nestor
Kompetenznetzwerk Langzeitarchivierung

nestor Handbuch:

Eine kleine Enzyklopädie der digitalen Langzeitarchivierung

8.3 Evaluierung der Vertrauenswürdigkeit digitaler Archive

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/

8.3 Evaluierung der Vertrauenswürdigkeit digitaler Archive

Susanne Dobratz und Astrid Schoger

8.3.1 Vertrauenswürdige Digitale Langzeitarchive

Die Anwendung von Methoden der IT-Sicherheit wie Hashfunktion und digitale Signatur kann bestimmte Risiken minimieren, insbesondere sie, die die Integrität, Authentizität und Vertraulichkeit digitaler Objekte betreffen. Das Problem einer breiten anwendung derartiger Technologien für die Langzeiterhaltung und vor allem für die Gewährleistung der Langzeitverfügbarkeit digitaler Objekte aus heutiger Sicht besteht vor allem darin, dass die langfristige Archivierung digitaler Signaturen technologisch nicht für beliebig große Datenmengen, komplexe Objekte und beliebig lange Zeiträume erprobt ist. Aktive Trustcenter, wie z.B. die Telesec GmbH[16] vergeben Zertifikate mit einer Gültigkeitsdauer von 3 Jahren. Danach müssen neue Zertifikate, die mit den alten verknüpft sind, ausgegeben und angewandt werden. Es handelt sich demnach um relativ kurzfristige Verfahren.

Dem gegenüber steht die Jahrhunderte lange Erfahrung der Archivare und Bibliothekare, die aus dem Blickpunkt der langfristigen Gewährleistung der Benutzbarkeit digitaler Objekte dem Einsatz digitaler Signaturen skeptisch gegenüberstehen.

Hier stehen die Aspekte der Verfügbarkeit und der Interpretierbarkeit digitaler Objekte in der Zukunft eine übergeordnete Rolle. Aus diesem Grunde konzentriert man sich darauf, organisatorische, wirtschaftlich-finanzielle Aspekte hervorzuheben und sich bei den technischen Aspekten auf die Methoden zu fokussieren, die die Anwendung von Normen und Standards bei der Abspeicherung der Objekte betreffen. Konkret sind dies die Aspekte des **Datenformats** und der **Metadaten** sowie der **Datenträger**, denen man eine besondere Bedeutung beimisst.

Daher haben verschiedene Organisationen und Initiativen mit der Formulierung von Anforderungen an vertrauenswürdige digitale Langzeitarchive begonnen. Diese Kriterien betreffen sowohl organisatorische als auch technische Rahmenbedingungen, die erfüllt werden müssen, um die Aufgabe der Erhaltung (der Interpretierbarkeit) digitaler Objekte gerecht werden zu können.

Dabei spielt die sogenannte Zielgruppe (engl. designated community) eine

16 Siehe http://ww.telesec.de

besondere Rolle, da z.B. die Interpretierbarkeit und die Nutzbarkeit digitaler Objekte auf die Vorkenntnisse, organisatorische und technische Benutzungsbedingungen und Nutzungsszenarien dieser Zielgruppe optimiert werden müssen. Die Anwendung konkreter Kriterien bzw. Anforderungen an das digitale Langzeitarchiv ist abhängig von der jeweiligen Zielgruppe.

Daher können allgemeingültige Anforderungen, wie sie die derzeit existierenden Kriterienkataloge darstellen, nur auf einem relativ abstrakten Niveau formuliert werden.

So hat die *nestor Arbeitsgruppe Vertrauenswürdige Archive – Zertifizierung* daher **Grundprinzipien bei der Herleitung und Anwendung der nestor-Kriterien** formuliert:

Abstraktion: Ziel des Kataloges ist es, Kriterien zu formulieren, die für ein breites Spektrum digitaler Langzeitarchive angewendet werden können und über längere Zeit Gültigkeit behalten sollen. Deshalb wird von relativ abstrakten Kriterien ausgegangen. Den Kriterien werden jeweils ausführliche Erläuterungen und konkrete Beispiele aus verschiedenen Bereichen mitgegeben. Die Beispiele entsprechen dem heutigen Stand der Technik und Organisation und sind unter Umständen nur im Kontext einer spezifischen Archivierungsaufgabe sinnvoll. Sie haben keinen Anspruch auf Vollständigkeit.

