Empfehlungen zur Gewährleistung der Authentizität und Integrität archivierter digitaler Unterlagen

AutorInnen Arbeitsgruppe Authentizität des Pilotprojekts KOSTPROBE:
Georg Büchler, KOST
Markus Lischer, StALU
Claudia Schmucki, StAZH
Peter Witschi, StAAR
Redaktion Georg Büchler
Version 1.1

Disposition

0 Die wichtigsten Punkte

- Jedes Archiv braucht normative Grundlagen (Policies) für seine Arbeit. Diese tragen auch zur Authentizität der archivierten Unterlagen bei.
- Archivierte Unterlagen müssen in einen neuen strukturellen Kontext eingebettet werden, nämlich in die Archivtektonik.
- Bei einer Migration muss sich das Archiv anhand einer vorhergehenden Analyse und einer nachfolgenden Stichprobe überzeugen können, dass die essentiellen Eigenschaften der Unterlagen erhalten geblieben sind.
- Die Integrität archivierter Unterlagen wird mit einer Reihe von Massnahmen gewährleistet:
  - Grundprinzip ist die redundante Speicherung.
  - Der Zugriff auf die gespeicherten Unterlagen muss strikt kontrolliert sein.
  - Veränderungen an den archivierten Daten müssen protokolliert werden.
  - Die Verwendung digitaler Signaturen ist für die Langzeitarchivierung nicht sinnvoll.
  - Um die Integrität der archivierten Unterlagen jederzeit einfach kontrollieren zu können, empfiehlt es sich, auf kryptografische Algorithmen, insbesondere Hash-Werte, zurückzugreifen.

1 Ausgangslage

Die Arbeitsgruppe hat den Auftrag, Empfehlungen zu formulieren, wie die Authentizität archivierter digitaler Unterlagen nachvollziehbar gewährleistet werden kann.
Authentizität bedingt Identität und Integrität. Authentische Unterlagen sind, was sie zu sein vorgeben (Identität), und sind nicht unbefugt und undokumentiert verändert worden (Integrität)[1].
Nachvollziehbar meint, dass sowohl Verwaltungsstellen und Archivpersonal als auch BenutzerInnen die Möglichkeit haben, sich zu vergewissern, ob vom Archiv aufbewahrte digitale Unterlagen authentisch sind.

Jedes Archiv braucht normative Grundlagen (Policies) für seine Arbeit. Im Besonderen müssen offizielle, genehmigte und bekannte Richtlinien existieren, die festlegen, wie mit archivierten digitalen Unterlagen umgegangen werden soll. Normative Grundlagen erfüllen eine doppelte Funktion: Sie tragen wesentlich zur Glaubwürdigkeit sowie zur Nachvollziehbarkeit der Arbeit bei, und sie sind eine Voraussetzung für rationelles Arbeiten, da sie Einzelfallentscheidungen deutlich verringern. Die vorliegende Empfehlung versteht sich deshalb als Ausgangspunkt für eine Richtlinie zur Gewährleistung der Authentizität archivierter Unterlagen.

2 Risikoanalyse

Grundsätzlich versteht sich von selbst, dass sämtliche Stadien der digitalen Archivierung Risiken für die Authentizität der archivierten Unterlagen beinhalten. Insbesondere sind die Hauptprobleme der technischen und intellektuellen Obsoleszenz von Bedeutung: Wenn Unterlagen nicht mehr gelesen oder verstanden werden können, ist ihre Integrität nicht mehr gewährleistet. In diesem Fall ist jedoch der Verlust der *Benutzbarkeit* das eigentliche Problem, das in anderem Zusammenhang angegangen wird, besonders im OAIS-Funktionsbereich Preservation Planning.

