

KOST.07

Kolloquium "Archivtaugliche Speicherinfrastruktur"

Abschlussveranstaltung

Inhalt

1	Einleitung	2
2	Berichterstattung über die thematischen Veranstaltungen.....	2
2.1	Speicherplatz mieten	2
2.2	Speicherzentrum	3
2.3	Blackbox	3
2.4	Datenauslagerung.....	4
2.5	Speichern im Netzwerk	5
3	Analyse an Hand der Gesichtspunkte aus der Einführungsveranstaltung	6
3.1	Archivische Anforderungen	6
3.2	Technische Aspekte.....	6
3.2.1	Innovation.....	6
3.2.2	Best Practice	7
3.2.3	Speichermedien	7
3.2.4	Datenmengen.....	7
3.3	Organisatorisch-rechtliche Aspekte	8
3.4	Kosten	8
4	Fazit.....	9
4.1	Forschungsbedarf	9
4.2	Entscheidungsmatrix.....	10
4.3	Szenarien	10
4.3.1	Speicherplatz mieten.....	10
4.3.2	Speicherbox teilen.....	11
4.3.3	Archivnetzwerk.....	11
5	Ausblick.....	12

1 Einleitung

Das KOST-Kolloquium "Archivtaugliche Speicherinfrastruktur" hat versucht, ein vertieftes Wissen über verschiedene archivtaugliche Speicherkonzepte zu vermitteln. Einleitend muss noch einmal präzisiert werden, dass es dabei einzig um die Speicherung der archivierten Unterlagen geht, also um die sogenannte *Bitstream Preservation*, nicht aber um die archivische Bewirtschaftung der Unterlagen (Metadatenverwaltung, Migrationen etc.).

Mit der Abschlussveranstaltung verfolgte die KOST nun zwei Ziele. Einerseits sollten möglichst alle Interessierten einen Überblick über die Inhalte des Kolloquiums gewinnen. Andererseits wurden Erkenntnisse und Schlussfolgerungen präsentiert, welche für die weiteren Schritte hin zu einer archivtauglichen Speicherinfrastruktur in den Staatsarchiven als Leitlinien dienen sollen.

2 Berichterstattung über die thematischen Veranstaltungen

2.1 Speicherplatz mieten

Das Archiv mietet den benötigten Speicherplatz bei einem (kommerziellen oder öffentlichen) Anbieter. Der dabei ins Auge gefasste Zeithorizont ist üblicherweise und realistischerweise etwa 5 Jahre. Der Anbieter ist für die Datenpersistenz verantwortlich; das Archiv hat damit nichts zu tun. (Es muss auch nicht unbedingt im Detail Bescheid wissen, kann dies hingegen anstreben, was allerdings aufwendig ist.) Die Anforderungen des Archivs werden in einem *Service Level Agreement* (SLA)¹ festgehalten.

Als Referenten eingeladen waren ein Leistungserbringer, COLT Telecom² (im Wesentlichen Betreiber von Datacentern), sowie ein Produktanbieter, FAST LTA³ (mit eigener Archivierungslösung basierend auf Diskcontroller; bietet in Zusammenarbeit mit Swisscom eine konkrete Lösung an). Als vermutlich relativ naheliegende Lösung sind auch kantonale Informatikdienste als Anbieter denkbar.

Die Angebote sind sehr unterschiedlich gestaltet und müssen deshalb auf die einzelnen Kunden massgeschneidert werden. Die Referenten betonen die Wichtigkeit einer individuellen Situations- und Risikoanalyse, damit ein den Anforderungen entsprechendes Produkt angeboten werden kann. Das Anforderungsprofil von Archiven deckt sich nur schlecht mit herkömmlichen Angeboten von Anbietern von Speicherplatz. Problematisch kann sein, dass die Anstrengungen für die Datenverfügbarkeit oder für tägliches Backup standardmässig im Angebot enthalten sind, für Archive jedoch nicht notwendig wären. Deshalb ist die Formulierung dieser Anforderungen und die Kostenkontrolle sehr wichtig. Zwar bieten gewisse Firmen Speicherplatz als "Click-and-go" an. Im Normalfall werden jedoch eine genaue Situationsanalyse und eine massgeschneiderte Lösung anzustreben sein. Dies verteuert den Speicherplatz bei kleinem Volumen

¹ Ein Service-Level-Agreement (SLA) oder eine Dienstgütevereinbarung (DGV) legt fest, mit welchen Leistungseigenschaften (Ausfallzeit, Reaktionszeit, Geschwindigkeit etc.) eine wiederkehrende Leistung angeboten wird.

