

KOST.07

Kolloquium „Archivtaugliche Speicherinfrastruktur“

Einführungsveranstaltung

Inhalt

1	Zur Einleitung	2
1.1	Das Kolloquium.....	2
1.2	Ziel der Einführungsveranstaltung	3
1.3	Dokumentation des Kolloquiums	3
2	Archivische Anforderungen	4
2.1	Grundsatz	4
2.2	Exkurs zur Vertrauenswürdigkeit	4
2.3	Gliederung der Anforderungen	4
2.4	Anforderungskatalog.....	5
3	Technische Aspekte	6
3.1	Speicherplatz mieten	7
3.2	Speicherzentrum.....	7
3.3	Blackbox	8
3.4	Datenauslagerung.....	8
3.5	Speichern im Netzwerk	9
4	Organisatorische und rechtliche Fragen	10
4.1	Organisatorische Fragen	10
4.2	Rechtliche Fragen.....	10
4.2.1	Archivgesetzliche Grundlagen	11
4.2.2	Datenschutz	11
5	Kostenaspekte digitaler Archivierung	12
5.1	Speicherbewirtschaftung / Lifecycle Management.....	12
5.2	Kostenvergleich digitales Speichersystem – Papiermagazin.....	13
5.3	Datenmenge	14
5.4	Zusammenfassung	14

1 Zur Einleitung

1.1 Das Kolloquium

In den ersten Pilotprojekten der KOST wurde erkannt, dass das Thema der archivtauglichen Speicherinfrastruktur nicht im Rahmen eines Archivierungsprojekts, sondern separat angegangen werden sollte. Die KOST hat deshalb letztes Jahr vorgeschlagen, sich in ihrem Programm 2007 mit diesem Thema zu befassen. Zur Abgrenzung präzisieren wir, dass es dabei einzig um die Speicherung der archivierten Unterlagen geht, also um die *Bitstream Preservation*, nicht aber um die Bewirtschaftung der Unterlagen (Metadatenverwaltung, Migrationen etc.).

Die Situation bei den KOST-Mitgliedsarchiven im Bereich der digitalen Archivierung präsentiert sich zurzeit folgendermassen: Die Archive sind noch mit wenig bis keinen Angeboten digitaler Unterlagen konfrontiert. Wenige Archive haben bereits gewisse digitale Unterlagen übernommen; diese sind jeweils in einer provisorischen Speicherinfrastruktur abgelegt. Wir befinden uns also noch in einer gewissen Schonfrist; es ist allerdings absehbar, dass in den nächsten paar Jahren einerseits digitale Unterlagen in grösserem Umfang angeboten werden, andererseits die jetzigen Provisorien abgelöst werden müssen. Deshalb ist jetzt der richtige Moment, sich im Bewusstsein des Problems, aber noch ohne dringenden Druck mit der Beschaffung einer archivtauglichen Speicherinfrastruktur zu befassen zu beginnen. Die KOST hat sich in der Detailplanung dafür entschieden, dies mit einer Art Auslegeordnung in Form eines Kolloquiums zu tun. Dieses Kolloquium hat zum Ziel, den Mitgliedsarchiven fünf verschiedene Speicherkonzepte an Hand von Anbieterpräsentationen und Erfahrungsberichten vorzustellen und sie auf ihre Eignung für die konkreten Bedürfnisse zu evaluieren. Jedes dieser Konzepte wird in einer zweieinhalbstündigen Veranstaltung von zwei Referenten vorgestellt und im Licht archivischer Anforderungen diskutiert. Diese fünf thematischen Veranstaltungen werden gerahmt von einer Eröffnungs- und einer Abschlussveranstaltung.

Das Projekt hat mehrere Ziele. Zum einen werden sich die Teilnehmenden vertieftes Wissen über verschiedene Speicherkonzepte aneignen können. Zum anderen wird ein Projektbericht die vorgestellten Konzepte zusammenstellen, vergleichen und aus archivischer Sicht analysieren, um so den Verantwortlichen in den Kantonen als Entscheidungsunterstützung zu dienen. Besonders wichtig scheint uns zudem zu sein, dass mit der Untersuchung dieser Fragestellung ein erster Schritt zu einem dauerhaften, ununterbrochenen Engagement getan wird. Es ist unbestritten, dass digitale Archivalien konstante Betreuung verlangen. Wenn wir uns also mit der Implementation einer Speicherinfrastruktur befassen, müssen wir das mit Blick auf ein dauerndes Engagement in einem langen Zeithorizont tun.

