



Speicherinfrastruktur und Langzeitspeicherung von Digitalisaten

Innenansicht der bestückten Tape Library

Aus Digitalisierungsprojekten und der nunmehr digitalen Bundessicherungsverfilmung entstandene Datenmengen stellen die Archive bei der Langzeitspeicherung und Datenverwaltung vor erhebliche Herausforderungen. Das Landesarchiv Sachsen-Anhalt reagiert darauf mit einer redundanten Speicherinfrastruktur und einem dezidierten Speicherkonzept.

Das Jahr 2013 stellte für das Landesarchiv Sachsen-Anhalt eine Zäsur dar: Erstmals wurden im größten Maßstab Digitalisierungsprojekte mit dem Ziel durchgeführt, einen webbasierten Zugriff auf das Archivgut zu ermöglichen. Für die eigentliche Digitalisierung und die Online-Präsentation lagen bereits definierte Prozesse vor, für die Speicherung der Digitalisate standen sie zu diesem Zeitpunkt noch aus. Während die Speicherbedarfe kontinuierlich wuchsen, verblieben die Master- und Nutzungsderivate auf Wechseldatenträgern sowie einem Netzwerkspeicher. Dies änderte sich 2017 mit der Beantragung von Fördermitteln bei der DFG zur Digitalisierung und Onlinestellung des Kernbestandes C 20 I Oberpräsident Magdeburg. Einen Teil des Antrags bildeten die Planungen zur Langzeitspeicherung der mehr als 3 Mio. zu erwartenden Einzelobjekte, woraus ein generalisiertes Speicherkonzept entstand, in dem sämtliche Digitalisate gleichwertig behandelt werden. Die konzeptionellen Arbeiten des Landesarchivs enthalten dabei die organisatorischen Anforderungen, ein Rechte- und Rollenkonzept, konkrete Handlungsan-

weisungen und eine Übersicht über die geforderte technische Infrastruktur.

Dateiformate

Einen wichtigen Teilaspekt der Langzeitspeicherung stellen die verwendeten langzeitstabilen Dateiformate dar. Das Landesarchiv verwendet folgende Dateiformate:

Rasterbilder

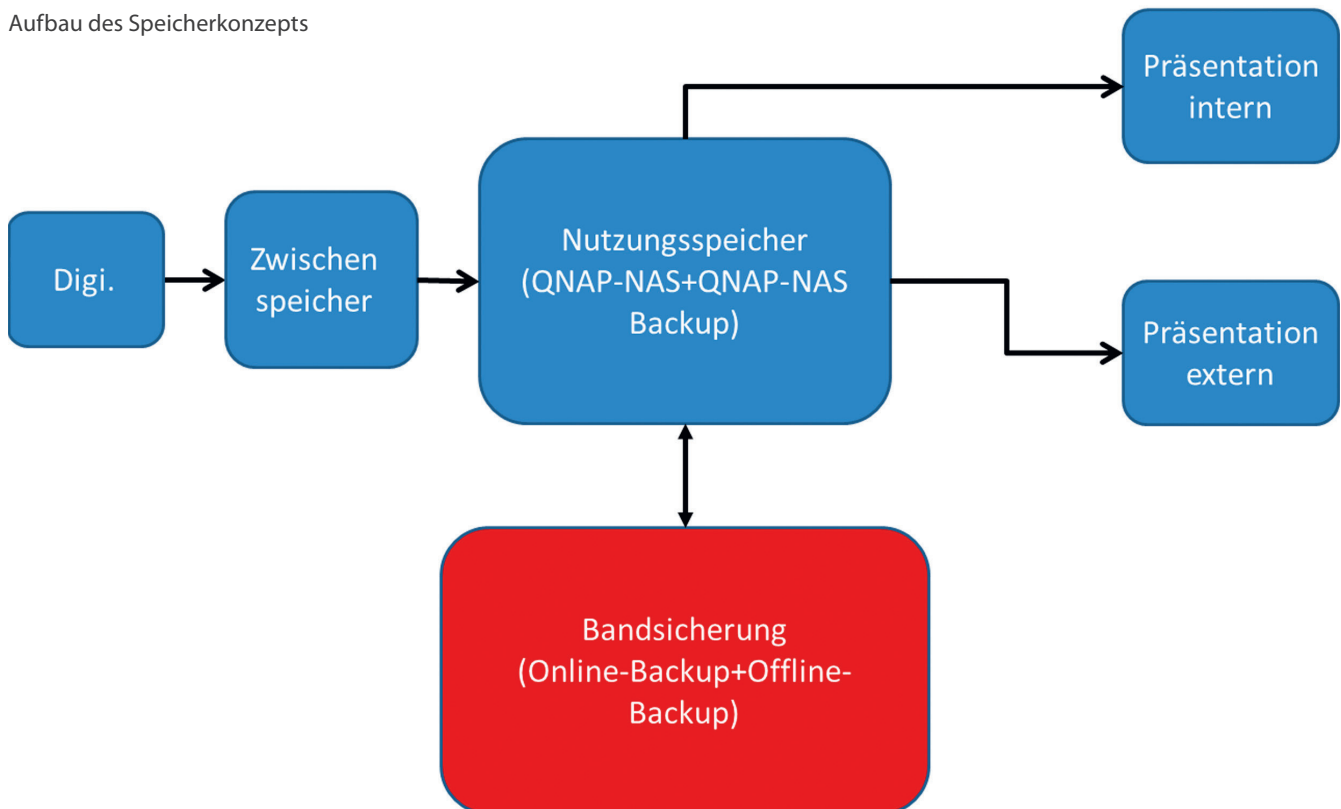
- Masterderivat: JPEG2000, alternativ TIFF
- Nutzungsderivat: JPEG

