

Der Langzeitspeicher des Stadtarchivs Stuttgart – ein strukturintegrativer Lösungsweg

HEIKE MAIER

1. Einleitung

Dem Stadtarchiv Stuttgart wurden 2006 umfangreiche digitale Unterlagen von Seiten der Fachämter angeboten. Für diese Menge, von der aus einem Amt alleine 1 TB als archivwürdig bewertet wurde, musste zeitnah eine finanzierbare, modular erweiterbare Speicherstruktur geschaffen werden. Es wird dargestellt, welche Optionen diskutiert, wie die Entscheidungsprozesse für LTO3 WORM bzw. Tivoli Storage Manager begründet wurden, und auf welche Herausforderungen das Stadtarchiv zu reagieren hatte.

2. Ausgangssituation

Seit dem Jahr 2005 beschäftigt sich das Stadtarchiv Stuttgart strukturiert und systematisch mit der Thematik digitale Langzeitarchivierung. In diesem Jahr wurde gemeinsam mit Unilog Avinci ein Rahmenkonzept zur digitalen Langzeitarchivierung¹ formuliert. Wesentliche Erkenntnisse des Stadtarchivs waren:

1. dass die zu diesem Zeitpunkt noch sehr kleinen Mengen von digitalen Beständen exponentiell wachsen werden, und dass
2. die notwendigen Kompetenzen, um diesen Beständen qualitativ nachhaltig zu sichern, nicht komplett im Stadtarchiv selbst abrufbar waren.

Neben dieser praktischen Aufgabe, eigene Wissens- und Erfahrungslücken mit der Unterstützung einer externen Beratungsfirma zu schließen, hatte dieses Rahmenkonzept auch eine strategische Bedeutung für das Stadtarchiv. Es machte auf die Thematik digitale Langzeitarchivierung aufmerksam, und zwar in der gesamten Verwaltung. Es zeigte darüber hinaus finanzielle und personelle Ressourcenbedürfnisse auf und dokumentierte gegenüber Dritten den Anspruch des Stadtarchivs, diese Thematik zukünftig zu besetzen. Ergänzend zu diesem ersten konzeptionellen Schritt wurde - ebenfalls gemeinsam mit Unilog Avinci - ein Konzept zur „Vorbereitung der Pilotübernahme von digitalen Daten durch das Stadtarchiv“² erarbeitet.

Zu diesem Zeitpunkt (2006) wurde in der Stadtverwaltung Stuttgart LTO 2 im Backup-Bereich eingesetzt. Das Stadtarchiv selbst verfügte über keinen spezifischen Speicherplatz. Digitale Objekte gab es in nennenswertem Umfang nicht, so dass diese provisorisch auf dem Server der Landeshauptstadt Stuttgart gelegt und ganz klassisch mit dem Backup-Verfahren gesichert wurden.

Ausgangspunkt für die Überlegung war daher, unter Berücksichtigung, dass in der Verwaltung der Landeshauptstadt Stuttgart ein elektronisches Dokumentenmanagement eingeführt und im Rahmen dieses Vorhabens eine physische Speicherlösung geschaffen wird, die Suche nach einem Speicher, der bis zur Einführung des DMS ohne Refreshing der Datenträger und der Datenformate mit möglichst geringen Investitionskosten genutzt werden kann. Datenvolumen, Verfügbarkeit, Sicherheit und Langlebigkeit wurden in der Analyse betrachtet.

Im Rahmen des Konzeptes hatte das Stadtarchiv drei Muss-Kriterien formuliert:

1. Objekte werden auf zwei Speichermedien archiviert, einem Mastermedium und einem Benutzungsmedium.
2. Mindestens ein Speichermedium muss über eine WORM-Technologie verfügen.
3. Die beiden Speichermedien müssen räumlich getrennt aufbewahrt werden.