² <http://www.colt.net/>

³ <http://www.fast.ag/>

enorm. Die Referenten empfehlen deshalb, Möglichkeiten der Zusammenarbeit zu prüfen, um gemeinsam Lösungen einzukaufen. Als Richtpreis kann von FAST LTA die Angabe von 2,70 SFr./GB/Jahr für eine mehr oder weniger archivoptimierte Lösung dienen.

2.2 Speicherzentrum

Ein Speicherzentrum ist ein Rechenzentrum, das primär Dienstleistungen im Bereich der Datenspeicherung und Langzeitarchivierung erbringt.

Zwei Rechenzentren aus dem öffentlich-rechtlichen Bereich stellten ihre Erfahrungen beim Aufbau und Betrieb dieser Dienstleistung vor: Die Referenten waren Dagmar Ullrich von der Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG)⁴ und Jürg Gemeinder vom Verwaltungsrechenzentrum St.Gallen (VRSG)⁵.

Folgendes Fazit lässt sich aus den Ausführungen und der anschliessenden Diskussion ziehen:

- Der Aufbau eines Speicherzentrums setzt ein minimales (Daten-)Volumen voraus und darf in seiner Komplexität nicht unterschätzt werden.
- Vom Neuaufbau eines Rechenzentrums raten die Referenten grundsätzlich ab. Das Mieten bestehender Rechenzentrumsinfrastruktur scheint sinnvoller. Dies ist umfassender als das, was oben unter "Speicherplatz mieten" beschrieben wurde: Während es dort einzig darum geht, eine bestimmte Menge von Speicherplatz in TB zu mieten, wobei alle Details in der Verantwortung des Anbieters sind, beinhaltet das Mieten von Rechenzentrumsinfrastruktur eine grössere Verantwortung des Mieters.
- Das Beherrschen der Hardwaremigrationszyklen ist eine der grössten Herausforderungen für ein Rechenzentrum im Bereich der Langzeitarchivierung.
- Eine klare Trennung von Speichermanagement im Speicherzentrum und Archivsoftware unter Aufsicht des Archives ist unerlässlich.

2.3 Blackbox

Unter einer Blackbox verstehen wir eine Speicherbox, d.h. ein physisch greifbares Gerät, das im Archiv selber oder bei einem Raumanbieter stehen kann. "Black" deshalb, weil die interne Funktionsweise dieser Box unbekannt und irrelevant ist. Wichtig ist, dass der Benutzer mit der Box über ein API (Application Programming Interface [Programmierschnittstelle]⁶, meist proprietär) kommunizieren kann, um Daten abzuspeichern und wieder anzufordern. Die Datenpersistenz wird vom Hersteller garantiert; vom Anwender muss alle paar Jahre nur die Migration auf eine neue Box gelöst werden.

Das Kolloquium bot dazu zwei Referate, einmal von einem Lösungsanbieter (Raphael Mettan und Christian Kübler von IBM mit der DR550⁷) und dann von einem archivi-

⁴ <http://www.gwdg.de/>

⁵ <http://www.vrsg.ch/>

⁶ Siehe <http://de.wikipedia.org/wiki/Programmierschnittstelle>.

⁷ <http://www-03.ibm.com/systems/de/storage/disk/dr/550/index.html>

schen Anwender dieses Konzepts (Ilka Stahlberg und Jörg Homberg vom Brandenburgischen Landeshauptarchiv⁸ mit der Centera von EMC⁹).

Ein Hauptvorteil der Blackbox ist der relativ geringe Administrationsaufwand im laufenden Betrieb. Nach Auskunft von Herrn Homberg beschränkt sich dieser im Normalfall auf etwa eine Minute pro Tag, um das Statusmail zu lesen; im Havariefall bietet der Anbieter umfassenden Service. Ein weiterer, vielleicht eher psychologischer Vorteil kann die physische Kontrolle über die Speicherinfrastruktur sein. Investitionen fallen schubweise immer dann an, wenn auf eine neue Blackbox migriert werden muss.

Technisch kann eine Blackbox auf verschiedenen Trägermedien beruhen: im Normalfall auf Disks; diese können aber auch ergänzt oder ersetzt werden durch eine Tape-Library. Vor allem in letzterem Fall sollte eine Datenmenge im mittleren zweistelligen TB-Bereich vorhanden sein, damit die Lösung Sinn macht. Allerdings scheinen im Licht des Anforderungsprofils des Archivs Tapes besser geeignet, insbesondere weil die hohe Verfügbarkeit und kleine Reaktionszeit von Disks nicht notwendig sind. Zu beachten ist dabei jedoch, dass die automatische Integritätskontrolle auf Disks einfacher und gründlicher machbar ist, was die Datensicherheit vergrößert.

2.4 Datenauslagerung

Die digitalen Daten werden auf ein langzeitstabiles Speichermedium umkopiert und in dieser Form aus der operativen Informatikinfrastruktur ausgelagert. Während der Phase der Auslagerung fallen nur die Kosten für die Lagerung des Trägermediums an.