Audio

- Masterderivat: WAVE
- Nutzungsderivat: MP3

Video

- Masterderivat: MXF
- Nutzungsderivat: MPEG-2



Im Rahmen des DFG-Projektes wurde untersucht, in welchem Dateiformat die Masterderivate von C 20 I zu liefern sind. Vor dem Hintergrund, dass das bisher verwendete Dateiformat TIFF sehr speicherplatzintensiv ist, sollte die von den DFG-Praxisregeln zur Digitalisierung aufgezeigte Alternative JPEG2000 auf seine Verwendungsfähigkeit als Masterformat geprüft werden. Als Ergebnis der Prüfung wurde für Rasterbildformate JPEG2000 eingeführt, was eine Ersparnis des prognostizierten Speicherbedarfs von ca. 40-60 % gegenüber dem bisherigen Masterformat TIFF bedeutete.

Hinsichtlich der Qualität zeigte sich, dass JPEG2000 sämtliche technischen Werte aus den äquivalenten TIFF-Dateien übernommen hat (Anzahl Farben, Farbtiefe, Skalierung etc.). Im Tiefenzoom sind keine visuellen Unterschiede mehr wahrnehmbar. Gleichzeitig ist zu berücksichtigen, dass JPEG2000 gegenüber TIFF eine höhere Toleranz bei Übertragungsfehlern aufweist. Hinzu kommt, dass im Fehlerfall JPEG2000-Dateien zwar unschärfer werden, im Gegensatz zu TIFF-Dateien aber keinen kompletten Informationsverlust erleiden.

Speicherinfrastruktur

Zwischenspeicher

Der Zwischenspeicher stellt die erste Station der Speicherinfrastruktur dar. Als "digitale Laderampe" dient dieser Speicher dazu, angelieferte Digitalisate zu empfangen und eine Qualitätssicherung auf eine korrekte und vollständige Lieferung durchzuführen. Technisch besteht der Zwischenspeicher aus einem Netzwerkspeicher (NAS), ohne zusätzliche Sicherung, da die Digitalisate hier zwischengelagert werden. Mit einer erfolgreichen Qualitätssicherung wird die Ablieferung in den Nutzungsspeicher importiert. Die Ablage der digitalisierten Verzeichnungseinheiten erfolgt auf dem Zwischenspeicher unterteilt nach den Archivaliengattungen Bild, Audio und Video, hierunter vornehmlich in einer vergabeorientierten Verzeichnisstruktur.

Nutzungsspeicher

Für den unmittelbaren Zugriff auf die digitalen Objekte durch die Facharchivar*innen und eine mittelfristige Aufbewahrung besteht im Landesarchiv der sogenannte Nutzungsspeicher. Auch dessen Speichertechnologie basiert auf einem NAS-System mit RAID 6-Level, welcher jedoch im Gegensatz zum Zwischenspeicher redundant vorgehalten wird, um

einem möglichen Komplettausfall des Systems entgegenzuwirken. Die Sicherung des zweiten Speichers erfolgt durch eine 1:1-Kopie der vorgehaltenen Digitalisate und nicht durch eine klassische Backup-Software, um Abhängigkeiten zu proprietärer Software zu vermeiden. Beide NAS-Speicher befinden sich räumlich gut 50 Kilometer Luftlinie entfernt voneinander, was die wichtigsten, örtlichen Risiken minimiert. Die Digitalisate sind hier in einer bestandsorientierten Ablage organisiert.