Unter der Annahme, dass das Stadtarchiv über einen Zeitraum von sieben Jahren 1 TB Daten übernehmen würde, ergab die von der Unilog Avinci durchgeführte Wirtschaftlichkeitsberechnung folgende Kosten für verschiedene Speichermedien:

¹ Rahmenkonzept zur Langzeitarchivierung digitaler Daten für die Landeshauptstadt Stuttgart. Dieses Konzept kann von anderen Archiven zu ihrer internen Verwendung beim Stadtarchiv Stuttgart angefordert werden (stadtarchiv@stuttgart.de). Beteiligt am Konzept waren die Unilog Systems Avinci GmbH, Informations- und Kommunikationstechnik und das Stadtarchiv der Landeshauptstadt Stuttgart.

² Vorbereitung der Pilotübernahme von digitalen Daten durch das Stadtarchiv (Konzeptentwurf Version 0.06) der Unilog Systems Avinci GmbH.

lichkeit nicht in Betracht. Darüberhinaus gab es innerhalb des Stadtarchivs keinen einzigen Rechner oder eine externe Festplatte für Ablieferungen in diesem Umfang. Das Transferproblem lösen wir mittlerweile über zwei mobile Festplatten à 600 GB. Für die Eingangsbearbeitung bzw. für die Bereitstellung zur Langzeitarchivierung wurde ein leistungsfähiger Rechner⁷ mit 3 TB Speicherplatz gekauft. Das Stadtarchiv hatte sein geplantes Speichermedium für die Langzeitarchivierung genau in dem Moment aufgegeben, als die größte digitale Ablieferung bevorstand, ohne dass eine wirkliche Alternative absehbar war. Eine durchaus kritische Situation. Obwohl wir die dortige Empfehlung nicht umsetzen, war das Speicherkonzept bei den weiteren Schritten dennoch nützlich. Wir hatten dort wichtige Grundsätze festgelegt, an denen wir uns auch weiterhin orientierten. Es waren hier unterschiedliche Speichermedien bewertet worden. Wir hatten gelernt, dass Technologiereife ein nicht zu unterschätzendes Kriterium ist, und dass unsere Volumeneinschätzungen nicht zu zaghaft sein durften. Mit diesem Erfahrungshintergrund und durchaus unter dem Druck, zeitnah eine Lösung finden zu müssen, wurde das Stadtarchiv mit einer neuen innerstädtischen Situation konfrontiert. Denn unabhängig von den o.g. Ereignissen, aber nahezu zeitgleich, beschloss die Stadtverwaltung Stuttgart, für das Backup von LTO2 auf LTO3 umzusteigen und neue Tape Libraries zu beschaffen. Im Gegensatz zu LTO2 sind LTO3 WORM-fähig. Damit ergab sich für uns eine neue Option, nämlich den Versuch zu unternehmen, die geplante Backup-Umgebung, insbesondere die LTO3 WORM, als Infrastruktur für die eigenen archivischen Zwecke mitzunutzen.⁸

3. Umsetzung einer Speicherlösung im Rahmen der vorhandenen Backup-Lösung

Das Stadtarchiv setzte sich daraufhin im Juli 2006 mit der Abteilung Information und Kommunikation (IuK)⁹ in Verbindung, um zu prüfen, ob die Nutzung von LTO3 WORM für archivische Zwecke möglich wäre. Die Reaktionen auf diese Anfragen waren positiv. Prozessual arbeiteten sich in den folgenden Monaten sowohl Stadtarchiv wie auch IuK an die derzeitige Lösung heran.

Tape Libraries / Standortfrage

Die IuK plante zunächst den Kauf einer neuen Tape Library. Diese sollte die bisherige ersetzen und LTO3 WORM-fähig sein.

Hieraus ergaben sich unter Berücksichtigung der Muss-Kriterien folgende Nutzungsoptionen für das Stadtarchiv:

- a) Für den Zeitraum der Archivierung müssen Master und Nutzungskopie in die neue Tape Library eingelegt werden. Sobald der Archivierungsprozess abgeschlossen ist, muss die Masterkopie aus der Tape Library manuell entfernt werden und an einen Standort in einem anderen Gebäude transportiert werden.
- b) Die alte Tape Library wird parallel zur neu gekauften Tape Library weiterhin genutzt. Diese ist LTO3-fähig, kann aber nicht mit einer WORM-Emulation umgehen. Die Tape Libraries stehen jeweils in getrennten Gebäuden.