Die zwei Referenten präsentierten die Lösung aus Sicht der Forschung und in der kommerziellen Anwendung: René Meier, Swiss Data Safe AG¹⁰ und Peter Fornaro, Peviar¹¹, Imaging & Media Lab (IML).

Diese Lösung wird schon länger diskutiert und eingesetzt, z.B. bei der COLD Archivierung (*Computer Output to Laser Disk*), erlangt aber mit dem Einsatz von Mikrofilm als Trägermedium eine neue Dimension bezüglich Auslagerungszeitraum (> 100 Jahre).

Mit den heutigen Methoden können auf einen 35-mm Film von 600m Länge (entspricht etwa einer Rolle von 40 cm Durchmesser) ungefähr 250 GB Nutzdaten geschrieben werden. Diese Daten lassen sich mit einem handelsüblichen Scanner und einer *Open-Source*-Software wieder zurückgewinnen. Damit entfällt die Notwendigkeit, ein spezielles Lesegerät während des ganzen Auslagerungszeitraums bereit zu halten. Metadaten zu den verwendeten Formaten (z.B. Formatspezifikationen) können in menschenlesbarer Form ebenfalls auf den Film geschrieben werden.

Eine Standardisierung, sowohl was das Format auf dem Mikrofilm betrifft, wie auch der notwendigen Prozesse, hat bis jetzt noch nicht stattgefunden. Die aktuellen Auslagerungsprojekte sind kundenspezifisch und befinden sich eigentlich noch im Experimentierstadium. Swiss Data Safe kann aber dennoch einen Richtwert von 25'000 € für die Ausbelichtung von 1 TB Daten auf Film angeben.

Beide Referenten verstehen ihre Lösung nur als Baustein einer umfassenden Speicherlösung, nämlich als Backupkopie oder Dark Archive, das nicht bewirtschaftet wer-

⁸ <http://www.landeshauptarchiv-brandenburg.de/>

⁹ <http://germany.emc.com/products/systems/centera.jsp>

¹⁰ <http://www.swissdatasafe.ch/>

¹¹ <http://www.peviar.ch/>

den muss, es den Archiven jedoch ermöglicht, für die Gebrauchskopie ein geringeres Sicherheits- und damit Preisniveau zu wählen. (Vergleichbare Lösungen existieren auch für die Ausbelichtung von digitalen Daten auf Papier.) Einzelne Teilnehmer regten in der Diskussion an, die Auslagerung auf Mikrofilm auch als Primärkopie für gewisse kaum gebrauchte Archivbestände zu erwägen. Tatsächlich ist in der Regel eine grobe Unterteilung der Archivalien machbar in Bestände, die häufig oder gelegentlich, und solche, die praktisch nie nachgefragt werden. Zu klären ist dabei, ob für letztere tatsächlich eine Zugriffszeit im Wochenbereich vertretbar ist.

2.5 Speichern im Netzwerk

Diesen Begriff haben wir verwendet für die verteilte Speicherung in Peer-to-Peer-Netzwerken. Das Prinzip ist allgemein bekannt und bewährt (beispielsweise bei Tauschbörsen im Internet)¹²; Hauptcharakteristikum ist die sehr grosse Ausfallsicherheit, die mit dem Fehlen eines zentralen Verwaltungspunktes einhergeht. Das Konzept ist im archivischen Umfeld weder bekannt noch verbreitet. Im Bibliotheksbereich wird es jedoch seit einigen Jahren erfolgreich für die Archivierung von e-journals angewendet. Eingeladen waren deshalb einerseits Simon Margulies vom Forschungsprojekt distarnet¹³ der Universität Basel (Anwendung des Prinzips auf archivische Bedürfnisse), andererseits Jan Dirk Brinksma von der Bibliothek für exakte Wissenschaften der Universität Bern¹⁴, die das LOCKSS-Konzept¹⁵ (Lots Of Copies Keep Stuff Safe) für e-journals implementiert hat.

Das Konzept beruht auf einem Verbund einzelner Computer (Server), die über ein Netzwerk (privat oder verschlüsselt über das Internet) miteinander verbunden sind. Ein Protokoll sorgt für die nötige Redundanz und Datenintegrität: Inhalte werden von einem Knoten eingespielen, durch das Protokoll kopiert und verteilt und laufend miteinander abgeglichen. Unterschiedliche Risikoanalysen führen zu unterschiedlichen Abwehrszenarien. Beide vorgestellten Projekte zielen auf einfache Administration und geringen Aufwand ab: LOCKSS z.B. kommt mit einem handelsüblichen Computer aus, auf dem die Software installiert wird.

