

KOST.07

Kolloquium „Archivtaugliche Speicherinfrastruktur“

Thema: Speicherzentrum

Inhalt

1	Einleitung	2
2	Die Referate	2
2.1	Dagmar Ullrich von der Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG)	2
2.2	Jürg Gemeinder vom Verwaltungsrechenzentrum St. Gallen (VRSG)	3
3	Diskussion.....	3
3.1	Archivische Anforderungen?	3
3.2	Technische Aspekte?	4
3.3	Kosten: Mit welcher Grössenordnung ist zu rechnen?	4
3.4	Realisierbarkeit: Ist das etwas für die Staatsarchive?	5

1 Einleitung

Die zweite Veranstaltung des KOST-Kolloquiums "Archivtaugliche Speicherinfrastruktur" befasste sich mit dem Thema "Speicherzentren".

Ein Speicherzentrum ist ein Rechenzentrum, das primär Dienstleistungen im Bereich Langzeitarchivierung erbringt. Zwei Rechenzentren aus dem öffentlich-rechtlichen Bereich wurden eingeladen, in kurzen Referaten ihre Erfahrungen vorzustellen: die Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG)¹ und das Verwaltungsrechenzentrum St. Gallen (VRSG)².

Beide Rechenzentren bieten die ganze Palette der Informatikdienstleistungen an (Datenverarbeitung, Datenspeicherung, Applikationsbetreuung und Beratung), sind also nicht nur auf Langzeitarchivierung spezialisiert, erbringen aber für ihre öffentlich-rechtlichen Auftraggeber gewisse Archivierungsdienstleistungen.

Wir erfahren in den Referaten und in der anschliessenden Diskussion, wie im Speziellen die Archivierung technisch und organisatorisch gelöst ist und welche zusätzlichen Vorkehrungen im Bereich Langzeitarchivierung getroffen worden sind.

2 Die Referate

2.1 Dagmar Ullrich von der Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG)

Die Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen betreibt das Hochschulrechenzentrum der Universität Göttingen und ein Rechenzentrum für die Max-Planck-Gesellschaft. Die Spannweite der Aufgaben reicht vom Betrieb von Hochleistungsrechnern bis zur Ausbildung von Informatikfachkräften. Im Bereich Langzeitarchivierung bietet die GWDG einerseits die grundlegende *Bitstream Preservation* für wissenschaftliche oder kulturell bedeutende nicht-reproduzierbare Daten im Terabyte-Bereich und andererseits mit dem Projekt kopal³ eine Archivierung nach OAIS-Standards an. Kopal wird von der Niedersächsischen Staats- und Universitätsbibliothek Göttingen und der Deutschen Nationalbibliothek Frankfurt genutzt.

Frau Ullrich erläutert erst den technischen und organisatorischen Hintergrund der *Bitstream Preservation* (Hardware und Prozesse). Hier handelt es sich nur teilweise um eine eigentliche Archivierung, weil mit und an den grossen wissenschaftlichen Datenbeständen der Max-Planck-Gesellschaft oft noch gearbeitet wird.

¹ <http://www.gwdg.de/>

² <http://www.vrsg.ch/>

³ <http://kopal.langzeitarchivierung.de/>
http://kopal.langzeitarchivierung.de/downloads/kopal_Demonstrator_en.swf

Anschliessend zeigt sie, wie das kopal-System aufgebaut ist: koLibRi Frontend, kopal System basierend auf DIAS⁴ und Content Manager, mit Backend Tivoli Storage Manager von IBM.

Das MPG Bitstream-Archiv umfasst im Augenblick etwa 22 Terabyte, die gesamte Tape Library der GDWD nähert sich bereits der Pentabyte Grenze.

2.2 Jürg Gemeinder vom Verwaltungsrechenzentrum St. Gallen (VRSG)

Das Verwaltungsrechenzentrum St. Gallen erbringt mit 180 Mitarbeitern Informatikdienstleistungen für Kommunen in der ganzen Ostschweiz. Dabei werden Fachapplikationen von Anlagebuchhaltung bis Zivilschutz für Gemeinden und Städte zur Verfügung gestellt. Ein Teil der Aufgabe der VRSG ist dabei auch die Archivierung der anfallenden elektronischen Belege. Herr Gemeinder zeigt in seiner Präsentation, wie die revisionssichere Archivierung mit WORM Tapes realisiert wird. Zum Einsatz kommen dabei Content Manager OnDemand⁵ von IBM und ebenfalls Tivoli Storage Manager⁶.

Die Archivierung der verschiedenen Datenströme erfolgt in unterschiedlichen Formaten:

- COLD Archivierung (elektronische Belege) AFP⁷
- Scanning (Altdatenbestände der Gemeinden) TIFF
- PC Dokumente diverse Office-Formate und PDF

Von besonderem Interesse ist in seinem Vortrag die Darlegungen zur kürzlich erfolgten Migration von Optical Disk zu WORM Tape. Eine interessante Diskussion entspinnt sich hier zu den Themen Hardware-, Medien- und Software-Migration.

