



nestor

Auswahl von Systemen
und Systemkomponenten
für die digitale Langzeitarchivierung

nestor checkliste

urn:nbn:de:0008-2011071203



Auswahl von Systemen
und Systemkomponenten
für die digitale
Langzeitarchivierung

nestor checkliste

Herausgegeben von der AG Kooperation und Vernetzung

nestor - Kompetenznetzwerk Langzeitarchivierung und
Langzeitverfügbarkeit Digitaler Ressourcen für Deutschland

nestor - Network of Expertise in Long-Term Storage
of Digital Resources

<http://www.langzeitarchivierung.de>

nestor Kooperationspartner:

- Bayerische Staatsbibliothek
- Deutsche Nationalbibliothek
- FernUniversität Hagen
- Georg-August-Universität Göttingen / Niedersächsische Staats- und
Universitätsbibliothek Göttingen
- Humboldt-Universität zu Berlin
- Landesarchiv Baden-Württemberg
- Stiftung Preußischer Kulturbesitz / SMB - Institut für Museumsforschung
- Bibliothekservice-Zentrum Baden-Württemberg
- Institut für Deutsche Sprache
- Computerspiele Museum Berlin
- Goportis
- PDF/A Competence Center

© 2011

nestor - Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit
Digitaler Ressourcen für Deutschland

Der Inhalt dieser Veröffentlichung darf vervielfältigt und verbreitet werden, sofern der Name des
Rechteinhabers "nestor - Kompetenznetzwerk Langzeitarchivierung" genannt wird. Eine
kommerzielle Nutzung ist nur mit Zustimmung des Rechteinhabers zulässig.

URN: urn:nbn:de:0008-2011071203

[<http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:0008-2011071203>]

nestor-Checkliste: Auswahl von Systemen und Systemkomponenten für die digitale Langzeitarchivierung

Digitale Aufzeichnungen spielen eine zunehmend zentrale Rolle in unserem gesellschaftlichen, politischen und kulturellen Leben. Ob digitalisierte Bücher oder E-Journals, Webseiten, E-Mails oder elektronische Akten: die Bewahrung ist eine gemeinsame Herausforderung für eine Vielzahl von Gedächtnisinstitutionen. Mittlerweile existieren einige technische Systeme, die die digitale Langzeitarchivierung unterstützen. Diese Checkliste, die sich primär an kulturbewahrende Einrichtungen (Bibliotheken, Archive und Museen) richtet, kann anhand konkreter Fragen beim Einstieg in den Auswahlprozess für ein passendes Langzeitarchivierungssystem behilflich sein, will dabei aber natürlich keinesfalls einen Anspruch auf Vollständigkeit oder ein hohes Maß an Detaillierung erheben. Jede Einrichtung hat ganz eigene Anforderungen an ein Archivierungssystem und diese Checkliste soll sie zunächst dabei unterstützen, diese Anforderungen zu strukturieren und zu gewichten. Natürlich können im Einzelfall bestimmte Kriterien auch irrelevant oder vernachlässigbar sein oder zusätzlich andere Aspekte hinzukommen. In der Praxis wird zudem oftmals ein modularer Aufbau eines Archivierungssystems realisiert werden, d.h. ein solches System wird sich aus einer Vielzahl von unterschiedlichen Komponenten mit unterschiedlichen Funktionalitäten zusammensetzen. Die Checkliste versucht auch hier, mögliche Abhängigkeiten zu berücksichtigen. In dieser Liste sind sowohl Fragen eingebunden, die sich der Auftraggeber selbst beantworten muss, ehe er sich nach Anbietern umschaute, als auch Fragen, die der Auftraggeber dem Anbieter bei der Begutachtung des Systems stellen kann.