Aus archivischer Sicht sind zwei Varianten dieses Konzepts denkbar, nämlich ein Verbund mehrerer Archive oder ein internes Netzwerk. Beide sind interessant und würden eine nähere Betrachtung rechtfertigen. Interessant sind vor allem die geografische Redundanz und die intellektuelle Kontrolle über die vom Träger getrennte Speicherlogik. Allerdings ist dieses Konzept für Staatsarchive im Moment nicht anwendungsreif: Es existiert keine Lösung, die einfach so implementiert werden könnte. Weitere Forschung scheint nötig, welche auch herkömmliche, in der Praxis bewährte P2P-Netzwerke einbeziehen sollte.

¹² Siehe dazu auch <http://de.wikipedia.org/wiki/Peer-to-Peer>.

¹³ <http://www.distarnet.ch/>

¹⁴ <http://www.bewi.unibe.ch/>

¹⁵ <http://www.lockss.org/>

3 Analyse an Hand der Gesichtspunkte aus der Einführungsveranstaltung

3.1 Archivische Anforderungen

Als zentrale Erkenntnis stellt sich heraus, dass die Archive ein relativ spezielles Anforderungsprofil an ihre Speicherinfrastruktur haben, welches von gängigen Speicherangeboten (insbesondere der Gattungen "Speicherplatz mieten" und "Blackbox") schlecht abgedeckt wird:

- Die Verfügbarkeit kann niedrig sein: In den Nächten und am Wochenende können die Daten problemlos offline sein; zwischendurch kann auch ein Tag Ausfall verkraftet werden.
- Die Zugriffszeit muss nicht optimiert werden. Archivbenutzer sind sich Zugriffszeiten von mehreren Minuten gewöhnt; eine Zugriffszeit von etwa einer Minute, wie sie Nearline-Speicherung¹⁶ bietet, ist deshalb absolut akzeptabel, Online-Zugriff ist unnötiger Luxus.
- Es gibt keine Transaktionen auf den archivierten Daten, deshalb sind die Backup-Anforderungen einfach: Nur inkrementelles Backup ist notwendig.
- Der Zugriff auf die Daten erfolgt arbiträr; ein hierarchisches Speichermodell, d.h. verschiedene Arten von Speicherplatz mit unterschiedlichen Zugriffszeiten, macht deshalb keinen Sinn.
- Datenintegrität und Datensicherheit sind absolut zentral.

Weil diese Anforderungen so speziell sind, existieren kaum Speicherkonzepte und -angebote, die darauf zugeschnitten sind. Es besteht deshalb immer die Gefahr, entweder für Leistungen zu bezahlen, die man nicht braucht, oder zu wenig Leistungen zu erhalten. Das archivische Anforderungsprofil ist so das wichtigste Werkzeug bei der Analyse und Beschaffung von Speicherlösungen¹⁷.

3.2 Technische Aspekte

3.2.1 Innovation

Betrachten wir die fünf Lösungen aus Sicht des Innovationspotentials und der vermutlichen Entwicklungsmöglichkeit, ist *Speichern im Netzwerk* wahrscheinlich die innovativste Lösung und hat in der Form der *Filesharing*-Lösungen im Internet schon sein Potential gezeigt. Es besteht aber noch keine im eigentlichen Sinne archivtaugliche Variante.

Datenauslagerung besitzt ebenfalls ein hohes Innovationspotential und könnte in standardisierter Form langfristig die Archive finanziell entlasten, indem die Bewirtschaftungskosten für Teile der Bestände gesenkt werden können.

¹⁶ Nearline-Speicherung ist zwischen Online (Sofortzugriff, hohe Kosten) und Offline (ausgelagert, günstig) angesiedelt; in der Regel handelt es sich dabei um Systeme, die auf Tape-Libraries basieren.

¹⁷ Von den Speicherplatzanbietern der ersten Veranstaltung wurde dazu die "Zauberformel der Informationssicherheit ins Spiel gebracht: CIA, kurz für confidentiality – integrity – availability (Vertraulichkeit – Integrität – Verfügbarkeit). Die Gewichtung dieser drei Aspekte charakterisiert ein Anforderungsprofil.

3.2.2 Best Practice

Speicherplatz mieten und *Blackbox*-Lösungen sind die aktuellen Angebote der Informatikindustrie:

Blackbox-Lösungen können heute als ausgereift betrachtet werden und lassen dem Archiv viel Gestaltungsfreiraum. Im Einsatz sind solche Systeme vornehmlich im KMU-Bereich, bei Firmen und Institutionen die sich auf Grund ihrer Grösse kein eigenes Rechenzentrum leisten können, dennoch aber einen digitalen Archivierungsbedarf haben.