Dokumentation: Die Ziele, die Konzeption und Spezifikation sowie die Implementierung des digitalen Langzeitarchivs sind angemessen zu dokumentieren. Anhand der Dokumentation kann der Entwicklungsstand intern und extern bewertet werden. Eine frühzeitige Bewertung kann auch dazu dienen, Fehler durch eine ungeeignete Implementierung zu vermeiden. Insbesondere erlaubt es eine angemessene Dokumentation aller Stufen, die Schlüssigkeit eines digitales Langzeitarchiv umfassend zu bewerten. Auch alle Qualitäts- und Sicherheitsnormen fordern eine angemessene Dokumentation.

Transparenz: Transparenz wird realisiert durch die Veröffentlichung geeigneter Teile der Dokumentation. Transparenz nach außen gegenüber Nutzern und Partnern ermöglicht diesen, selbst den Grad an Vertrauenswürdigkeit festzustellen. Transparenz gegenüber Produzenten und Lieferanten bietet diesen die Möglichkeit zu bewerten, wem sie ihre digitalen Objekte anvertrauen. Die Transparenz nach innen dokumentiert gegenüber den Betreibern, den Trägern, dem Management sowie den Mitarbeitern die angemessene Qualität des digitalen Langzeitarchivs und sichert die Nachvollziehbarkeit der Maßnahmen. Bei denjenigen Teilen der Dokumentation, die für die breite Öffentlichkeit nicht geeignet sind (z.B. Firmengeheimnisse, Informationen mit Sicherheitsbezug), kann die Transparenz auf einen ausgewählten Kreis (z.B. zertifizierende Stelle) beschränkt werden. Durch das Prinzip der Transparenz wird Vertrauen aufge-

baut, da es die unmittelbare Bewertung der Qualität eines digitalen Langzeitarchivs durch Interessierte zulässt.
Angemessenheit: Das Prinzip der Angemessenheit berücksichtigt die Tatsache, dass keine absoluten Maßstäbe möglich sind, sondern dass sich die Bewertung immer an den Zielen und Aufgaben des jeweiligen digitalen Langzeitarchivs ausrichtet. Die Kriterien müssen im Kontext der jeweiligen Archivierungsaufgabe gesehen werden. Deshalb können ggf. einzelne Kriterien irrelevant sein. Auch der notwendige Erfüllungsgrad eines Kriteriums kann – je nach den Zielen und Aufgaben des digitalen Langzeitarchivs – unterschiedlich ausfallen.
Bewertbarkeit: Für die Vertrauenswürdigkeit existieren zum Teil - insbesondere unter Langzeitaspekten - keine objektiv bewertbaren (messbaren) Merkmale. In diesen Fällen ist man auf Indikatoren angewiesen, die den Grad der Vertrauenswürdigkeit repräsentieren. Transparenz macht auch die Indikatoren für eine Bewertung zugänglich.

8.3.2 Einige Definitionen

Die folgenden Begriffe sind im Zusammenhang mit vertrauenswürdigen digitalen Langzeitarchiven essentiell und orientieren sich am OAIS-Modell (siehe entsprechendes Kapitel im
Handbuch).

Digitales Objekt, Metadaten

Ein digitales Objekt ist eine logisch abgegrenzte Informationseinheit in der Form digitaler Daten. Daten sind maschinenlesbare und –bearbeitbare Repräsentationen von Information, in digitaler Form (eine Bitfolge, also eine Folge von Nullen und Einsen). Zur Nutzung der Informationen müssen die digitalen Daten interpretiert (dekodiert) werden.

Der Informationsbegriff umfasst hier jeden Typ von Wissen, der ausgetauscht werden kann; zum Beispiel aus inhaltlicher Sicht etwa Werke geistiger Schöpfung, Ergebnisse der Forschung und Entwicklung, Dokumentationen des politischen, sozialen und wirtschaftlichen Handelns.

Zu den Daten, die die Inhaltsinformation repräsentieren (Inhaltsdaten), können weitere Daten hinzukommen, die z.B. der Identifizierung, der Auffindbarkeit, der Rekonstruktion und Interpretation oder dem Nachweis der Integrität und Authentizität sowie der Kontrolle der Nutzungsrechte dienen (Metadaten). Metadaten können zu unterschiedlichen Zeiten im Lebenszyklus digitaler Objekte

entstehen (z.B. bei der Produktion, bei der Archivierung, bei der Bereitstellung für die Nutzung). Sie werden als Teile der logischen Einheit „digitales Objekt" aufgefasst und können sowohl getrennt als auch gemeinsam mit den Inhaltsdaten verwaltetet werden.