1	Allgemeine Merkmale/Komponenten	Kommentar
1.1	Welche Teilbereiche des OAIS-Modells (Datenübernahme, Archivspeicherung, Datenmanagement, Administration, Bereitstellung/Zugriff, Planung der Erhaltung) werden vom Langzeitarchivierungssystem bzw. seinen Komponenten in welcher Breite funktional abgedeckt?	
1.2	Welche Systemkomponenten und Leistungen umfasst das jeweilige System des Anbieters (Rechner-Hardware, Speicher-Hardware, Betriebssystem, Speichermanagement, Datenbank, Anwendungssoftware, Wartung, Anwender-Schulung, Administratoren-Schulung, Systemadministration)	
1.3	Welche der genannten Systemkomponenten und Teilbereiche sollen/müssen in lokaler Verantwortung, welche müssen/können von externen Anbietern betrieben bzw. betreut werden?	
1.4	Falls wesentliche Teile extern betrieben werden (insbesondere als "Cloud"): Wo (in welchen Ländern) erfolgt die physische Speicherung? Welche Rechtsnormen für Datenschutz und Zugriff durch Ermittlungsbehörden gelten dort? Wie wird die Vertraulichkeit und Integrität der Datenbestände sichergestellt?	
1.5	Falls wesentliche Teile extern betrieben werden: Verfügt der Anbieter über relevante Zertifikate?	
1.6	Falls wesentliche Teile extern betrieben werden: Welche Netzbandbreite zum Internet muss ich mindestens zur Verfügung stellen?	
1.7	Zu welchen Bedingungen (Lizenzmodell) wird das System bzw. seine Komponenten angeboten? Gibt es möglicherweise Konflikte zwischen Lizenzen für einzelne Komponenten oder existieren für bestimmte Produkte bereits entsprechende Lizenzen?	
1.8	Inwieweit und auf welche Weise können Softwareprodukte oder Services anderer Anbieter integriert werden? Wer übernimmt dafür hinsichtlich Einbindung und Aktualisierung die Verantwortung? Ergeben sich daraus Konsequenzen für den Support?	
1.9	Welcher Entwicklungsaufwand ist für eine lokale Anpassung des Systems bzw. einzelner Komponenten zu leisten?	
1.10	Inwieweit ist das System für die Nutzung durch mehrere Institutionen/Mandanten geeignet?	

2	Dokumentation und Support	Kommentar
2.1	Für welche Bereiche/Komponenten und in welchem Umfang, in welcher Tiefe und Qualität ist das System dokumentiert?	
2.2	Welche Formen des Supports sind für welche Komponenten des Systems verfügbar? Wie sind Erreichbarkeit und Reaktionszeiten? Welches Level an Service benötigt meine Institution?	
2.3	Gibt es eine aktive Anwender-Community auf deren Erfahrungsschatz zurückgegriffen werden kann?	
2.4	Welche technischen Kenntnisse auf Seiten der Betreuer und Anwender des Systems werden vorausgesetzt bzw. benötigt?	

3	Verbreitung, Nachhaltigkeit und Transparenz des Produkts	Kommentar
3.1	In welcher Version liegt das Produkt derzeit vor, wie viele Versionen sind bereits vorausgegangen und liegen Release Notes für die einzelnen Versionen vor?	
3.2	In wie vielen Instanzen läuft es derzeit?	
3.3	Inwiefern ist eine dauerhafte Betreuung/Weiterentwicklung durch den Anbieter bzw. eine Anwender-Community im Sinne der Nachhaltigkeit sichergestellt?	
3.4	Welche weiteren Entwicklungsschritte bzw. neuen Funktionalitäten sind für die nächsten Versionen vorgesehen? Gibt es eine offizielle Roadmap?	
3.5	Wie geht der Anbieter ggf. mit gemeldeten Fehlern um? Inwieweit werden diese transparent dokumentiert? Gibt es ein standardisiertes, transparentes Verfahren für die Fehlermeldung?	
3.6	Verfügt der Anbieter über einen Nachweis seiner erforderlichen Fachkunde, Zuverlässigkeit und Leistungsfähigkeit?	

4	Systemvoraussetzungen	Kommentar
4.1	Welche Betriebssysteme werden unterstützt bzw. gefordert?	
4.2	Welche Datenbanksysteme werden unterstützt bzw. gefordert?	
4.3	Welche Skriptsprachen werden unterstützt bzw. gefordert?	
4.4	Liegt eine Spezifikation von Art und Umfang der benötigten Hardware-Ressourcen vor?	
4.5	Welche Hardwaresysteme zur physischen Archivspeicherung werden unterstützt (z. B. Plattensysteme/Bandspeichersysteme)?	
4.6	Wie ist die Verknüpfung/Steuerung/Anknüpfung des Archivsystems an die Hardwareschicht gestaltet?	
4.7	Welche Bandbreite bzw. Übertragungsgeschwindigkeit wird für den Betrieb des Systems bzw. einzelner Komponenten benötigt?	
4.8	Inwieweit ist eine mehrfache (redundante) Speicherung möglich: a) auf physisch getrennten Anlagen b) auf technisch unterschiedlicher Speicherhardware c) auf räumlich getrennten Speicheranlagen	