Spezifische Risiken umfassen:
Verlust des Kontexts. In der archivischen Theorie kommt dem Kontext eine grosse Bedeutung für die Garantie der Authentizität zu. Für sich selbst genommen sind Dokumente oder Daten nicht authentisch; Authentizität können sie nur beanspruchen, wenn sie in einem Zusammenhang stehen. Da ihr Ursprungskontext im Rahmen der Archivierung bewusst und notwendigerweise aufgebrochen wird, besteht dabei ein inhärentes Risiko für die Authentizität digitaler Unterlagen.
Bewusste Manipulation. Es kann davon ausgegangen werden, dass die Unterlagen so lange in der Verwaltungsstelle verbleiben, wie sie für deren Geschäfte von Bedeutung sind. Erst wenn sie keine organisatorische Funktion

[1] Gemäss der Definition des InterPARES-Projekts, siehe INTERPARES 1, Authenticity Task Force Report (2003) p. 2 (online unter http://www.interpares.org/book/interpares_book_d_part1.pdf).

mehr erfüllen, werden sie in der Regel dem Archiv angeboten. Da es damit praktisch ausgeschlossen ist, dass sich jemand durch bewusste und absichtliche Manipulation archivierter Unterlagen einen Vorteil finanzieller oder anderer Art verschaffen könnte, kann das Risiko solcher Manipulationen in dieser Empfehlung wenn nicht vernachlässigt, so doch von den restlichen Empfehlungen abgedeckt werden. Als Sonderfall stellt jedoch die Migration archivierter Unterlagen einen bewussten, aber notwendigen Eingriff in deren Integrität dar.
Unbeabsichtigte Manipulation. Das grösste Risiko für die Authentizität archivierter Unterlagen sind unbeabsichtigte und fahrlässige Manipulationen.

3 Problemanalyse

Aufbauend auf den oben identifizierten Risiken können nun konkrete Probleme analysiert werden, die im Rahmen der digitalen Archivierung zu bestimmten Zeiten gelöst werden müssen.

1. Wie kann der Verlust des Kontexts aufgefangen werden? Entsprechend den Verfahren bei herkömmlichen Akten setzt die Verantwortung des Archivs für die Unterlagen grundsätzlich im Moment der Ablieferung ein. Da die Ablieferung digitaler Unterlagen jedoch einen grösseren Eingriff in deren Kontext darstellt als diejenige analoger Unterlagen, besteht bereits bei der Vorbereitung der Ablieferung Handlungsbedarf. Dabei muss das Archiv sicherstellen, dass der ursprüngliche Kontext erhalten oder mindestens dokumentiert wird.
2. Wie kann die Authentizität archivierter Unterlagen nach einer Migration gewährleistet werden? Dieses Problem betrifft auch die allfällige Extraktion der Daten aus der Ursprungsanwendung in ein archivtaugliches Format.
3. Wie kann die Integrität archivierter Unterlagen in der Zeitspanne zwischen zwei Migrationen nachgewiesen werden? D.h., wie kann der/die Benutzer/in sich vergewissern, dass die ausgelieferte Datei gleich ist, wie sie nach der Übernahme bzw. der letzten Migration war?
4. Wie können Kopien von vom Archiv nach aussen gegebenen Unterlagen authentifiziert werden, so dass Dritte eine Möglichkeit zur Verifikation haben, dass es sich dabei um eine der authentischen gleichwertige Unterlage handelt?

Diese vier Problemstellungen werden in der Folge behandelt.

4 Kontextinformationen

In der Literatur herrscht Einigkeit, dass zur Bewahrung des Kontexts Dokumentation und Metadaten von entscheidender Bedeutung sind. Da die Unterlagen aus ihrem Kontext zum Teil herausgelöst werden, müssen verloren gegangene Beziehungen dokumentiert und nachvollziehbar gemacht werden. Die folgenden Elemente tragen dazu bei:

- Archivierte Unterlagen müssen in einen neuen strukturellen Kontext eingebettet werden, nämlich in die Archivtektonik, die im Normalfall in

einem Archivinformationssystem abgebildet ist. Aus der Archivtektonik lassen sich implizite Kontextinformationen herauslesen. Auf diese Weise wird der Ursprungskontext der archivierten Unterlagen nachgebildet.
- In den archivischen Metadaten werden unter anderem Informationen verzeichnet, die im ursprünglichen Kontext der Unterlagen zu deren Authentizität beitrugen, im archivierten Kontext jedoch nicht mehr vorhanden wären. Der Metadatenkatalog der Arbeitsgruppe Metadaten beinhaltet die wichtigsten derartigen Metadaten.
- Eine Dokumentation des Ursprungskontexts muss im Rahmen der Übernahme erstellt werden. Ihr Inhalt hängt von der Art der archivierten Unterlagen ab. Beispielsweise ist es für Unterlagen aus datenbankgestützten Informationssysteme nützlich zu dokumentieren, wie mit diesen Systemen gearbeitet wurde: Eingabemasken, Workflow, Berechtigungen etc.