Beide Lösungen erschienen nicht ideal, wobei die *Lösung a)*, die einen hohen Organisationsaufwand erfordert, als störanfälliger betrachtet wurde. Archivierungsprozesse im Stadtarchiv fallen nicht regelmäßig an, sondern unregelmäßig und zum Teil auch ad hoc, so dass kein wöchentlicher Plan für die Herausnahme der Bänder formuliert werden konnte. Hierfür einen zuverlässigen Workflow zu entwickeln, erschien allen Beteiligten problematisch. Die *Lösung b)* erschien weniger kritisch, auch, wenn so nur ein LTO mit einer WORM-Emulation versehen gewesen wäre. Aber die Muss-Kriterien waren generell erfüllt.

⁷ WortmanTerra BTO Server Intel S5000 VSATA /FBDIMM/ 6-Port SATA Mainboard, Prozessor: Intel XEON 5130 2 GHz (Doppelkernprozessor) FSB 1333 4 MBcache aktiv, Arbeitsspeicher: 2x 1GB Festplatten: 4x 750 GB SATAII HDD 2x 250 GB SATA II HDD 3Ware 965SE PCI Express SATAII RAID-Controller, Betriebssystem Win XP mit Service Pack 2; RAID Konfiguration: Konfiguration RAID: RAID 0 (Stripping) 4x 750 GB HD ergibt ein logisches Laufwerk; max. 3 TB Datenvolumen, RAID1 (Spiegelung) 2x 250 GB HD max. 250 GB für Betriebssystem und Programme.

⁸ Die Entscheidung gegen das Medium UDO ist keine grundsätzlich negative Aussage gegenüber der Qualität des Mediums als Langzeitspeicher. In unserem Kontext: Startvolumen > 1 TB, damit die Notwendigkeit der Beschaffung einer Jukebox, bei gleichzeitiger Möglichkeit mit LTO3 ein WORM-fähiges Medium kostengünstiger zu nutzen und die etwas kritische Marktlage in Bezug auf die quasi Monopolstellung der Firma Plasmon, kam UDO für das Stadtarchiv als Archivierungsmedium nicht mehr in Frage.

⁹ Die Abteilung Information und Kommunikation strukturiert, organisiert und administriert die zentralen EDV-Dienstleistungen in der Stadt.

– Speicherung –

Das Stadtarchiv entschied sich für letztere Variante, wurde dann aber mit einer neuerlich veränderten Situation bezüglich der technischen Infrastruktur konfrontiert. Dabei konnte das Stadtarchiv letztlich von einer Entwicklung profitieren, die nicht zuletzt durch den gesteigerten Volumenbedarf des Stadtarchivs selbst ausgelöst wurde: Die Stadtverwaltung entschied sich, nicht nur eine, sondern zwei neue Tape Libraries zu beschaffen. Beide Tape Libraries (HP ESL 322e, mit Platz für 322 Bänder,¹⁰ 5 Laufwerke ; HP MSL 4048 mit Platz für 48 Bänder,¹¹ 1 Laufwerk) sind WORM-fähig und stehen in getrennten Gebäuden. Ein Vereinbarung mit der IuK hält für das Stadtarchiv 10 TB Bandplatz in jeder dieser Libraries verfügbar, dieser Speicherplatz ist perspektivisch ausbaubar. Mit dieser Lösung hat das Stadtarchiv ein geeignetes Medium für die Archivierung im Einsatz, die Standortfrage der Medien gut gelöst und mit 10 TB ein ausreichendes Startvolumen für die Langzeitarchivierung gefunden.

Datentransfer zum Speicher

Neben dem Medium und dem Standort war die Frage des Datentransfers zum Langzeitspeicher zentral. Auch diese Problemlösung durchlief verschiedene Stadien. Die Stadtverwaltung entschied sich Mitte 2006 für den Tivoli Storage Manager (TSM) als neues Produkt für das Backup. Dieses Produkt sollte auch für den Transfer der Archivobjekte genutzt werden.