Digitales Langzeitarchiv, Vertrauenswürdigkeit

Unter einem digitalen Langzeitarchiv wird eine Organisation (bestehend aus Personen und technischen Systemen) verstanden, die die Verantwortung für den Langzeiterhalt und die Langzeitverfügbarkeit digitaler Objekte sowie für ihre Interpretierbarkeit zum Zwecke der Nutzung durch eine bestimmte Zielgruppe übernommen hat. Dabei bedeutet „Langzeit" über Veränderungen in der Technik (Soft- und Hardware) hinweg und auch unter Berücksichtigung möglicher Änderungen der Zielgruppe. Vertrauenswürdigkeit (engl. trustworthiness) wird als Eigenschaft eines Systems angesehen, gemäß seinen Zielen und Spezifikationen zu operieren (d.h. es tut genau das, was es zu tun vorgibt). Aus Sicht der IT-Sicherheit stellen Integrität, Authentizität, Vertraulichkeit und Verfügbarkeit Grundwerte dar. IT-Sicherheit ist somit ein wichtiger Baustein für vertrauenswürdige digitale Langzeitarchive.

8.3.3 Kriterienkataloge für vertrauenswürdige digitale Archive

Die Überprüfung und Bewertung der eingesetzten Maßnahmen zur Minimierung der Risiken, die den Langzeiterhalt der durch die digitalen Objekte repräsentierten Information bedrohen, erzeugt Vertrauenswürdigkeit. Diese kann anhand eines Kriterienkatalogs für Vertrauenswürdige digitaler Langzeitarchive geprüft und bewertet werden.
Dabei existieren internationale mehrer Ansätze.
Die Grundvoraussetzung für die Vertrauenswürdigkeit aller digitalen Langzeitarchive ist die, dass jedes nach seinen Zielen und Spezifikationen operiert. Diese sind durch die jeweilige Zielgruppe bestimmt.

Ein digitales Langzeitarchiv entsteht als komplexer Gesamtzusammenhang. Die Umsetzung der einzelnen Kriterien muss stets vor dem Hintergrund der Ziele des Gesamtsystems gesehen werden. Sowohl die Realisierung des digitalen Langzeitarchivs als Ganzes als auch die Erfüllung der einzelnen Kriterien läuft als Prozess in mehreren Stufen ab:

1. Konzeption

2. Planung und Spezifikation
3. Umsetzung und Implementation
4. Evaluierung

Im Zuge der ständigen Verbesserung sind diese Stufen nicht als starres Phasenmodell zu
betrachten sondern zu wiederholen.

Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC)

Auf der Grundlage der Eigenschaften und Verantwortlichkeiten eines digitalen Langzeitarchivs, die bereits 2002 im Bericht der RLG/OCLC Working Group on Digital Archive Attributes als wesentlich für deren Vertrauenswürdigkeit aufgeführt wurden, hat die RLG-NARA Task Force on Digital Repository Certification 2006 als Entwurf in überarbeiteter Fassung 2007 eine Liste von Kriterien erarbeitet, die ein vertrauenswürdiges digitales Langzeitarchiv erfüllen müssen. Diese Liste dient der Orientierung, kann als Checkliste auch zur Selbstevaluierung und zum externen Audit eingesetzt werden, vgl. RLG-NARA 2007.

Nestor-Kriterienkatalog

nestor hat unter Berücksichtigung nationaler Ansätze und Arbeitsergebnisse wie des „DINI-Zertifikats für Dokumenten- und Publikationsserver“, vgl. DINI 2006, die Kriterien nationalen Rahmenbedingungen und den Bedürfnissen der deutschen Gedächtnisorganisationen angepasst und im Sommer 2006 als Entwurf zur öffentlichen Kommentierung veröffentlicht.
Überblick über die nestor-Kriterien:
Aus Gründen der Übersichtlichkeit wird im Folgenden der Term „digitales Langzeitarchiv“ mit „dLZA“ abgekürzt.

A. Organisatorischer Rahmen

1. Das dLZA hat seine Ziele definiert.

1.1 Das dLZA hat Kriterien für die Auswahl seiner digitalen Objekte entwickelt.

1.2 Das dLZA übernimmt die Verantwortung für den dauerhaften Erhalt der durch die digitalen Objekte repräsentierten Informationen.