Speicherplatz mieten verursacht wenig Aufwand im Archiv, schafft aber externe Abhängigkeiten. Diese Lösung ist erst seit kurzer Zeit wieder am Markt, nachdem die ersten Versuche in diese Richtung Ende der 1990er Jahre gescheitert waren.

Das *Speicherzentrum* (Rechenzentrum) ist die klassische Lösung für die digitale Archivierung, ohne besonderes Entwicklungspotential aber mit viel Erfahrungshintergrund. In der Regel ist es für kleinere Datenvolumen sehr teuer.

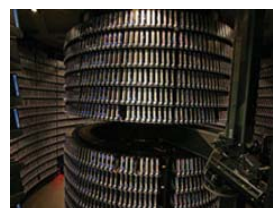
3.2.3 Speichermedien

Betrachten wir die Lösungen aus Sicht der eingesetzten Speichermedien können wir folgendes Fazit ziehen:

- Für Datenbestände von 1 bis 20 TB, wie wir sie für die nächsten fünf Jahre bei Staatsarchiven erwarten, sind *spinning Disk*¹⁸ das Medium der Wahl.
- Für Datenbestände >20 TB muss aus ökonomischen und ökologischen Überlegungen *Tape* als Medium gewählt werden.
- Mikrofilm, Rosetta Disk¹⁹, Hologramme etc. sind noch weitgehend exotische Medien für die digitale Archivierung.



EMC CLARiiON



SDSC Tim Mcnew



Imaging & Media Lab

3.2.4 Datenmengen

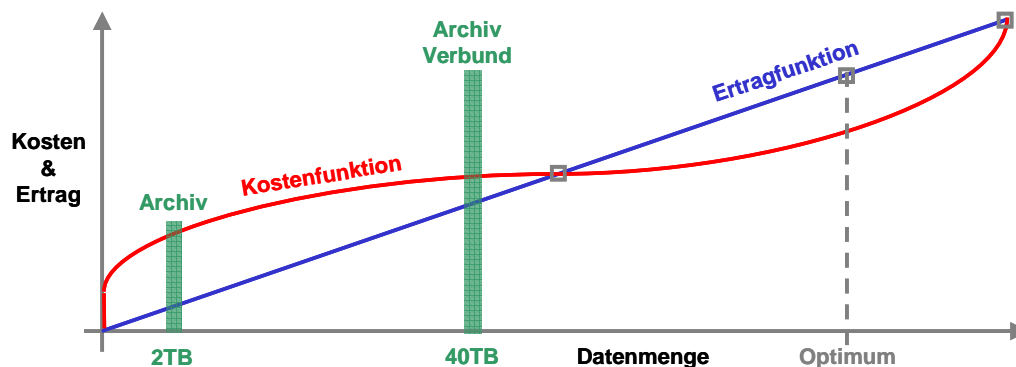
Was das Datenvolumen betrifft, welches auf die Staatsarchive zukommt, sind wir auf Schätzungen angewiesen. Wir können aber folgende Rechnung anstellen: Die wenigsten kantonalen Verwaltungen haben heute insgesamt mehr als 1 TB Daten online. Diese Daten stammen sowohl aus Geschäftsverwaltungssystemen und Fachanwendungen, vor allem aber aus Datei-Ablagen. Dieses Terabyte repräsentiert wahrscheinlich etwa 5 Jahre digitale Aktenführung und Verwaltungstätigkeit. Wenn wir weiter

¹⁸ Festplatten im Onlinebetrieb; Festplatten können im Prinzip auch als Auslagerungsmedium verwendet werden.

¹⁹ <http://www.rosettaproject.org/about-us/disk/concept>

grosszügig annehmen, dass die Hälfte der angesammelten Unterlagen als archivwürdig bewertet werden, ergibt das eine Ablieferung von 100 GB pro Jahr. Daraus folgt, dass eine Systemauslegung im Archiv von 1-10 TB vollauf genügt.

Gerade die Datenspeicherung unterliegt einem Skaleneffekt (Ertragsgesetz der Betriebswirtschaft), deshalb sind sich alle Referenten einig, dass die Langzeitarchivierung ein Gemeinschaftsprojekt der Staatsarchive sein muss:



3.3 Organisatorisch-rechtliche Aspekte

Wie bereits erwähnt ist in verschiedenen Veranstaltungen klar zum Ausdruck gekommen, dass die Datenmengen, die bei einzelnen Staatsarchiven in den nächsten zehn Jahren zu erwarten sind, unter der kritischen Masse liegen, die bei verschiedenen Lösungen für einen wirtschaftlichen Betrieb nötig wäre. Mehrere Referenten wiesen eindringlich darauf hin, dass Modelle der Zusammenarbeit, der gemeinsamen Bestellung von Leistungen geprüft werden sollten. Dies gilt selbstredend für Speicherzentren, aber ebenfalls für "Speicherplatz mieten" und sogar für "Blackbox". Aus dem Publikum wurde dazu angeregt, neben der interkantonalen auch die innerkantonale Zusammenarbeit zu prüfen, also zwischen dem Archiv und anderen kantonalen Institutionen, die ähnliche Anforderungen an die Speicherung grosser Datenmengen stellen.