Bandsicherung

Die Kernkomponente für die langfristige Bewahrung der Digitalisate ist die Bandsicherung. In 2020 konnte mit der Installation dieser wichtigen Komponente in Form einer Tape Library der Firma Quantum die praktische Implementierung des Speicherkonzepts abgeschlossen werden. Als Speichermedium eignet sich insbesondere Magnetband (Tape), da es die nötige Langzeitstabilität aufweist, um auch in mehreren Jahren, unter den regulierten Bedingungen eines Magazinraums bis zu 30 Jahre, noch lesbar zu sein. Die optimalen klimatischen Bedingungen zur Lagerung der Tapes liegen bei 5-23 Grad Celsius und einer relativen Luftfeuchte im Bereich von 10-50 % (nicht kondensierend), die durch Magazinräume des Landesarchivs abgedeckt wird.

Ein weiteres, maßgebliches Argument für den Einsatz einer Tape Library und die Speicherung auf Tape ist der Kostenfaktor. Über eine Wirtschaftlichkeitsbetrachtung wurde festgestellt, dass eine Speicherung auf Tape über eine inhouse betriebene Lösung die im Vergleich wirtschaftlichste Variante unter den gestellten Rahmenbedingungen darstellt. Die Anforderungen des Landesarchivs an die Langzeitspeicherungskomponente orientieren sich an den Schutzziele der IT-Grundschutzkataloge des BSI:

- Verfügbarkeit: normal
- Integrität: hoch
- Vertraulichkeit: hoch

In der Langzeitspeicherung muss eine unbeabsichtigte Veränderung der digitalen Objekte verhindert werden. Ebenso sind schutzfristenbehaftete Digitalisate vor unberechtigtem Zugriff durch Dritte zu schützen. Normale Zugriffszeiten von bis zu 72 Stunden sind hingegen als tolerabel anzusehen. Durch die planerische Unsicherheit, inwieweit drittmittelgeförderte Digitalisierungsprojekte tatsächlich in der beantragten Form bewilligt werden, entstehen starke Schwankungen in den prognostizierten Speicherbedarfen, die ggf. auch kurzfristig anfallen. Bei der Aus-



Ansicht der Tape Library im Serverschrank

wahl der Speichertechnologie muss dies ebenfalls beachtet werden. In der praktischen Umsetzung kann die Tape Library sämtliche Anforderungen erfüllen. Bereits über das Speichermedium des Magnetbands (hier LTO-8) werden Veränderungen des Bitstroms unterbunden, zusätzliche Integritätschecks decken mögliche Veränderungen während einer Revitalisierung in den Nutzungsspeicher auf. Das Offline-Backup an den Standorten sichert die Digitalisate zudem im Falle eines Cyberangriffs – das Einspielen der Tapes ist prinzipiell in jede andere LTO-8-fähige Tape Library möglich. Über die Lagerung in den archiveigenen Magazinräumen ist der Zugriff durch das berechnete Fachpersonal gesichert.

Die Tapes werden redundant an den Standorten Magdeburg und Dessau gelagert, wobei sichergestellt wird, dass sich ein Digitalisat stets mindestens auf zwei Tapes befindet, indem gleichartige Kopien erzeugt werden. Über die interne Speicherlogik der Library als auch über die Barcodes sind die Tapes einzeln identifizierbar. Sollten einzelne Digitalisate oder ganze Verzeichnungseinheiten vom Tape wieder für eine Benutzung benötigt werden, werden die Tape Library mit dem entsprechenden Satz an Tapes bestückt und die Daten auf den Nutzungsspeicher revitalisiert. Ein integrierter Cache-Speicher sorgt für schnelle Bereitstellungszeiten. Als Langzeitspeicher sollen die Tapes die Verzeichnisstruktur des Nutzungsspeichers 1:1 reproduzieren, weshalb auch hier auf proprietäre Backup-Software und Verschlüsselungen verzichtet wird.