1.3 Das dLZA hat seine Zielgruppe(n) definiert.

2. Das dLZA ermöglicht seinen Zielgruppe(n) eine angemessene Nutzung der durch die digitalen Objekte repräsentierten Informationen.

2.1 Das dLZA ermöglicht seinen Zielgruppe(n) den Zugang zu den durch die digitalen Objekte repräsentierten Informationen.

2.2 Das dLZA stellt die Interpretierbarkeit der digitalen Objekte durch seine Zielgruppe(n) sicher.

3. Gesetzliche und vertragliche Regelungen werden eingehalten.

3.1 Es bestehen rechtliche Regelungen zwischen Produzenten und dem digitalen Langzeitarchiv.

3.2 Das dLZA handelt bei der Archivierung auf der Basis rechtlicher Regelungen.

3.3 Das dLZA handelt bei der Nutzung auf der Basis rechtlicher Regelungen.

4. Die Organisationsform ist für das dLZA angemessen.

4.1 Die Finanzierung des digitalen Langzeitarchivs ist sichergestellt.

4.2 Es steht Personal mit angemessener Qualifikation in ausreichendem Umfang zur Verfügung.

4.3 Es bestehen angemessene Organisationsstrukturen für das dLZA.

4.4 Das dLZA betreibt eine langfristige Planung.

4.5 Die Fortführung der festgelegten Aufgaben ist auch über das Bestehen des digitalen Langzeitarchivs hinaus sichergestellt.

5. Es wird ein angemessenes Qualitätsmanagement durchgeführt.

5.1 Alle Prozesse und Verantwortlichkeiten sind definiert.

5.2 Das dLZA dokumentiert alle seine Elemente nach einem definierten Verfahren.

5.3 Das dLZA reagiert auf substantielle Veränderungen.

B. Umgang mit Objekten

6. Das dLZA stellt die Integrität der digitalen Objekte auf allen Stufen der Verarbeitung sicher.

6.1 Aufnahme (Ingest): Das dLZA sichert die Integrität der digitalen Objekte.

6.2 Archivablage (Archival Storage): Das dLZA sichert die Integrität der digitalen Objekte .

6.3 Nutzung (Access): Das dLZA sichert die Integrität der digitalen Objekte.

7. Das dLZA stellt die Authentizität der digitalen Objekte und Metadaten auf allen Stufen der Verarbeitung sicher.

7.1 Aufnahme (Ingest): Das dLZA sichert die Authentizität der digitalen Objekte.

7.2 Archivablage (Archival Storage): Das dLZA sichert die Authentizität der digitalen Objekte.

7.3 Nutzung (Access): Das dLZA sichert die Authentizität der digitalen Objekte.

8. Das dLZA betreibt eine langfristige Planung seiner technischen Langzeiterhaltungsmaßnahmen.

9. Das dLZA übernimmt digitale Objekte von den Produzenten nach definierten Vorgaben.

9.1 Das dLZA spezifiziert seine Übergabeobjekte (Submission Information Packages, SIPs).

9.2Das dLZA identifiziert, welche Eigenschaften der digitalen Objekte für den Erhalt von Informationen signifikant sind.

9.3Das dLZA erhält die physische Kontrolle über die digitalen Objekte, um Lang-zeit-archivierungsmaßnahmen durchführen zu können.

10. Die Archivierung digitaler Objekte erfolgt nach definierten Vorgaben.

10.1 Das dLZA definiert seine Archivobjekte (Archival Information Packages, AIPs).

10.2 Das dLZA sorgt für eine Transformation der Übergabeobjekte in Archivobjekte.

10.3 Das dLZA gewährleistet die Speicherung und Lesbarkeit der Archivobjekte.

10.4 Das dLZA setzt Strategien zum Langzeiterhalt für jedes Archivobjekt um.

11. Das dLZA ermöglicht die Nutzung der digitalen Objekte nach definierten Vorgaben.

11.1 Das dLZA definiert seine Nutzungsobjekte (Dissemination Information Packages, DIPs).

11.2 Das dLZA gewährleistet eine Transformation der Archivobjekte in Nutzungsobjekte.

12 Das Datenmanagement ist dazu geeignet, die notwendigen Funktionalitäten des digitalen Langzeitarchivs zu gewährleisten.