Was den rechtlichen Aspekt betrifft, ist an der Einführungsveranstaltung klar geworden, dass dieser nicht das Hauptproblem sein wird. Zwar können gewisse Lösungsansätze im Konflikt mit rechtlichen Vorgaben stehen, aber es wird nicht als Problem betrachtet, diese bei Bedarf anzupassen. Aufgepasst werden muss allerdings auf möglichen politischen Widerstand gegen bestimmte Lösungen bzw. politischen Druck auf bestimmte Lösungen. Z.B. sehen sich gewisse Archive der Erwartung gegenüber, ihren Speicherbedarf beim kantonalen Informatikanbieter zu decken, womit viele der hier angestellten Überlegungen wegfallen. Zu beachten sind ferner mögliche "psychologische" Probleme, die mit der Datenspeicherung ausserhalb der physischen Kontrolle des Archivs oder der Verwaltung verbunden sind.

3.4 Kosten

Es hat sich als sehr schwierig erwiesen, vergleichbare Kostenangaben für die Langzeitarchivierung zu erhalten. Einerseits herrscht ein gewisses Zögern, die Kosten mitzuteilen; andererseits handelt es sich in vielen Fällen nicht um fixe Preislisten, sondern um individuell verhandelte Kosten.

Grundsätzlich lässt sich immerhin sagen, dass die digitale Speicherung kostengünstiger ist als konventionelle Archivierung (siehe Präsentation); wir können aber davon

ausgehen, dass die Datenmengen, die von der Verwaltung produziert werden, laufend steigen.

Ein Kostenvergleich von *Bitstream-Preservation*-Anbietern zeigt grosse Unterschiede:

- Verbindliches Angebot eines kantonalen Informatikdienstleisters (30 SFr./GB/Jahr bei 10 Jahren Laufzeit) => 30'000 SFr./TB/Jahr
- Kostenangaben BAR (30 TB für 600'000 SFr./Jahr beim BIT) => 20'000 SFr./TB/Jahr
- Kauf einer Blackbox (Pauschalpreis 100'000 €/8TB/3Jahre für Kauf und Unterhalt) => 6'850 SFr./TB/Jahr
- Kauf von zwei Tape-Systemen (2 mal Server, Tape Library, Software = 140'000 SFr. ergibt bei der Annahme von 25% Hardwarekosten an den Gesamtkosten 560'000 SFr./40TB/3Jahre) => 4'800 SFr./TB/Jahr
- Günstigster Anbieter von archivtauglichem Speicherplatz in unserem Kolloquium (0.23 SFr./GB/Monat, 5 Jahren Laufzeit) => 2'760 SFr./TB/Jahr
- Amazon S3 Speicherdienst (0.20 SFr./GB/Monat) => 2'400 Sfr./TB/Jahr

4 Fazit

Die KOST stellt einige ihrer Schlussfolgerungen aus dem Kolloquium "Archivtaugliche Speicherinfrastruktur" vor. Weitere, über die Abschlussveranstaltung hinausgehende Überlegungen folgen im Kapitel 5 unter "Ausblick".