Die Tatsache, dass wir die Speicherung in diesem Projekt separat untersuchen, soll nicht so verstanden werden, dass wir die Speicherung losgelöst vom übrigen Kontext sehen. Wir wollen vielmehr die vorgestellten Speicherkonzepte aus der Sicht der archivischen Anforderungen betrachten und die Einbettung der Speicherlösung in das Gesamtkonzept der Langzeitarchivierung im Auge behalten. Wir sind uns bewusst,

dass es nicht *eine* richtige Lösung für das Problem gibt, sondern dass es für jedes Archiv mehr oder weniger adäquate Lösungen geben wird

1.2 Ziel der Einführungsveranstaltung

Das Vorgehen ist also folgendes: In der Einführungsveranstaltung erarbeiten wir die Fragestellungen für die fünf thematischen Veranstaltungen. Wir müssen uns über unsere Bedürfnisse als Archive im Klaren sein; zudem müssen wir die vorgestellten Speicherkonzepte in einigermaßen strukturierter Weise analysieren können, damit wir sie miteinander vergleichen können. Die KOST schlägt vor, diese Analyse jeweils in vier Gesichtspunkte zu gliedern:

- Welche Anforderungen der Archive an die Speicherung werden erfüllt?
- Wie funktioniert die Lösung technisch?
- Welche organisatorischen und rechtlichen Implikationen bestehen?
- Welche Kosten verursacht die Lösung?

Dies ist deshalb zugleich die Struktur der Einführungsveranstaltung.

1.3 Dokumentation des Kolloquiums

Die KOST wird Unterlagen zu den einzelnen Veranstaltungen jeweils den Anwesenden zusenden sowie auf ihrer Website veröffentlichen. Dabei wird es sich jeweils um eine Zusammenfassung der Diskussion handeln, ergänzt durch die Präsentationen der ReferentInnen. Am Ende wird ein Schlussbericht die Erkenntnisse zusammenfassen.

2 Archivische Anforderungen

2.1 Grundsatz

Die Anforderungen, die das Archiv an die Speicherinfrastruktur stellt, lassen sich in einem Satz zusammenfassen: Die Speicherinfrastruktur muss Dateien entgegennehmen, ihnen einen eindeutigen und fixen Identifikator zuweisen und sie gegen Vorweisen dieses Identifikators jederzeit unverändert wieder ausliefern. Man kann zuspitzend anfügen: Wie sie das macht, kann dem Archiv egal sein, solange sie es macht.

Um verschiedene Speicherkonzepte analysieren und vergleichen zu können, müssen diese Anforderungen in zwei Hinsichten detailliert formuliert werden. Einerseits ergeben sich Anforderungen aus dem archivischen Betrieb im weitesten Sinn. Andererseits soll die Infrastruktur Anforderungen erfüllen, die zu unserem Vertrauen in ihre Langzeitarchivtauglichkeit beitragen.

2.2 Exkurs zur Vertrauenswürdigkeit

Die Frage der Vertrauenswürdigkeit ist auf einem umfassenderen Niveau, für digitale Archive, bereits mehrfach analysiert worden. Verschiedene Projekte haben entsprechende Kriterienkataloge und Checklisten erarbeitet. Zu erwähnen sind hier die Publikationen von RLG¹ und des nestor-Projekts². Vertrauenswürdigkeit wird in diesen Katalogen als Kategorie verstanden, die auf eine gesamte Institution angewendet werden muss. So gliedert der nestor-Katalog seine Kriterien in drei Gruppen zum organisatorischen Rahmen, zum Umgang mit Objekten und zur Infrastruktur und Sicherheit. Der Katalog soll es einerseits Archiven (und verwandten Organisationen) erlauben, ihre eigenen Anstrengungen zu evaluieren oder evaluieren zu lassen. Andererseits soll er den Besitzern von archivischen Beständen eine Wegleitung in die Hand geben, welchen Institutionen sie diese zur Aufbewahrung anvertrauen können.

Unser Projekt hat einen engeren Fokus, nämlich allein die Speicherinfrastruktur. Die Frage stellt sich jedoch ähnlich: Können wir Kriterien festlegen, an Hand derer wir die Vertrauenswürdigkeit verschiedener Speicherinfrastrukturen bewerten können?

2.3 Gliederung der Anforderungen

Zum besseren Verständnis gliedern wir die Anforderungen an Hand zweier Unterscheidungskriterien:

1. *funktionale und nicht-funktionale Anforderungen:*
 - Was muss die Speicherinfrastruktur können (funktionale Anforderungen)?
 - Wie muss die Speicherinfrastruktur beschaffen sein (nicht-funktionale Anforderungen)?
2. *archivbetriebsspezifische und langzeitarchivierungsspezifische Anforderungen:*

¹ *Trusted Digital Repositories. Attributes and Responsibilities.* An RLG-OCLC Report (2002). <http://www.rlg.org/legacy/longterm/repositories.pdf>

² *Kriterienkatalog vertrauenswürdige digitale Langzeitarchive* (2006). <http://nbn-resolving.de/urn:nbn:de:0008-2006060710>

- Welche Anforderungen muss die Infrastruktur erfüllen, um den geordneten Betrieb des Archivs (Übernahme, Verwaltung, Benutzung etc.) zu erfüllen (archivbetriebsspezifische Anforderungen)?
- Welche Anforderungen muss die Infrastruktur erfüllen, damit wir Vertrauen haben können in ihre Eignung für die Langzeitarchivierung (langzeitarchivierungsspezifische Anforderungen)?