Zu dem Zeitpunkt, als das Stadtarchiv eine Speicherlösung im Rahmen der Backup-Struktur der LHS zu entwickeln versuchte, wurde diese bezüglich der Hardware und der Software umgestellt. In Bezug auf die Hardware konnte das Stadtarchiv hiervon (s.o) positiv profitieren. In Bezug auf die Software TSM erwies sich der Zeitpunkt einerseits als günstig – man konnte auf vorhandene externe Beraterleistung zurückgreifen, die von der IuK selbst genutzt wurde –, andererseits gab es innerhalb der IuK mit dem neuen Produkt bis dato keine langfristigen Erfahrungen, was zu einem erhöhten Kommunikationsbedarf führte.

Das Stadtarchiv formulierte folgende KO-Kriterien, die erfüllt sein sollten, um TSM als Werkzeug nutzen zu können; sie wurden zusammen mit der IuK in der TSM-Archivierungspolicy¹² umgesetzt:

1. Archivpakete müssen über die Suche auffindbar sein, gleichgültig in welcher Weise sie gepackt sind (Ordnerstruktur, tar, zip oder sonstige Verpackungsstruktur).

Dieser Punkt erscheint vordergründig selbstverständlich. TSM sucht aber standardmäßig mit der graphischen Oberfläche nur über einzelne Dateien, d.h. tar und zip-files sind auffindbar, nicht aber Ordner. Über die Befehlszeile und die hier erweiterten Suchoptionen lässt sich dieses Problem, wenn auch nicht sehr komfortabel, lösen.

2. Aktionen / Protokolle, die die Archivierung betreffen, müssen automatisiert in einer durch das Stadtarchiv nachvollziehbaren Weise just in time zur Verfügung gestellt werden.

Die TSM-Administratoren erhalten routinemäßig über die ablaufenden Prozesse Protokolle. TSM stellt Mandanten – wie zum Beispiel dem Stadtarchiv – ein derartiges Protokoll ohne Admin-Rechte nur bei einer automatisierten Abholung der Objekte zur Verfügung. Bei einer automatischen Abholung ist aber die Nutzung des TSM-Beschreibungsfeldes nicht möglich. Diese Nutzung war jedoch beabsichtigt, um den Objekten die Bestandsbezeichnungen mitgeben zu können. Das Stadtarchiv entschied sich daraufhin, seine Objekte manuell zu transferieren. Die Protokollproblematik wurde mit einem speziellen für das Stadtarchiv geschriebenen Skript gelöst: Dieses erstellt einen Auszug aus dem Gesamt-Admin-Protokoll und schickt täglich ein TSM-Protokoll automatisiert per E-Mail an das Stadtarchiv.

3. Objekte des Stadtarchivs dürfen nicht komprimiert archiviert werden.

Im Backup wird standardmäßig komprimiert, um den Speicherplatz effektiv zu nutzen. Hier musste geklärt werden, ob das Ausschalten der Komprimierung hardwareseitig bzw. softwareseitig notwendig ist. Im TSM können pro definiertem Mandanten die Komprimierungsparameter hinterlegt werden, so dass hardwareseitig keine Modifikationen notwendig sind. TSM kann also parallel archivieren und ein Backup betreiben.

¹⁰ Siehe: <http://h10010.www1.hp.com/wwpc/de/de/sm/WF25a/24251-24383-1336321-1336321-12326042-12326048.html> (Stand 26.02.07).

¹¹ Siehe: <http://h10010.www1.hp.com/wwpc/de/de/sm/WF25a/24251-24383-1336321-1336321-12326030-12400880.html> (Stand 26.02.07).

¹² Die von der IuK und dem Stadtarchiv Stuttgart definierte TSM-Archivierungspolicy kann bei Bedarf beim Stadtarchiv angefordert werden (stadtarchiv@stuttgart.de).

– Speicherung –

4. Objekte des Stadtarchivs müssen separat auf nur für das Stadtarchiv bestimmte Medien archiviert werden.

Auch diese Forderung ist gut umsetzbar. In der Archivpolicy werden die dem Stadtarchiv zugeordneten Bänder definiert. Das Stadtarchiv hat sich zusätzlich entschieden, mit einem speziellen Label (Code) versehene Bänder zu nutzen.