5 Migrationen

Bei einer Migration aus einem Dateiformat in ein anderes wird die Dateistruktur der archivierten Unterlagen geändert, damit ihr Inhalt weiterhin lesbar und verständlich ist. Dies bedeutet eine gewollte und notwendige Veränderung der Unterlagen. Dabei muss gewährleistet werden, dass alle essentiellen Eigenschaften (siehe unten) erhalten bleiben, da nur so die Unterlagen weiterhin als authentisch gelten können.

Grundsatz
Das Archiv muss sich anhand einer vorhergehenden Analyse und einer nachfolgenden Stichprobe überzeugen können, dass die wichtigen Eigenschaften der Unterlagen erhalten geblieben sind. Diese Kontrolle muss wie die ganze Migration auch dokumentiert werden.

Vorgehen:
1. Vor der Migration: Definition der essentiellen Eigenschaften der Unterlagen:
   - einerseits (statistische) Kerngrössen auf Bestandesniveau: Anzahl und Art der Datensätze, Minimal- und Maximalwerte, Median- und Durchschnittswerte, etc.
   - anderseits auf Datensatz- bzw. Dokumentniveau: Welche Information ist essentiell? Dies ergibt sich zu einem guten Teil aus der Bewertung, zusätzlich allenfalls auch aus einer Neubewertung und aus Benutzerfeedback. Essentielle Informationen können unter verschiedene Kategorien fallen, insbesondere Inhalt, Struktur, Darstellung. Idealerweise werden diese essentiellen Informationen nach Kategorien gegliedert in den Metadaten zu den Unterlagen festgehalten.
2. Nach der Migration: Kontrolle der essentiellen Eigenschaften:
   - Überprüfung sämtlicher Kerngrössen auf Bestandesniveau.

- Stichprobe auf Datensatz- bzw. Dokumentniveau. Dabei muss zweierlei beachtet werden: (1) Es ist nicht möglich, mit Stichproben hundertprozentige Sicherheit zu gewährleisten, sondern nur eine Abschätzung der Sicherheit. (2) Die Auswahl der Stichproben muss im Einzelfall entschieden werden.
- Festhalten des Kontrollergebnisses.

Bei Misserfolg muss die Migration wiederholt werden. Die Unterlagen im Ursprungsformat müssen mindestens solange gespeichert bleiben, bis die Migration vollständig und erfolgreich abgeschlossen ist.
3. Protokollierung und Dokumentation der Migration: Der Metadatenkatalog sieht dazu Elemente vor. Die Arbeitsgruppe schlägt vor, darin nur wichtige Kenndaten festzuhalten, und daneben ein eigenes Dossier für die Migration zu führen. Dies ist analog zum herkömmlichen Vorgehen bei Bewertungen, wo die entsprechende Dokumentation ebenfalls getrennt von den archivierten Unterlagen aufbewahrt wird. Zudem ist es wahrscheinlich, dass sich ein Migrationsvorhaben auf mehr als eine Ablieferung erstreckt, da es wohl durch die verwendeten Formate gesteuert wird. Digitale Werkzeuge für die (teil)automatisierte Migration müssen alle relevanten Abläufe und eventuelle Fehlschläge angemessen protokollieren.

6 Integrität

Zur Gewährleistung der Integrität archivierter Unterlagen wird eine Reihe von Massnahmen vorgeschlagen, die auf zwei Ansatzpunkten beruhen: Mechanismen zur Datensicherheit sowie Kontrollmöglichkeiten.

- Grundprinzip ist die redundante Speicherung, welche die Wahrscheinlichkeit von Integritätsverletzungen entscheidend verringert. Eine Kopie wird als Master benutzt, die andere(n) dient (dienen) als Sicherheitskopie. Die Kopien sollen auf verschiedenen Datenträgern gespeichert werden, um Abhängigkeiten von Technologien und Anbietern zu verringern. Darunter befindet sich mit Vorteil ein nicht veränderbarer Datenträger[2]. Die Kopien dürfen nicht am gleichen Ort aufbewahrt werden.
  Es ist zu beachten, dass herkömmliche Backup-Technologien diesen Anforderungen nicht genügen.
- Der Zugriff auf die gespeicherten Unterlagen muss strikt kontrolliert sein. Längerfristig wird via Speichersoftware auf die archivierten Daten zugegriffen. Nur so wenige Personen wie nötig sollen zum direkten Zugriff auf die Daten berechtigt sein. Unter Umständen ist eine Verteilung der Zugriffsrechte beider Kopien auf verschiedene Personen in Betracht zu ziehen. Für die Kontrolle dieser Rechte können die Zugriffe geloggt und die entsprechenden Logfiles konsultiert werden (siehe dazu auch den folgenden Abschnitt).