2.4 Anforderungskatalog

Die Anforderungen wurden an der Einführungsveranstaltung brainstorming-mässig zusammengetragen und von der KOST anschliessend geordnet.

(aSI = archivtaugliche Speicherinfrastruktur)

	archivbetriebsspezifisch	langzeitarchivierungsspezifisch
funktional	<ul style="list-style-type: none"> • Die aSI muss eine Zugriffszeit auf die Archivalien in der maximalen Grössenordnung von Stunden bieten. • Die aSI muss nur zu bestimmten Zeiten verfügbar sein. • Die aSI muss jeder Datei einen eindeutigen Schlüssel geben, aber auch existierende Schlüssel annehmen können. 	<ul style="list-style-type: none"> • Die aSI soll Mechanismen zur Integritätsprüfung der archivierten Dateien und Monitoringtools anbieten oder deren Einsatz ermöglichen.
nicht-funktional	<ul style="list-style-type: none"> • Die aSI muss im Rahmen der Entwicklung des Speicherplatzbedarfs (Grössenordnung 1-10 TB für die nächsten 5 Jahre) zu linearen Kosten skalierbar sein. • Die Limiten für Datengrössen müssen genügend hoch sein. • Zusatzfunktionalitäten dürfen die Grundfunktionalitäten nicht beeinträchtigen. 	<ul style="list-style-type: none"> • Die aSI muss eine kontrollierte Redundanz anbieten oder ermöglichen (3 Kopien von jedem Archivobjekt) • Die aSI soll eine Politik weniger Speichermigrationen ermöglichen, um die damit verbundenen Risiken und Kosten zu senken (Einsatzzeit 5-7 Jahre). • Die Speicherlogik soll nicht an die Hardware gebunden sein. • Zwischen der Hardware und der Speicherlogik soll eine offene Schnittstelle existieren. • Der Zugang zur aSI muss kontrolliert sein. • Allfällige Abhängigkeiten von weiterer Software oder von Herstellern müssen offen gelegt sein.

3 Technische Aspekte

Bei der Bereitstellung einer langzeitauglichen Speicherinfrastruktur stellt sich die Frage nach dem technischen Funktionieren.

Zuerst eine Begriffsklärung: Wir verstehen hier unter Speicherinfrastruktur nicht nur die *Speicherhardware*, sondern auch die *Storage Management Software* gemäss der untenstehenden Tabelle. Diese beiden Schichten sind für die *Persistenz*, das heisst das langfristige Fortbestehen des zu archivierenden Bitstromes zuständig.

Datensicherheit bedeutet, dass die Verarbeitung, Speicherung und Kommunikation so gestaltet sind, dass die Vertraulichkeit, Verfügbarkeit und Integrität des archivierten Bitstromes in ausreichendem Maß sichergestellt ist.

Unter *Datensicherung (Backup)* versteht man das teilweise oder gesamte Kopieren der gespeicherten Daten auf ein alternatives Speichermedium. Datensicherung ist ein Teilaspekt der Datensicherheit, eine mögliche Strategie, die Datensicherheit zu erhöhen.

Die *Storage Hardware* ist die klassische Speicherinfrastruktur: Server, Plattenspeicher, Bandspeicher etc.

Die *Storage Management Software* stellt den Speicherplatz der Hardware in für die verwendete Applikation geeigneter Form dar und übernimmt die notwendigen Verwaltungsaspekte. Durch dieses Software-Layer wird der Speicherplatz für die Benutzung vollständig transparent; das heisst, sämtliche technischen Hardwareaspekte sind bei der Benutzung unsichtbar. Das *Storage Management Layer* stellt als einziges Objekt die Datei mit einem *Locator* (Dateiname und Pfad) und einem Set von Metadaten (Zugriffsrechte, Speicherdatum etc.) zur Verfügung. Durch diese Virtualisierung kann die *Storage Hardware* ohne Einfluss auf die Anwendung organisiert, erneuert oder aber auch ausgelagert werden.

Layer	Funktion und Beschreibung
Archival Management Software	Findmittel, Archivinformationssystem, Katalog: erlaubt den Zugriff auf die archivierten Daten
Digital Archiving Software	Entgegennehmen, Speichern, Erhalten und Ausliefern von digitalen Objekten nach OAIS
Storage Management Software	Verwaltung des Speicherplatzes
Storage Hardware	Physisches Speichermedium

Wenn wir also von *Speicherplatz* reden, meinen wir heute eigentlich immer Speicherplatz unabhängig von der darunter liegenden *Storage Hardware*. Der Grund dafür ist, dass der heute benötigte Speicherplatz schon seit einiger Zeit weit grösser ist als die Speicherkapazität eines einzelnen Speichermediums. Speicherplatz ist deshalb heute eher als eine virtuelle Ressource zu verstehen, die für die Archivierung völlig transparent zur Verfügung gestellt werden kann (wie Strom aus der Steckdose).