5	Installation und Konfigurierbarkeit	Kommentar
5.1	Wie aufwändig und komplex ist die Installation?	
5.2	Wie aufwändig ist es, komplett getrennte Test- bzw. Produktionssysteme aufzusetzen?	
5.3	Existiert eine Dokumentation über die Abhängigkeiten innerhalb des Systems?	
5.4	Wie aufwändig und wie oft sind Updates der Software durchzuführen? In welcher Form gibt es hierfür Unterstützung von Seiten des Anbieters bzw. eines Dienstleisters? Ergeben sich durch Updates von einzelnen Komponenten des Systems Auswirkungen auf das Gesamtsystem?	
5.5	Inwieweit bzw. in welchen Bereichen kann das System an die spezifischen Anforderungen des einzelnen Kunden angepasst werden?	
5.6	Welche Backup- und Rücksetzungsfunktionen gibt es für die Konfiguration?	

6	Systemsicherheit	Kommentar
6.1	Wie ist die Übertragung bzw. der Datentransfer von und zu externen Stellen abgesichert?	
6.2	Wie sind Datentransfers innerhalb der Systems selbst abgesichert?	
6.3	Welche gesetzlichen und/oder institutionellen Vorgaben (z.B. bzgl. Datenschutz, Datensicherheit) müssen vom System bzw. den einzelnen Komponenten abgedeckt sein?	
6.4	Inwieweit ist eine Fernwartung des Systems bzw. einzelner Komponenten durch eine externe Instanz möglich? Welche Auswirkungen ergeben sich dadurch in punkto Systemsicherheit?	
6.5	Sind verschlüsselte Supportanfragen möglich?	
6.6	Werden Sicherheits-Updates von Dritt-Anbietern zeitnah eingespielt (z.B. Java-Patches)?	

7	Interoperabilität/Schnittstellen zu anderen internen oder externen Systemen	Kommentar
7.1	Existieren geeignete Mechanismen für die Einbindung externer Software bzw. Tools (z. B. für Dateiformatmigration) oder eigener Entwicklungen (z.B. Skripte zur Validierung) und welche Folgen hat dies ggf. für den Support?	
7.2	Besteht auf Anwenderseite die Möglichkeit, die Software selbst weiter zu entwickeln und welche Folgen hat dies ggf. für den Support?	
7.3	Inwieweit bestehen bereits Schnittstellen zu anderen Systemen (z.B. Bibliotheks- bzw. Katalogsystemen, archivarischen Verzeichnungsdatenbanken)?	
7.4	Wie wird der Austausch (Import und Export) von Daten und Metadaten mit anderen Archivsystemen und/oder Suchmaschinenanwendungen gewährleistet (z.B. durch Unterstützung für OAI-PMH/OAI-ORE-Publishing, SOAP)?	
7.5	Welche Web User Interfaces sind standardmäßig in die Software integriert?	
7.6	Wie lassen sich vorhandene User Interfaces anbinden?	

8	Skalierbarkeit	Kommentar
8.1	Welche Dateizahlen und Datenmengen können beim Ingest innerhalb eines bestimmten Zeitraums in das System geladen werden?	
8.2	Welche Dateizahlen und Datenmengen können innerhalb eines bestimmten Zeitraums bereitgestellt werden?	
8.3	Was ist die maximal zu verarbeitende Anzahl an Dateien, was die maximale Größe für eine einzelne Datei?	

9	Systemperformance	Kommentar
9.1	Welche Zuverlässigkeit/Verfügbarkeit des Systems wird benötigt? Wie kann diese sichergestellt werden?	
9.2	Bis zu welcher Zahl von gleichzeitigen Nutzern kann man mit dem System sinnvoll arbeiten?	
9.3	Welche Zahl an gleichzeitig laufenden Prozessen kann das System bearbeiten?	
9.4	Wie sind die Antwortzeiten bei einer hohen Zahl an gleichzeitigen Nutzern bzw. zeitgleich laufenden Prozessen?	

10	Unterstützte Objekttypen	Kommentar
10.1	Welche Objekttypen (z.B. Images, Video, Audio, Forschungsdaten, E-Learning, archivierte Websites) werden vom System unterstützt?	
10.2	Welche Dateiformate werden vom System standardmäßig unterstützt?	
10.3	Welche Dateiformate werden auf Systemen, die im Echtbetrieb sind, bereits verarbeitet?	