3 Diskussion

3.1 Archivische Anforderungen?

Die spezifischen Eigenarten der archivischen Langzeitarchivierung, so wie wir sie im ersten Kolloquium ermittelt haben (keine Transaktionen, kein Zugriffsmuster, wenig Benutzer und Zugriffe, kein Löschen, lange Zugriffs- und Ausfallzeiten tolerierbar) werden von einem Rechenzentrum in der Regel nicht direkt widerspiegelt. Das Rechenzentrum bietet in der Regel hohe Transaktionssicherheit, schnellen Zugriff, geringe Ausfallzeiten und beschleunigt den Zugriff auf die langsamen Tapes mit Cache-Speichern. Die archivischen Anforderungen werden also übererfüllt. Frau Ullrich wendet hier ein, dass, sobald die Daten digital verfügbar sind, auch die Anforderungen bezüglich Zugriffszeiten und -häufigkeit bis hin zur gewohnten Webserver-Performance steigen werden.

An der Erfüllung der Forderung nach einem Trusted Repository⁸ wird in Göttingen mit dem Erstellen entsprechender Policies (Richtlinien) gearbeitet. Heute basiert das

⁴ <http://www-05.ibm.com/nl/dias/>

⁵ <http://www.ibm.com/software/data/content-management/products/>

⁶ <http://www.ibm.com/software/tivoli/products/storage-mgr/>

⁷ http://de.wikipedia.org/wiki/Advanced_Function_Presentation

⁸ <http://www.oclc.org/programs/ourwork/past/trustedrep/repositories.pdf>

"Vertrauen" auf Best Practice und der Zusammenarbeit mit einem entsprechend potenten kommerziellen Partner (in diesem Falle IBM).

3.2 Technische Aspekte?

Die Rechenzentren sind heute vorwiegend auf die *Bitstream Preservation* ausgerichtet und sehen dort ihre Kompetenz, das heisst der Datenowner, bzw. das Archiv ist in diesem Falle für die langfristige Interpretierbarkeit der Daten zuständig.

Unabdingbar ist heute die Trennung von Archivsoftware und Speichermanagement (in beiden vorgestellten Fällen Tivoli von IBM). Bei einer klaren Trennung können in der Speichermanagement-Schicht Medien- bzw. Hardware-Migrationen ohne Einfluss auf die Archivsoftware vorgenommen werden; das heisst, das digitale Objekt bleibt auch nach der Migration unter der gleichen logischen ID erreichbar.

Eine grosse Herausforderung für das Speicherzentrum stellt die Beherrschung der Medien- bzw. Hardware-Migrationen dar. In der Regel haben die Medien eine längere Lebensdauer oder Verfügbarkeit als die entsprechenden Lesegeräte (z.B. Optical Disk). Zu berücksichtigen gilt auch, dass das Lesen eines Vorgängermediums in der Regel 10 Mal langsamer als das Schreiben des aktuellen Mediums ist (Beispiel VRSG Medienmigration: 2 Wochen Lesen der Optical Disks und anschliessend 2 Tage Schreiben der Daten auf WORM Tapes). In der Regel erfolgt die Medienmigration heute weitgehend realitätsgetrieben; primär dadurch, dass ein Hersteller den *end of life cycle* seines Produktes ankündigt. Wünschenswert wäre eine langfristige Planung dieser Migrationen.

Gefährlich ist das Ausreizen der Lebensdauer von Gerät und Medien, wie sie in gewissen Archiven vielleicht gerne betrieben wird – nach dem Motto: "Es läuft ja noch und wir haben sicherheitshalber mehrere Kopien, was sollen wir jetzt schon migrieren?" Hier ist schnell der allerletzte Migrationszeitpunkt verpasst, denn das notwendige Lesen aller Daten kann leicht zu einem unwiederbringlichen Defekt beim Medium oder beim Lesegerät führen.

Die Referenten weisen darauf hin, dass bei sehr geringen Datenzugriffen die Lesbarkeit der Bänder softwaretechnisch abgesichert werden muss und dass mit entsprechenden Prozessen diese Lesbarkeitsprüfung überwacht werden muss, damit nicht unerkannte technische Defekte über längere Zeit unwiederbringliche Schäden anrichten können

3.3 Kosten: Mit welcher Grössenordnung ist zu rechnen?

Zu den Kosten, die für die Speicherung im Rechenzentrum entstehen, können sich die Referenten nicht detailliert äussern, da es sich um ein Teil des Betriebsgeheimnisses handelt. Frau Ullrich kennt nur Arbeitseinheiten pro archiviertem Terabyte, das sind interne Verrechnungseinheiten.

Die VRSG berechnet bei der Belegarchivierung 0.10 CHF pro Beleg fürs Einstellen und 0.03 CHF für ein Jahr Archivierung. Ein Beleg ist im Schnitt 3 Seiten (schwarz/weiss) und etwa 150 Kilobyte gross.

Für kopal gibt es ein Rechnungsmodell:

<http://kopal.langzeitarchivierung.de/downloads/kopal-goes-live-IBM-DIAS-Perspektive-und-Preismodell.pdf>

3.4 Realisierbarkeit: Ist das etwas für die Staatsarchive?

Beide Referenten raten vom Aufbau eines eigenen Archiv-Rechenzentrum ab und sind der Meinung, dass ein Anschluss an ein bestehendes staatliches Rechenzentrum sicher besser sei, da die ganze Infrastruktur, wie Liegenschaft, Netzwerk etc. mitgenutzt werden kann.