Wenn der Aspekt der Datensicherung ebenfalls durch die *Storage Management Software* gewährleistet wird, reden wir von *Managed Storage*.

Weil wir als Archiv mit unserem Anspruch einer Langzeitarchivierung und mit einer atypischen Benutzung andere Anforderungen an die Speicherbewirtschaftung stellen als die meisten Benutzer, ist es sinnvoll und notwendig, uns ein genaueres Bild von den Charakteristika der verschiedenen heute möglichen Lösungen zu machen. Wir können die Frage der Datensicherheit gerade im Hinblick auf die zeitliche Entwicklung besser beurteilen und wir können uns auch eine bessere Kostentransparenz verschaffen, wenn wir wissen, wie Speicherplatz der eigentlichen Archiv-Software zur Verfügung gestellt werden kann.

3.1 Speicherplatz mieten

Der Speicherplatz wird, abgesichert durch langfristige Verträge, bei einem staatlichen oder privaten Anbieter gemietet. Der Anbieter übernimmt alle notwendigen Vorkehrungen für die Erhaltung des Bitstromes. Das ist die klassische Form von *Managed Storage*.

Da wir dem Speicherplatzanbieter alle technischen Aspekte überlassen, ist das Vertrauen in den Anbieter die grundlegende Grösse für die Datensicherheit. Dieses Vertrauen kann vertraglich abgesichert werden und stützt sich andererseits auch auf ein Vertrauen in den Anbieter an sich (Grösse, Marktführerschaft etc). Wir können zudem vom Anbieter eine gewisse Offenlegung bis hin zur Zertifizierung seiner Prozesse verlangen, um auf diesem Wege das Vertrauen in die Datensicherheit herzustellen.

Grundsätzlich ist es aber das Kerngeschäft des *Managed Storage Providers*, im Hinblick auf die getroffene *Service-Level-Vereinbarung* hinreichend sicheren Speicherplatz zu möglichst günstigem Preis anzubieten, und er wird uns nur bedingt Einblick in sein Geschäfts-Know-how gewähren.

Zudem hat die kurze Geschichte der Informatik bereits den Absturz der ersten Anbieter von Speicherplatz (unter dem Slogan "Speicherplatz aus der Steckdose") Ende 2000 erlebt. Die grossen Anbieter haben sich dann alle aus diesem Geschäftsfeld zurückgezogen.

3.2 Speicherzentrum

Das Archiv oder eine Gruppe von Archiven zusammen betreibt ein eigenes Archivrechenzentrum und baut dort das entsprechende Informatik-Know-how auf. Hierbei sind im Prinzip alle Prozesse und technischen Aspekte bekannt und dokumentierbar. Audits, Qualitätsmanagement und Zertifizierung sind auf jeder Stufe möglich.

Die Referenten werden uns von ihrer Erfahrung beim Aufbau und Betrieb eines Rechenzentrums im Bereich Bibliotheken und Universitäten berichten.

Der Betrieb eines eigenen Rechenzentrums mit den entsprechenden Qualitätsanforderungen ist aber erst ab einer gewissen Grösse (Speichervolumen und finanzielle Ausstattung) sinnvoll. In einem kleinen Rechenzentrum kann das technische und organisatorische Wissen nicht in entsprechender Tiefe und Breite aufgebaut werden. Es besteht zudem die Gefahr, dass aus Kostengründen mit veralteter Technologie gearbeitet werden muss.

3.3 Blackbox

Im Archiv wird eine Speicherbox ohne den Aufbau eines eigenen Rechenzentrums betrieben. Die Vorkehrungen für die Erhaltung des Bitstromes werden hardwareseitig als gegeben vorausgesetzt. Die Speicherbox ist die technische Antwort auf die Probleme, die sich beim Speicherplatz Mieten oder beim Betreiben eines eigenen Speicherzentrums ergeben.

Einerseits kann eine Speicherbox eingesetzt werden, wenn die kritische Grösse für ein eigenes Rechenzentrum noch nicht gegeben ist (heute bei einem Datenvolumen unter 50 TB). Andererseits entfällt die langfristige Bindung an einen Anbieter von *Managed Storage*, die Daten bleiben "im Haus" und unter direkter Kontrolle.