11	Datenmodell	Kommentar
11.1	Wird vom System die Langzeitarchivierung, Verwaltung und Bereitstellung aus mehreren unterschiedlichen Dateien bestehender komplexer Objekte (z.B. Zeitschriften, digitale Akten, A/V-Materialien) unterstützt?	
11.2	Wie wird die Bildung von Kollektionen innerhalb physisch oder virtuell getrennter Speicherbereiche (z. B. für Objekte mit besonderen Datenschutzbestimmungen oder einer anderen Institution) unterstützt?	
11.3	Inwieweit ist eine Versionskontrolle bzw. ein Versionenmanagement der archivierten digitalen Objekte vorgesehen?	
11.4	Inwieweit können verschiedene Repräsentationen einer intellektuellen Einheit verwaltet/archiviert werden (z.B. die Masterdateien eines digitalisierten Buches im TIFF-Format und ein PDF-Datei des gesamten Bandes)?	
11.5	Inwieweit besteht die Möglichkeit einer gekapselten Speicherung aller zu einem digitalen Objekt gehörigen Dateien sowie zugehöriger Metadaten in einem Containerformat für die Bereitstellung oder Aufbewahrung?	

12	Metadaten	Kommentar
12.1	Welche Formen von Metadaten werden unterstützt: a) deskriptiv (z.B. DC, MARC-XML, EAD, TEI, MODS, XDOMEA) b) strukturell (z.B. METS) c) LZA-spezifische Metadaten (z.B. PREMIS): technisch, eventbezogen, rechtlich d) Dateienmanagement	
12.2	Können bestimmte Teile der Metadaten automatisch generiert werden (z. B. durch die Einbindung von Tools wie JHOVE oder FITS)?	
12.3	Inwieweit ist das System für die Erzeugung, Verwaltung und Meldung von Persistenten Identifikatoren (z.B. URN, DOI) geeignet?	
12.4	Inwieweit können die vom System vorgegebenen Metadaten-Schemas angepasst oder frei konfiguriert werden?	
12.5	Inwieweit ist eine Kollektionsbildung nach Metadaten möglich?	
12.6	Wie und wo erfolgt die Ablage der Metadaten im Archivsystem? Ausschließlich in einer Datenbank oder auch zusätzlich im Filesystem selbst?	
12.7	Wie können Metadaten für die Suche indiziert werden?	

13	Authentifizierung/Rechteverwaltung	Kommentar
13.1	Ist eine einzelobjekt- und eine objektgruppenbezogene Rechtevergabe möglich? Wie können Parameter festgelegt werden?	
13.2	Welche externen Systeme zur Nutzerverwaltung/ Nutzerauthentifizierung lassen sich integrieren?	
13.3	Inwiefern kann jeweils eine spezifische Prüfung der Zugriffsrechte bei der Bereitstellung (View) eines digitalen Objekts erfolgen?	
13.4	Welche Differenzierung von unterschiedlichen Nutzer- und Gruppenrollen mit unterschiedlichen Berechtigungen in Bezug auf die Funktionalitäten des Systems ist möglich?	

14	Ablieferung/Ingest	Kommentar
14.1	Welche Schnittstellen und Verfahren stehen internen und externen Nutzern für den Upload von digitalen Objekten und Metadaten in das Archivsystem zur Verfügung? Stehen bereits grafische Nutzeroberflächen/Webformulare zum manuellen/automatischen Upload zur Verfügung?	
14.2	Sind Verfahren zum Massen-Upload vorhanden?	
14.3	Ist eine individuelle Anpassung des Workflows bzw. der Nutzeroberflächen für den Ingest möglich?	
14.4	Erfolgt nach dem Upload eine Identifizierung und Validierung der Dateien (z.B. mittels JHOVE, DROID) sowie eine Überprüfung von mitgelieferten Checksummen?	
14.5	Kann eine automatische Formatkonvertierung (z.B. Dateiformate, Metadaten) während des Ingests erfolgen oder ist sie integrierbar?	

15	Bearbeitung/Administration/Datenmanagement	Kommentar
15.1	Inwieweit sind Validierungsroutinen bzw. Qualitätskontrollen (manuell/automatisch) im System vorgesehen?	
15.2	Inwieweit ist für die internen Bearbeiter eine Editierbarkeit von digitalen Objekten bzw. der Metadaten gegeben und wie wird diese dokumentiert (Versionierung)?	
15.3	Steht ein Verfahren für den massenhaften Export bzw. Import (Batch-Processing) von digitalen Objekten zur Verfügung?	
15.4	Sind Massenupdates von Metadaten (z. B. Zugriffsrechte) als Prozesse möglich?	
15.5	Inwieweit kann das System statische Auswertungen bzw. sonstige Reports zur Verfügung stellen?	
15.6	Welche Formen der Benachrichtigungen/Systeminformationen können für interne bzw. externe Nutzer vom System selbst automatisch erzeugt werden?	