12.1 Das dLZA identifiziert seine Objekte und deren Beziehungen eindeutig und dauerhaft.

12.2 Das dLZA erhebt in ausreichendem Maße Metadaten für eine formale und inhaltliche Beschreibung und Identifizierung der digitalen Objekte.

12.3 Das dLZA erhebt in ausreichendem Maße Metadaten zur strukturellen Beschreibung der digitalen Objekte.

12.4 Das dLZA erhebt in ausreichendem Maße Metadaten, die die vom Archiv vorgenommenen Veränderungen an den digitalen Objekten verzeichnen.

12.5 Das dLZA erhebt in ausreichendem Maße Metadaten zur technischen Beschreibung der digitalen Objekte.

12.6 Das dLZA erhebt in ausreichendem Maße Metadaten, die die entsprechenden Nutzungsrechte und –bedingungen verzeichnen.

12.7 Die Zuordnung der Metadaten zu den Objekten ist zu jeder Zeit gegeben.

C. Infrastruktur und Sicherheit

13 Die ITInfrastruktur ist angemessen.

13.1 Die IT–Infrastruktur setzt die Forderungen aus dem Umgang mit Objekten um.

13.2 Die IT-Infrastruktur setzt die Sicherheitsanforderungen des IT-Sicherheitskonzeptes um.
14 Die Infrastruktur gewährleistet den Schutz des digitalen Langzeitarchivs und seiner digitalen Objekte.

10 gemeinsame Prinzipien und ISO

Die Kriterien für die Vertrauenswürdigkeit digitaler Langzeitarchive befinden sich zurzeit im Prozess internationaler Abstimmung und Standardisierung im Rahmen der ISO.
Wesentliche Vertreter des Themas Vertrauenswürdigkeit auf internationaler Ebene - Center of Research Libraries CRL, Digital Curation Centre DCC, Projekt Digital Preservation Europe DPE sowie nestor haben 10 gemeinsame Prinzipien herausgearbeitet, vgl. (CRL, DCC, DPE, nestor 2007), die den oben genannten Kriterienkatalogen und Audit Checklisten zu Grunde liegen. Diese stellen die Grundlage der weiteren Zusammenarbeit dar.
Ferner arbeitet eine internationale Arbeitsgruppe daran, die Kriterien für die Standardisierung im Rahmen der ISO vorzubereiten[17].

8.3.4 Wie wird evaluiert?

Orientierung, Selbstevaluierung, Audits

Die oben vorgestellten Kriterienkataloge und Checklisten dienen zurzeit zur Orientierung beim Aufbau digitaler Langzeitarchive und zur Selbstevaluierung sowie für externe Audits. Ein digitales Langzeitarchiv entsteht als komplexer Gesamtzusammenhang. Die Umsetzung der einzelnen Kriterien muss stets vor dem Hintergrund der Ziele des Gesamtsystems gesehen werden. Sowohl die Realisierung des digitales Langzeitarchivs als Ganzes als auch die Erfüllung der einzelnen Kriterien läuft als Prozess in mehreren Stufen ab: 1. Konzeption, 2. Planung und Spezifikation, 3. Umsetzung und Implementation, 4. Evaluierung. Im Zuge der ständigen Verbesserung sind diese Stufen nicht als starres Phasenmodell zu betrachten sondern zu wiederholen.

Digital Repository Audit Method based on Risk Assessment DRAMBORA

Im Rahmen des EU-Projektes Digitale Preservation Europe in Zusammenarbeit mit Digital Curation Centre wurde ein Tool zur Selbstevaluierung entwi-

17 In einer Birds Of Feather Gruppe unter Leitung von David Giaretta, vgl.

ckelt, das die Risikoanalyse als Methode einsetzt. Ausgehend von den Zielen eines digitalen Langzeitarchivs müssen zunächst die Werte und Aktivitäten spezifiziert, in einem weiteren Schritt dann die damit verbundenen Risiken identifiziert und bewertet werden.

Zertifizierung

Bevor ein international abgestimmtes Zertifizierungsverfahren für digitale Langzeitarchive entwickelt werden kann, muss zunächst ein internationaler Konsens über die Evaluierungskriterien gefunden werden. Ferner müssen aus den Erfahrungen mit der Anwendung der Kriterienkataloge und Evaluierungstools Bewertungsmaßstäbe für unterschiedliche Typen von digitalen Langzeitarchiven ausgearbeitet werden.