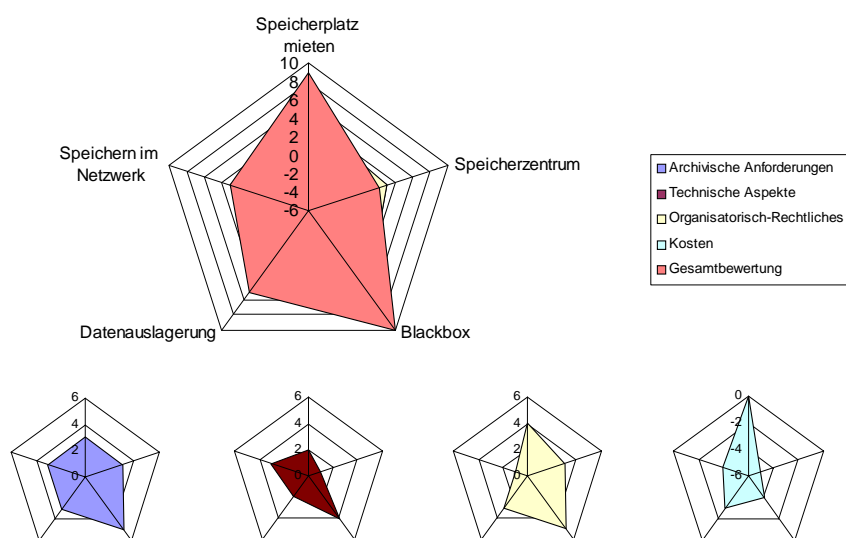
4.1 Forschungsbedarf

Die beiden im Projekt untersuchten Ansätze "Datenauslagerung" und "Speichern im Netzwerk" haben sich zwar in mancher Hinsicht als interessant erwiesen, sind aber wegen der fehlenden Reife noch nicht in den Archiven einsetzbar. Es könnte sich jedoch für die Archivwelt lohnen, sich weiter mit diesen Themen zu befassen. Dazu sind verschiedene Vorgehensweisen denkbar: Zum einen wäre ein verstärkter Kontakt interessant mit Forschungsinstitutionen (im Moment vor allem das Imaging & Media Lab der Universität Basel), die sich mit diesen Ansätzen befassen. Da das Hauptproblem für deren Einsatz in den Archiven jedoch die mangelnde Erfahrung und ungenügende Standardisierung sind, könnte vor allem ein Pilotprojekt von einem oder mehreren Archiven diese Ansätze entscheidend voranbringen.

4.2 Entscheidungsmatrix

Eine Entscheidungsmatrix vergleicht die fünf Speicherkonzepte aufgrund der besprochenen Kriterien (Skala 0-3, im Fall der Kosten invertiert):

Alternativen	Speicherplatz mieten	Speicherzentrum	Blackbox	Datenauslagerung	Speichern im Netzwerk
Archivische Anforderungen					
• Archivbetrieb	2	2	3	0	1
• Langzeitarchivierung	1	1	2	3	2
Technische Aspekte					
• Innovationspotentials	0	0	1	2	3
• Best Practice	2	1	3	0	0
Organisatorisch-Rechtliches					
• Organisatorisches	3	1	2	1	1
• Rechtliches	1	2	3	2	0
Kosten					
• Investitionen	0	-3	-3	-2	-1
• Administration	0	-2	-1	-1	-3
Bewertung	9	2	10	5	3



4.3 Szenarien

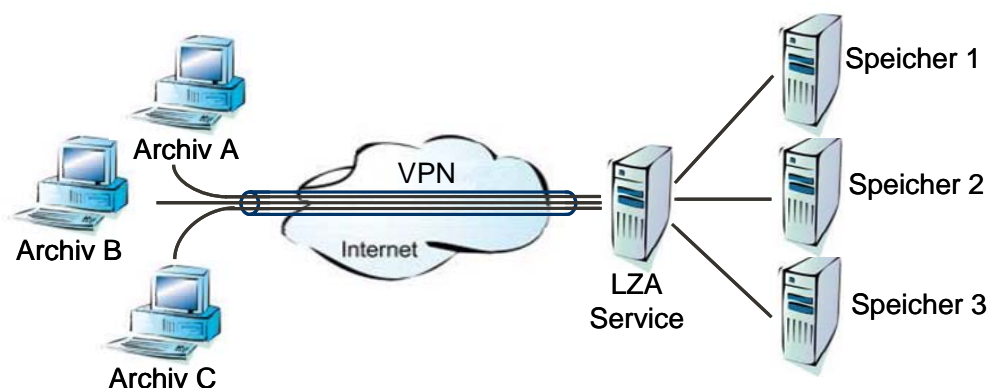
In drei Szenarien möchte die KOST vorstellen, wie eine gemeinsame Speicherstrategie der Staatsarchive auf Grund der in den fünf Kolloquien gewonnenen Einsicht aussehen könnte.

4.3.1 Speicherplatz mieten

1. Einige Archive geben der KOST den Auftrag zur Langzeitarchivierung (LZA) von minimal je einem TB. Als Zielgrösse für die Speicherkosten werden 4'500 SFr./TB/Jahr festgelegt (basierend auf der SDSC Studie²⁰, die von 1'500 SFr./TB/Jahr für eine Kopie ausgeht).

²⁰ IS&T Archiving Conference 2007: *Disk and Tape Storage Cost Models*, Richard L. Moore et al, University of California San Diego (USA)
<http://www.imaging.org/store/phypub.cfm?seriesid=28&pubid=765>

2. Die KOST spezifiziert für den LZA-Bereich eine Speicherschnittstelle (Amazon S3 API²¹ oder dCache²²) und bestimmt einen *Service Level* für den Dienst (dreifache Redundanz, periodischer Integritätscheck, Mandantenfähigkeit).
3. Die KOST holt Offerten bei kommerziellen oder öffentlich-rechtlichen Anbietern ein, allenfalls in Form einer WTO-Ausschreibung (BAR, Phonotheek, Swisscom, etc.).
4. Die KOST übernimmt Auditing und Inkasso für die gewählte Lösung.



