16 Vgl. National Library of Australia (2005).