16	Search/Access/Delivery:	Kommentar
16.1	Ist das System für den gewünschten Grad an Zugänglichkeit durch die Öffentlichkeit bzw. externe Nutzer geeignet?	
16.2	Welche Suchoptionen stehen im System zur Verfügung? Nach welchen Kriterien kann gesucht werden?	
16.3	Inwiefern ist ein Browsing innerhalb der archivierten digitalen Objekte möglich?	
16.4	Inwieweit sind bereits Viewer/Player für die Präsentationsformate der digitalen Objekte in das System eingebunden? Ist eine Anpassung an spezifische Anforderungen möglich?	
16.5	Kann das System für die externen Nutzer persistente Links erzeugen?	
16.6	Sind in das System bereits Web-2.0-Funktionalitäten für die Endnutzer integriert?	
16.7	Sind die Anforderungen der Barrierefreiheit erfüllt?	
16.8	Inwieweit stehen Funktionalitäten für einen Massenexport für eine externe Nutzung zur Verfügung?	

17	Preservation Planning und Action	Kommentar
17.1	Können Risiken (z.B. veraltete oder proprietäre Dateiformate) aus Sicht der Langzeitarchivierung innerhalb des Systems verzeichnet und dokumentiert werden oder entsprechende externe Services (z.B. PRONOM) eingebunden werden? Gibt es die Möglichkeit, alternative Erhaltungsszenarien anhand ausgewählter Testobjekte vorab zu testen und zu bewerten?	
17.2	Inwieweit stehen Funktionalitäten für einen Massenexport und -import für die externe Durchführung einer Dateiformatmigration zur Verfügung?	
17.3	Inwieweit wird eine Dateiformatkonvertierung innerhalb des Systems unterstützt?	
17.4	Wie werden bei der Nutzung Repräsentationen im Falle einer durchgeführten Dateiformatmigration unterschieden?	
17.5	Können ältere Repräsentationen gezielt gelöscht oder beibehalten werden?	
17.6	Inwiefern ermöglicht das System die Anbindung von Emulatoren?	

18	Sicherheit der Archivobjekte/Integrität	Kommentar
18.1	Inwieweit und in welcher Form finden regelmäßige Überprüfungen (z.B. Virenchecks oder Checksummenüberprüfungen) der gespeicherten Objekte statt? Gibt es systemseitig Routinen, die greifen, sofern hier Probleme festgestellt werden?	
18.2	Inwieweit können Back-Up-Routinen über das Archivsystem konfiguriert bzw. gesteuert werden?	
18.3	Kann ein bestimmter Speicherbereich oder ein bestimmtes Speichermedium für einzelne Objekte bzw. Objektgruppen zugewiesen werden?	

19	Risiko- und Exitszenario	Kommentar
19.1	Wie können verschiedene Risiken überwacht werden?	
19.2	Inwieweit sind bereits Risikoszenarien definiert? Sind hierfür Notfalllösungen vorgesehen?	
19.3	Steht ein eindeutiges Ausstiegsszenario für den Fall eines Systemwechsels zur Verfügung?	
19.4	Sind alle aus Sicht der Langzeitarchivierung relevanten Informationen und Daten vollständig exportierbar und wiederherstellbar (wenn Software und/oder Datenbank nicht länger zur Verfügung stehen)?	
19.5	In welchem Format erfolgt der Export, existiert eine Beschreibung des Export-Formates und wie allgemeingültig ist der Aufbau des Export-Formates?	

Weiterführende Lektüre, die neben der technischen auch die organisatorischen Aspekte eines digitalen Langzeitarchivs umfasst, bietet unter anderem auch der nestor-Kriterienkatalog (verfügbar unter: URN: urn:nbn:de:0008-2008021802). Kommentare, Ergänzungen und Anregungen zu dieser Liste können Sie an gerne an die nestor-AG Kooperation & Vernetzung Tobias Beinert: beinert@bsb-muenchen.de
Natascha Schumann: n.schumann@dnb.de richten.

Stand: Juli 2011