Insgesamt wurden die KO-Kriterien erfüllt. Dies soll aber nicht darüber hinwegtäuschen, dass TSM für das Stadtarchiv kein optimales Werkzeug darstellt. Es erfüllt bestimmte Mindestanforderungen, andere Anforderungen jedoch nicht.

Als besonders problematisch aus archivischer Sicht ist die Tatsache, dass TSM die zu archivierenden Objekte „kapselt“. Wenn TSM eine Datei speichert, setzt er davor und dahinter einen Marker. Man kann die Daten nur mit dem TSM wieder auslesen. Wenn die Stadtverwaltung auf ein anderes Programm umsteigt, müssen wir alle archivierten Objekte auslesen und neu archivieren. Problematisch ist ebenfalls, dass wir keinen Zugriff auf die Archivobjekte haben, wenn die TSM-Datenbank kaputt geht. Diese Gefahr betrifft aber auch die gesamten Backup-Daten der Stadt Stuttgart, so dass wir diese Risiko in Kauf nehmen. In diesem Verbund erscheint uns dies akzeptabel, als einzelner Akteur würden wir diese Risiko aber niemals tragen. Ein weiteres Problem sehen wir darin, dass TSM keine Aussagen über den Archivierungsort – also auf welcher LTO sich das digitale Archivgut befindet – macht.

Speicherworkflow¹³

Das digitale Archivgut wird mittels mobiler Festplatten ins Stadtarchiv transferiert. Dort werden die Daten auf die so genannte Archivwerkstation (W41ARCH) übertragen und bearbeitet. Dann werden sie dort in ein definiertes Verzeichnis (z.B. E:\Daten_TSM) eingestellt. Der eigentliche Archivierungsprozess wird anschließend manuell vom Stadtarchiv angestoßen. Innerhalb des TSM wird das digitale Archivgut zunächst auf S-ATA-Platten („Stadtarchiv_Diskpool“) zwischengespeichert, da für eine direkte Beschreibung auf LTO der erforderliche Input nicht immer gegeben ist. Würde von der Archivwerkstation (W41ARCH) direkt auf LTO archiviert, bestünde eine zu große Wahrscheinlichkeit, dass das LTO bei zu geringem Input in einen Start-/Stop-Modus wechselt und somit die Gefahr einer Beschädigung des LTO bestünde. Dem „Stadtarchiv_Diskpool“ werden S-ATA-Volumes zugewiesen, auf denen ausschließlich Daten des Stadtarchivs zwischengespeichert werden. Zu einem täglich festgelegten Zeitpunkt migriert TSM die Daten aus dem „Stadtarchiv_Diskpool“ in den ersten LTO-WORM-Pool („Stadtarchiv_Primarypool_ESL“) auf der HP ESL-Library. Diesem Band-Pool werden wieder genau definierte Bänder zugewiesen, die ausschließlich für die Archivierung des Stadtarchivs vorgesehen sind. Hierbei handelt es sich um die Nutzungsmedien aus Stadtarchivsicht. Nun folgt die TSM-seitige Erstellung der Kopie (Mastermedium aus Stadtarchivsicht) auf die Bänder des zweiten Bandpools („Stadtarchiv_Copypool_MSL“) in der HP MSL-Library, die genauso wie die Bänder im ersten Band-Pool genau definiert wurde und ausschließlich für die Archivierung des Stadtarchivs vorgesehen sind. Führt das Stadtarchiv Rückspielungen durch, werden die Nutzungsmedien (Bänder im „Stadtarchiv_Primarypool_ESL“) verwendet. Die Mastermedien bleiben unberührt und werden nur herangezogen, falls Daten auf den Arbeitsmedien fehlerhaft sind.

¹³ Text/Grafik Vitali Kraft, Haupt- und Personalamt, IuK.