[2] Siehe dazu Art. 9 der Verordnung über die Führung und Aufbewahrung der Geschäftsbücher (GeBüV) vom 24. April 2002, SR 221.431 (online unter http://www.admin.ch/ch/d/sr/c221_431.html).

- Veränderungen an den archivierten Daten müssen protokolliert werden. Speicherverwaltungssoftware muss deshalb eine detaillierte Logging-Funktion anbieten[3]. Festgehalten werden müssen bei Änderungen die folgenden Informationen: betroffenes Objekt, Art der Manipulation, Datum und Zeit, allenfalls ausführende Person oder Software.
  Die Logfiles geben Aufschluss über die Quelle allfälliger Fehler, können diese jedoch nicht korrigieren.
- Die Verwendung digitaler Signaturen ist für die Langzeitarchivierung nicht sinnvoll. Die Pflege solcher Signaturen ist für den Einsatz im archivischen Zeithorizont viel zu ressourcenintensiv. Die Arbeitsgruppe ist der Meinung, dass das Archiv selber die Authentizität seiner Unterlagen garantieren soll und dafür nicht auf einen externen Dienstleister wie einen Signaturprovider zurückgreifen soll[4].
- Um die Integrität der archivierten Unterlagen jederzeit einfach kontrollieren zu können, empfiehlt es sich, auf kryptografische Algorithmen, insbesondere Hash-Werte, zurückzugreifen. Zum Nachweis der Integrität wird der Hash-Wert einer Unterlage berechnet und mit dem separat hinterlegten Wert verglichen[5].

7 Beglaubigung

Für die Beglaubigung der Authentizität von Dokumenten (bzw. Kopien davon), die von Archiv ausgelieferten werden, eignet sich wiederum der Hash-basierte Ansatz. Zusammen mit der Kopie eines Dokuments soll sein Hash-Wert ausgeliefert werden. Falls die Authentizität dieses Dokumentes in Zweifel gezogen werden sollte, d.h. falls unklar sein sollte, ob das vorliegende Dokument tatsächlich das gleiche ist, welches von Archiv ausgeliefert wurde, kann sein Hash-Wert errechnet und mit dem mitgelieferten Hash-Wert verglichen werden.

[3] Als Beispiel können die OpenSource-Produkte Fedora oder DSpace gelten.
[4] Für die Langzeitarchivierung hochsensibler Unterlagen wird im Ausland gelegentlich auf digitale Signaturen zurückgegriffen, so im Urkundenarchiv der österreichischen Notare (siehe http://www.notar.at/de/portal/einrichtungen/cyberdocgmbhcokg/ und http://w4.siemens.de/de2/html/press/edesk/2001/adhm_011_01.html) sowie im Elektronischen Grundbuch von Baden-Württemberg (siehe http://www.elektronisches-grundbuch.de/, bzw. http://www.elektronisches-grundbuch.de/html/konzept.htm). Vergleichbare Anwendungen in der Schweiz verzichten auf digitale Signaturen: Sowohl das elektronische Zivilstandsregister Infostar als auch das digital geführte Grundbuch setzen auf (den hier skizzierten ähnliche) rigorose Massnahmen zur Datensicherheit (Auskünfte von Martin Jäger [Infostar] und Maria-Pia Portmann [Grundbuch], Bundesamt für Justiz).
[5] Die Überlegungen zur Verwendung von Hash-Werten werden René Quillet, Staatsarchiv Basel-Landschaft, verdankt. Siehe dazu auch T.C. Stein, E.A. Guinness, S.H.Slavney, „Establishing a Mechanism for Maintaining File Integrity within the Data Archive", paper presented at PV 2005, Edinburgh (vorläufige Publikation online unter http://www.ukoln.ac.uk/events/pv-2005/pv-2005-final-papers/039.pdf).