DIMAG

Das Digitale Magazin (DIMAG) ist von der Speicherinfrastruktur für Digitalisate abzugrenzen. Genuin digitales Schriftgut, welches für archivwürdig befunden wurde, muss in einer OAIS-konformen Infrastruktur auf Langzeit bewahrt werden. Hierzu gehören eine mindestens dreifach redundante Speicherinfrastruktur in jeweils unterschiedlichen Speichertechnologien, eine permanente Integritätskontrolle und regelmäßige Überwachung der Dateiformate auf Dateiformatobsoleszenz zur Gewährleistung der technischen Lesbarkeit einschließlich Massenkonzertierungen. Der komplette Verlust einer Verzeichnungseinheit ist auszuschließen.

Im Gegensatz dazu zielen Digitalisierungsprojekte im Archivwesen primär darauf ab, bildhafte Derivate von physischem Schriftgut für Nutzungszwecke in einer Online-Präsentation zu erzeugen. Eine Vorlage der analogen Archivalien im Lesesaal ist in der Regel nicht mehr erforderlich, wodurch bestandserhalterische Risiken, die durch regelmäßige Nutzung auftreten, vermieden werden. Das eigentliche Ziel der Erhaltung bleiben jedoch die analogen Archivalien. Wenn unwiederbringliche Informationsverluste drohen oder gar bereits geschehen sind, treten die digitalen Derivate an die Stelle des Archivals – dies gilt insbesondere für die Tonüberlieferung. Diese nun zu Ersatzdigitalisaten deklarierten Objekte werden aus der internen Speicherinfrastruktur in das im DAN-Verbund länderübergreifend betriebene digitale Magazin exportiert, um hier die entsprechenden Erhaltungs- und Überwachungsmaßnahmen durchzuführen.

Ausblick

Eine redundante Speicherinfrastruktur muss in Digitalisierungsprojekten stets mitgeplant werden, um auch die digitalen Derivate langfristig vorhalten zu können. Unter anderem können die Praxisregeln zur Digitalisierung der DFG oder auch die Best-Practice-Empfehlungen der „Handreichung zur Digitalisierung von Archivgut“ der BKK zum Aufbau und Betrieb eines solchen Speichers und dessen Organisation nützliche Hinweise geben. Dabei stehen Archive vor einer grundsätzlichen Herausforderung: Als Derivate des schriftlichen Archivguts werden Digitalisate nicht gleichartig zu genuin digitalem Archivgut OAIS-konform behandelt, weshalb parallele Infrastrukturen aufgebaut werden müssen, die insbesondere auf wirtschaftliche Effizienz ausgelegt sind, um in erster Linie der Speicherplatzproblematik zu begegnen. Hierbei kann ein Informationsverlust durch fehlende

Migrationsroutinen, bit-rot bzw. schleichende Datenkorruption oder schlicht fehlendes Management in der Praxis mitunter hingenommen werden, da das Archivgut gegebenenfalls erneut digitalisiert werden kann. Dies sollte aber aus bestandserhalterischen Aspekten vermieden werden. Der Transport- und insbesondere der Scanprozess stellen eine enorme Stresssituation für die Informationsträger dar.

Mit dem Einsatz einer Tape-Library kann der Kostendruck erheblich reduziert werden, während die Digitalisate nachhaltig gesichert werden. Mit seiner Speicherinfrastruktur besitzt das Landesarchiv nun eine solide technische als auch organisatorische Lösung zur Bewältigung der Speicherbedarfsproblematik für Digitalisate.

Da es sich bei der Langzeitspeicherung von Digitalisaten um eine konzeptionelle Herausforderung handelt, der sich alle Archive zu stellen haben, sollten die bestehenden best-practices in verbindliche Standards für die Langzeitspeicherung überführt werden. Durch die inhaltliche Nähe zur digitalen Archivierung ist durchaus ein „OAIS-light“ denkbar, welches in Abgrenzung zur ISO 14721 tragfähige Mindestvoraussetzungen für die Langzeitspeicherung definiert. Eine Verbundstruktur, wie sie bereits in der digitalen Archivierung mit DIMAG besteht, könnte Synergieeffekte schaffen und Belastungen verringern.

Björn Steffenhagen

Digitale Derivate sichern die Information alter Tonträger. Als Ersatzdigitalisate werden sie ins Digitale Magazin überführt.