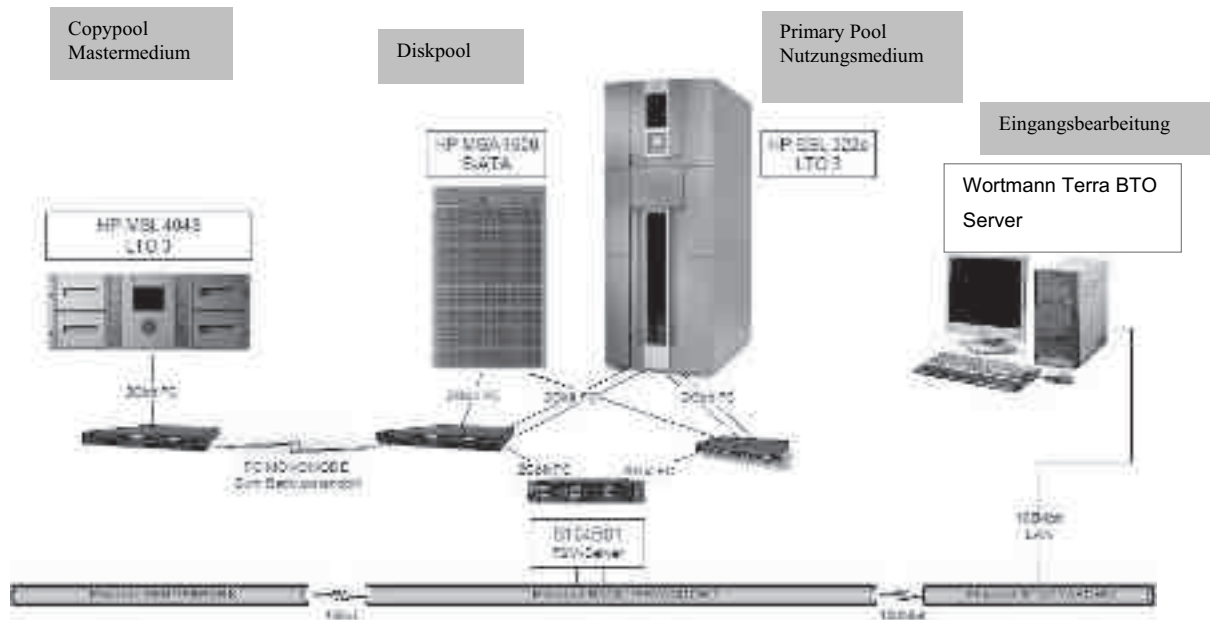
– Speicherung –

Speicherinfrastruktur Stadtarchiv Stuttgart

Standort:
Verwaltungsgebäude 1

Standort: Verwaltungsgebäude
2

Standort:
Stadtarchiv



4. Ausblick

Das Stadtarchiv Stuttgart hat sich mit dieser Lösung erstmals Platz für 10 TB digitaler Archivbestände geschaffen. Dies ist vergleichbar einem neu angemieteten Magazin für analoge Bestände. Das Archiv hat keine eigene autonome Speicherlösung entwickelt, und auch eine externe Lösung bei einem Dienstleister war aus Kosten- und Sicherheitsgründen nicht gewünscht. Es wurde vielmehr ein integrativer Lösungsweg umgesetzt. Die vorhandene technische Backup-Infrastruktur der Stadtverwaltung wurde für das Stadtarchiv modifiziert, um so eine Speicherlösung für die Langzeitarchivierung zu ermöglichen. So wurde ein sehr pragmatischer, an den Ressourcen des Stadtarchivs orientierter Lösungsweg umgesetzt. Wir haben damit keine optimale Lösung realisieren können. Als kleine Institutionen versuchen wir unsere Möglichkeiten maximal auszuschöpfen und uns die Defizite unserer Lösungswege transparent zu machen. Einerseits können wir nur so verantwortlich mit den Risiken umgehen, andererseits sehen wir uns so zumindest in der Lage, weiterhin sukzessive Verbesserungen – sobald diese für uns realisierbar sind – vorzunehmen.

Das nun anstehende Thema für das Stadtarchiv ist es, eine Lösung für die Eingangsbearbeitung (Ingest) der digitalen Daten zu finden.