4.3.2 Speicherbox teilen

1. Grundsätzlich gleiches Szenario wie bei Speicherplatz mieten, aber es findet sich kein Anbieter von Speicherplatz mit einer hinreichend interessanten Offerte. Mögliche Ursache dafür sind das kleine Datenvolumen oder zu spezielle *Service-Level*-Anforderungen.
2. Die KOST kauft eine Speicherinfrastruktur vom Typus Blackbox mit entsprechender Redundanz und lässt diese in einem Rechenzentrum betreiben.
3. Die KOST evaluiert Software für die Speicherschnittstelle (Amazon S3 API oder dCache), die Replikation und den Integritätscheck.
4. Die KOST übernimmt das Inkasso.

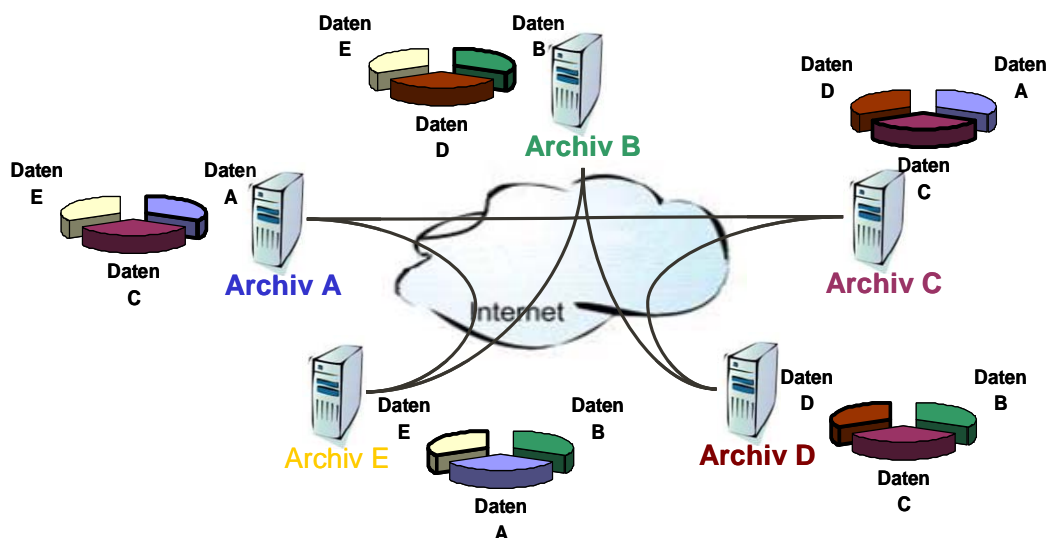
4.3.3 Archivnetzwerk

1. Die interessierten Archive kaufen je eine eigene Blackbox-Lösung und betreiben diese bei sich im Archiv oder bei ihrem jeweiligen Informatikdienstleister.
2. Die Archive stellen 2/3 ihrer Speicherkapazität anderen Archiven als *Dark Archive* zur Verfügung.
3. Die KOST evaluiert eine Softwarelösung für die Replikation der Daten von Archiv zu Archiv über VPN.

²¹ <http://docs.amazonwebservices.com/AmazonS3/2006-03-01/gsg/>

²² <http://www.dcache.org/>

4. Die KOST organisiert die Struktur des Netzwerks und eine Lösung für die sichere Übertragung der Daten via VPN.
5. Der Datenzugriff wird vertraglich zwischen den beteiligten Archiven geregelt.



5 Ausblick

Das Kolloquium hat einen Überblick über die aktuelle Lage auf dem Gebiet der archivtauglichen Speicherinfrastruktur geboten. Es ist sinnvoll, diese Erkenntnisse jetzt zu nutzen und die weiteren Schritte zu überlegen. Die KOST schlägt dazu eine Umfrage unter den Mitgliedsarchiven vor, ausgehend vom oben dargelegten Fazit. Die KOST-Archive sollen befragt werden, ob sie daran interessiert sind, sich an einer gemeinsamen Lösung gemäss Szenario 1 (Speicherplatz mieten) oder 3 (Archivnetzwerk) zu beteiligen. (Szenario 2, Speicherbox teilen, gilt dabei als möglicher Ersatz für Szenario 1.) Bei einem Interesse sollen sie auch beziffern, wie viel Speicherplatz sie in den nächsten fünf Jahren benötigen würden.

Details und Modalitäten dieser Umfrage wird die KOST mit dem Steuerungsausschuss klären. Die Ergebnisse einer solchen Umfrage werden dann gegebenenfalls in ein neues Projekt münden, das sich mit der Realisierung eines Szenarios befasst.