Die Speicherbox ist im Prinzip ein zertifiziertes, kleines Speicherzentrum mit einer einfachen Administration, die ganze Hardware (in der Regel *Disk Arrays*, aber auch *Tape Libraries*) ist in einer Box untergebracht. Technisches Know-how ist zum Betrieb in der Regel nicht notwendig oder entspricht etwa dem zum Betrieb eines grossen Drucker-/Kopierers. Die eingesetzte Hardware und integrierte Software garantieren einen Schutz vor Datenverlust, Überschreiben und Löschen. Durch eine einfache Administrationssoftware ist auch eine kleine Institution (aus Sicht der Anbieter die KMU) in der Lage, diese Lösung zu betreiben. Im Gegensatz zur einfacheren *File Server*-Lösung kommuniziert der Betreiber nur mit dem *Storage Management Layer*. Sämtliche Fragen der Datensicherheit sind für ihn in der Speicherbox bereits gelöst.

Im Gegensatz zum Mieten von Speicherplatz und zu einem eigenen Speicherzentrum sind die in der Speicherbox eingesetzte Technologie und die Überwachungsprozesse in der Regel dem Betreiber nicht bekannt. Die Speicherbox ist eine *Black Box*, in der Dateien abgelegt und wieder abgerufen werden können. Solange es nicht zu einem physischen Verlust der Box kommt, sind die Daten sicher.

3.4 Datenauslagerung

Die digitalen Daten werden auf ein langzeitstabiles Speichermedium kopiert und in dieser Form ausgelagert. Primär geht es bei der Datenauslagerung um die Sicherung von Daten ausserhalb der Onlineumgebung des Archivs. Das kann eine Datensicherung auf ein anderes Speichermedium im Speicherzentrum sein oder eine physische Auslagerung einer Kopie des Datenbestandes.

Auf dem ausgelagerten Medium entfallen im Prinzip die Kosten für die Datenpflege, es fallen nur reine Lagerkosten für das Medium an. Die Zugriffszeit kann aber je nach Auslagerungsort beträchtlich sein. Es muss zudem über die Zeit der Auslagerung ein Lesegerät für die ausgelagerten Daten bereitgehalten werden.

Sind die ausgelagerten Daten eine Kopie der online verfügbaren Daten, steigt die Datensicherheit erheblich, der Kostenvorteil der Auslagerung entfällt dann aber vollständig.

Das Speichermedium der Wahl können Magnetbänder oder Strichcode auf Mikrofilm sein. Magnetbänder haben eine grosse Verbreitung, eine technische Reife und Standardisierung (*Linear Tape Open*, kurz LTO) erreicht, können aber ohne Pflege nicht allzu lange gelagert werden. Bei Strichcode auf Mikrofilm ist die Standardisierung erst

in den Anfängen und die Erfahrungen damit sind gering, grundsätzlich ist hier aber die Lagerdauer ohne Pflege, das heisst ohne Umkopieren der Daten, beträchtlich.

3.5 Speichern im Netzwerk

Die Datenspeicherung erfolgt nicht lokal im Archiv oder in einem Rechenzentrum, sondern an mehreren durch ein Netzwerk verbundenen Stellen. Die Idee dabei ist, dass die Datensicherheit nicht durch Massnahmen am Speicherort gewährleistet wird, sondern durch die mehrfache Speicherung. Durch den periodischen Zugriff auf die Daten kann ein Datenverlust an einem Speicherort schnell kompensiert und korrigiert werden. Damit können äusserst günstige Speicherlösungen, die nur eine bescheidene Sicherheit und Zuverlässigkeit gewährleisten, zum Einsatz kommen. Im Prinzip kann das Netzwerk lokal sein und eine Menge von ans Netzwerk angeschlossenen Massenspeichereinheiten umfassen (*Network Attached Storage*, kurz NAS) oder aber aus Knoten im Internet bestehen. Eine Verwaltungssoftware kontrolliert die Integrität der Kopien und der Speicherknoten und führt automatisch Reparaturen an beschädigten Dateien durch.

Dieser Ansatz ist heute im Archivbereich noch sehr experimentell, hat aber ein nicht zu unterschätzendes Entwicklungspotential, da einerseits Speicherplatz im Internet als günstig verfügbare Ressource angeboten wird, andererseits die Preise für billige Speicherlösungen sich rasant nach unten entwickeln.

4 Organisatorische und rechtliche Fragen

Bei der Beschaffung einer archivtauglichen Speicherinfrastruktur müssen neben den archivischen und technischen auch organisatorische und rechtliche Aspekte beachtet werden. Im Rahmen der Auslegeordnung dieses Kolloquiums müssen diese Fragen zwar noch nicht geklärt werden. Sie sind aber in der Diskussion der verschiedenen Konzepte zu stellen und zu untersuchen.

4.1 Organisatorische Fragen

Ein wichtiger Punkt bei der Implementierung einer Speicherinfrastruktur werden Modelle der Zusammenarbeit sein. Wir gehen davon aus, dass alle Staatsarchive in absehbarer Zeit (in den nächsten fünf Jahren) eine archivtaugliche Speicherinfrastruktur benötigen werden. Wir gehen weiter davon aus, dass allenfalls wenige Archive diese in eigener Verantwortung aufbauen. Es bieten sich verschiedene Modelle zur Zusammenarbeit an:

- Der kantonale Informatikprovider stellt eine archivtaugliche Speicherinfrastruktur zur Verfügung.
- Ein Staatsarchiv mietet sich bei einem anderen Anbieter ein. Das kann ein anderes Archiv sein oder ein kommerzieller Anbieter.
- Mehrere Archive betreiben eine gemeinsame Infrastruktur.
- etc.

Diese verschiedenen Modelle stellen unterschiedliche Anforderungen an die verwendeten Lösungen. Wichtig sind vor allem die Mandantenfähigkeit und Fragen zum Datenschutz.

Als mandantenfähig wird eine Lösung bezeichnet, die auf demselben Server oder demselben Software-System mehrere Mandanten, also Kunden oder Auftraggeber, bedienen kann, ohne dass diese gegenseitigen Einblick in ihre Daten, Benutzerverwaltung und ähnliches haben. Wenn mehrere Archive also eine gemeinsame Infrastruktur betreiben, oder wenn ein Archiv seine Infrastruktur vermietet, muss diese mandantenfähig sein.

Datenschutzfragen müssen verstärkt beachtet werden, sobald ein Archiv seine Archivalien nicht in eigener Verantwortung und Umgebung aufbewahrt.

In der Diskussion wird darauf hingewiesen, dass eine Zusammenarbeit von Archiven und Bibliotheken politisch gewollt sein kann; als Beispiel wird die frühere gemeinsame Speicherlösung von Bundesarchiv und Nationalbibliothek erwähnt. Dabei muss allerdings beachtet werden, dass die Anforderungen von Bibliotheken an die Datensicherheit und an die Zugriffsperformanz unterschiedlich sein können. Inwieweit eine gemeinsame Plattform diese Unterschiede meistern kann, muss im Einzelfall analysiert werden. Gleiches gilt sinngemäss für alle Formen von geteilter Infrastruktur.

4.2 Rechtliche Fragen

Bei der Diskussion von Speicherkonzepten ist auf potentielle rechtliche Probleme zu achten: Sind das gewählte Konzept und die gewählte Organisationsform mit dem nor-

mativen Umfeld kompatibel? Dabei müssen sowohl die kantonale Archivgesetzgebung als auch die Datenschutzgesetze berücksichtigt werden.

Grundsätzlich soll auf jeden Fall von den archivischen Anforderungen und technischen Möglichkeiten her gedacht werden. Es macht keinen Sinn, gewisse Lösungsvarianten schon zum Vornherein aus rechtlichen Gründen auszuschliessen; vielmehr soll nach der Identifikation der geeigneten Lösung abgeklärt werden, ob dadurch Gesetzgebungsbedarf entsteht, und sollen entsprechende Schritte eingeleitet werden.

4.2.1 Archivgesetzliche Grundlagen

In einigen Archivgesetzen sind Passagen enthalten wie diejenige im Liechtensteinischen Archivgesetz³, Art. 5 Abs. 1: „Staatliches Archivgut wird im Landesarchiv verwahrt.“ oder im Archivgesetz des Kantons Basel-Stadt⁴ (§6, Ziff. 1): „Staatliches Archivgut wird im Staatsarchiv verwahrt.“ Es ist also von den Rechtsdiensten zu prüfen, ob eine eventuelle Speicherung ausserhalb des Archivs dieser Bestimmung gerecht wird. Allerdings werden diese und ähnliche Bestimmungen zum Teil im gleichen Gesetz relativiert. Die Kolloquiumsteilnehmer vermuten, dass sie für die digitale Archivierung wenig problematisch sind.

4.2.2 Datenschutz

Die kantonalen gesetzlichen Grundlagen für den Datenschutz machen teils präzise Vorgaben zum Umgang mit sensiblen Daten. Es ist zu prüfen, ob dadurch einzelne Lösungen verunmöglicht werden. Zu beachten ist dabei auch, dass gewisse Speicherkonzepte (wie das Mieten von Speicherplatz) zwar von den entsprechenden Gesetzen gedeckt sein, aber auf politischen Widerstand stossen können.

³ http://www.llv.li/rss/pdf-llv-la-archiv_g.pdf

⁴ <http://www.staatsarchiv.bs.ch/archivgesetz.pdf>

5 Kostenaspekte digitaler Archivierung

Das Erhalten von digitalen Archivdaten bedingt, im Gegensatz zu den klassischen Archivalien, nicht nur die Erschliessung und Aufbewahrung, sondern auch eine fortlaufende Pflege der Daten und der Infrastruktur, in der diese Daten erst überhaupt existieren. Das heisst, die Integrität der digitalen Archivdaten muss fortlaufend überwacht werden. Die Geräte, auf denen die Daten gespeichert sind, und die notwendige Infrastruktur zum Speichern und Anzeigen der Daten müssen gewartet und in relativ kurzen Abständen erneuert werden. Die Daten selber müssen in neue Formate migriert werden. Alle diese Pflgetätigkeiten, ohne welche digitale Daten sehr schnell unwiederbringlich verloren sind, bedingen einen dauernden Investitionsbedarf und Personaleinsatz.

Bei der Planung einer langzeittauglichen Speicherinfrastruktur müssen wir also nicht nur die Investitionskosten, sondern auch die fortlaufenden Kosten für Unterhalt und Pflege garantieren können, da sonst die bereits gesammelten digitalen Daten wieder verloren gehen.

5.1 Speicherbewirtschaftung / Lifecycle Management

Für die Kostenfrage bei der Speicherbewirtschaftung ist die Nutzung der gespeicherten Daten und das Lifecycle-Management der Daten von grosser Bedeutung. Die Informatik unterscheidet üblicherweise drei Kategorien, *Online*, *Nearline* und *Farline* Speicherung, im Bezug auf Zugriff und Kosten.

Die Benutzer von grossen digitalen Speicherlösungen (Electronic Records Management, kurz ERM), wie die Finanzindustrie, Telekom-Firmen etc., haben eine relativ gleich bleibende Menge von täglichen Neuzugängen, eine hohes tägliches Transaktionsaufkommen und eine abnehmende Zugriffshäufigkeit und Interesse in Relation zum Alter der Daten. Daraus ergeben sich folgende Anforderungen: Schnelle Verfügbarkeit aktueller Daten bei grosser Transaktionssicherheit, abnehmendes Interesse an Zugriffsgeschwindigkeit und Datensicherheit bei zunehmendem Alter der Daten bis hin zur Inkaufnahme von Datenverlust nach Ablauf der gesetzlichen Aufbewahrungsfristen.

Die Speicherlösungen auf dem Markt sind an diese Anforderungen angepasst:

Online-Speicherung der operativen Daten auf Harddisk mit Datensicherung (Backup) in kurzen Intervallen, um keine Transaktionen zu verlieren.

Nearline-Speicherung der älteren Daten auf Bändern oder WORMs ohne Backup, da hier keine Transaktionen mehr stattfinden.

Farline als letzter Schritt, die Auslagerung der Daten auf Bänder, wobei die notwendigen Lesegeräte meist zusammen mit dem Ende des Lifecycles der Daten nicht mehr verfügbar sind.

Im Archiv sehen die Speichernutzung und das Daten-Lifecycle-Management völlig anders aus. Wir haben periodisch grosse Zugänge, keine Transaktionen, Zugriffe auf die Daten völlig unabhängig von deren Alter. Lange Zugriffszeiten sind vertretbar und relativ grosse Systemunterbrüche können akzeptiert werden. Hingegen muss eine gleich

bleibende Datensicherheit unabhängig vom Alter der Daten gewährleistet werden können. Ein mehr oder weniger geplanter Datenverlust am Ende des Lifecycles der Daten ist völlig unakzeptabel, da es dieses Ende nicht gibt. Ein Speichersystem für Archive muss also nur teilweise den Anforderungen einer ERM-Lösung genügen und kann damit in gewissen Bereichen kostengünstiger angelegt sein. Datenverlust bei langer Auslagerung darf aber auf keinen Fall eintreten.

Fassen wir zusammen:

- Ein Archivsystem muss nicht hochverfügbar sein, relativ lange Ausfallfristen können toleriert werden.
- *Nearline*-Speicherung ist akzeptabel (Zugriff ~ 1 min), wenn damit eine Kostenersparnis verbunden ist.
- Datensicherung (*Backup*) ist nur bei Neuzugängen oder Migrationen notwendig.
- Die Ansprüche an die Datensicherheit verlangen, dass alle Daten, auch nicht benutzte Datenbestände, regelmässig auf ihre Integrität überprüft werden müssen.
- Da der Zugriff auf die Daten völlig arbiträr erfolgt, fällt eine *Farline*-Speicherung praktisch ausser Betracht.

5.2 Kostenvergleich digitales Speichersystem – Papiermagazin

Aus Sicht eines Archivs interessiert uns der Kostenvergleich zwischen dem Betrieb eines klassischen Papiermagazins und einem digitalen Speichersystem. Wenn wir in einer Annäherung davon ausgehen, dass die Kosten für die vorarchivische Bewertung, die Übernahme und die Benutzerbetreuung in etwa gleich bleiben, unterscheiden sich die beiden Arten der Archivierung durch die Kosten für die Lagerung des Archivgutes, beziehungsweise das Speichern der digitalen Daten. Wir nehmen in diesem Zusammenhang an, dass der physische Zugriff auf das Archivgut ein vernachlässigbar kleiner Aufwand bei der Gesamtbetreuung der Benutzer ist und durch den Aufwand der Datenerhaltung im digitalen Magazin wettgemacht wird.

Für unsere Rechnung gehen wir von folgenden Annahmen aus:

Die Miete für ein Lager von 10'000 m² verursacht Vollkosten von 250'000.- pro Jahr. Dieses Lager fasst etwa 10 Lkm Papierakten

Wir verwenden folgende Umrechnungsformel:

10 Lkm Papierarchivgut ≙ 3 TB Digitale Daten

(1 Dokument à 4 Seiten ≙ 50 KB → 1 Laufmeter (6000 Seiten) ≙ 300 MB)

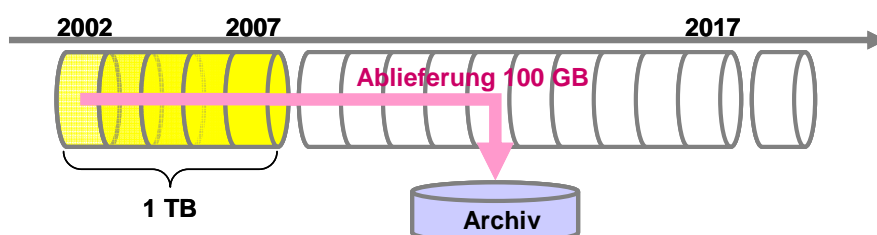
Die gleiche Datenmenge kostet nach der Vollkostenrechnung einer durchschnittlichen kantonalen Informatikdienststelle zwischen 90'000.- und 180'000.- pro Jahr (30 – 60 CHF/Jahr und GB *nearline* bzw. *online* Speicherung bei Aufbewahrung für 10 Jahre).

Mieten wir bei derselben Dienststelle einen bzw. mehrere Server (auch hier eine Vollkostenrechnung, aber ohne Datensicherung), kosten uns die 3 TB etwa 70'000.-.

Aktendepot (10 Lkm)	Speicherzentrum (3TB)
250'000.-	70'000.- bis 180'000.-

5.3 Datenmenge

Für unsere Kostenüberlegungen ebenfalls von Bedeutung ist die Frage, mit welchen Mengen von digitalen Daten wir in absehbarer Zeit rechnen müssen. Hier haben wir eine sehr schlechte Datenlage, praktisch keine Verwaltung und kein Archiv kann darauf fundiert antworten. Wir können uns also nur auf Hochrechnungen stützen. Die wenigsten kantonalen Verwaltungen haben heute mehr als 1 TB Daten online. Dieses eine TB repräsentiert wahrscheinlich etwa 5 Jahre digitale Aktenführung. Wir hätten also 2 TB in zehn Jahren ohne eine irgendwie geartete archivische Bewertung. Wenn wir nun grosszügig die Hälfte der angesammelten Unterlagen als archivwürdig bewerten, ist heute für ein Jahr mit 100 GB zu rechnen. Auch wenn wir mit einer gewissen Wachstumsrate in der Datenproduktion rechnen, sind wir für Papierakten mit einer Systemauslegung von 1TB bis 10 TB gut bedient.



Diese Annahme ändert sich jedoch sofort entscheidend, wenn wir Audio- oder Videodaten (z.B. Sitzungsaufzeichnungen) archivieren müssen⁵. In diesem Falle sind die jährlichen Datenmengen aber relativ einfach an der Quelle zu bestimmen.

5.4 Zusammenfassung

Wenn wir die Kosten der digitalen Archivierung mit den Kosten eines Papierarchivs vergleichen, sehen wir, dass die digitale Archivierung keinesfalls teurer ist als das bereits vorhandene Magazin. Es handelt sich aber um Kosten, die zum bereits vorhandenen Magazin hinzukommen und jedes Jahr anfallen werden. Wir haben auch keinerlei Anpassungen an archivische Bedürfnisse bei der Speicherstrategie vorgenommen, sondern die Kosten mit IT-Archivlösungen allgemeinsten Art verglichen.

Es lassen sich verschiedene kritische Einwände anbringen:

- 3 TB sind im Vergleich mit Multimedia-Archiven ein recht bescheidenes Archiv.
- Eine 2 TB NAS Festplatte kostet heute nur noch \$ 999⁶.
- Des Weiteren gibt es Web-Hosting-Anbieter, die 1 TB Speicherplatz für \$ 6.95/Monat anbieten⁷.

Wir werden die Kostenfrage bei allen betrachteten Lösungen genau im Auge behalten müssen.

⁵ So ergeben die Audioaufzeichnungen der National- und Ständeratsdebatten, die im BAR archiviert werden, jährlich Datenmengen von 2.4 TB. In solchen Fällen, aber auch bei Digitalisierungen von Papieroriginalen, muss individuell abgewogen werden, ob die Archivierungskosten für eine andere Erscheinungsform derselben Unterlagen gerechtfertigt sind.

⁶ LaCie, <http://www.lacie.com/us/products/product.htm?pid=10647>

⁷ glob@t.com, <http://www.globat.com/products/index.htm>