



Der Ingest in das Digitale Magazin des Bundesarchivs

Bestandsaufnahme und Ausblick



Akten



Bilder



Filme



Töne



Karten

Dr. Beatrix Dietel, Referat IT 4

Agenda

- I: Begriffsbestimmung und Infrastruktur
- II: Ausgangslage und Rahmenbedingungen der digitalen Langzeitarchivierung im Bundesarchiv
- III: Der Ingest ins Digitale Magazin
- IV: Ausblick

Begriffsbestimmung und Infrastruktur

Das „Digitale Magazin“ des Bundesarchivs

Das „Digitale Magazin“ des Bundesarchivs ist eine in den letzten Jahren aufgebaute, leistungsfähige Speicherinfrastruktur (Bandspeicher), die geeignet ist:

- *sehr* große Mengen digitaler Objekte dauerhaft zu sichern
- Grundprinzipien der digitalen Langzeitarchivierung umzusetzen
- technologische Weiterentwicklungen zu begünstigen.

Daneben:

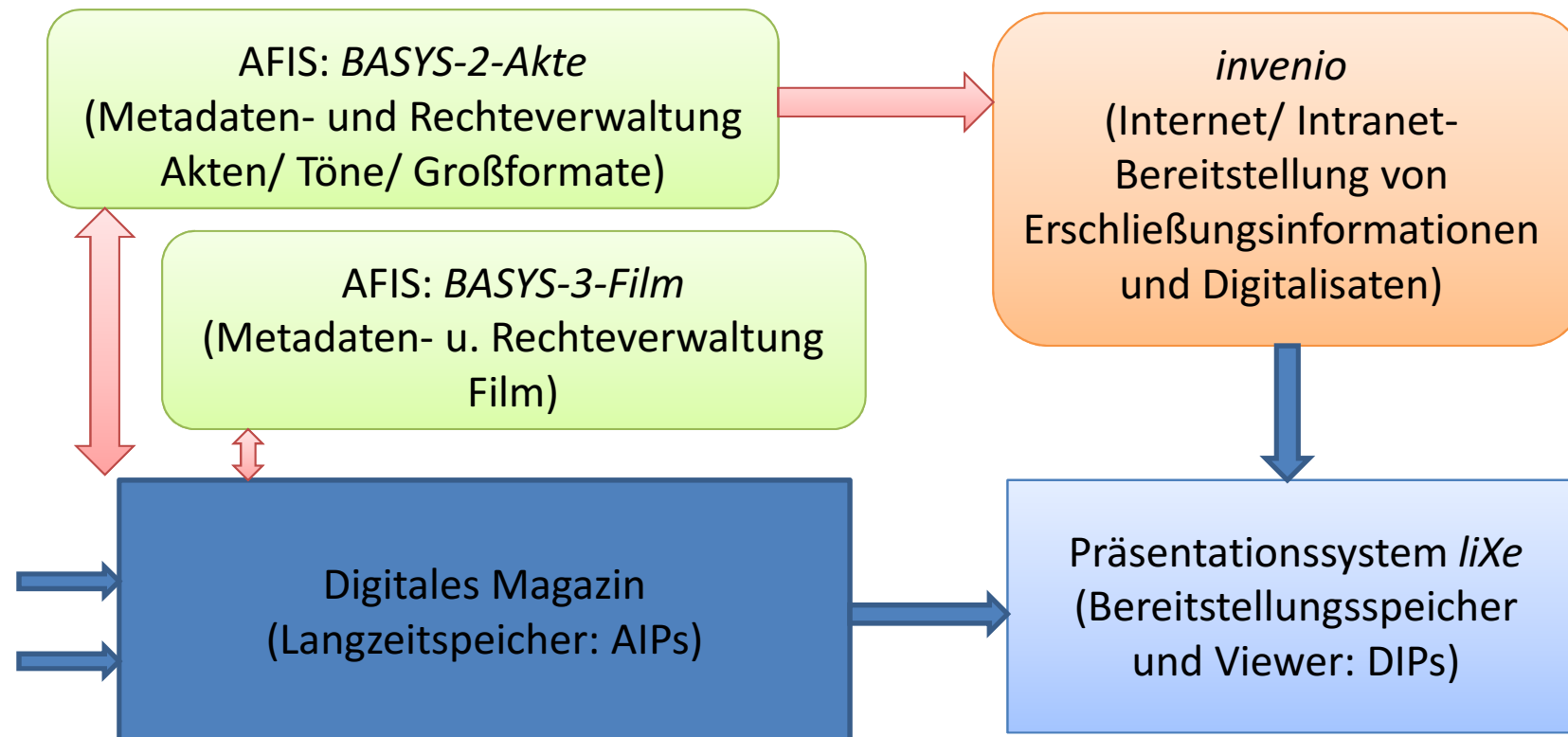
- weitere, vorläufige Speicher, die nicht den o.g. Kriterien entsprechen und deren Inhalte (Altdigitalisate, MD, genuin digitale Unterlagen aus rezenten Übernahmen) sukzessive ins DM überführt werden sollen
- separater Bereitstellungsspeicher für Benutzungsrepräsentationen (DIPs)

Hard- und Softwarestruktur des Digitalen Magazins

- Hardware
 - Bandspeicher Oracle SL 8500 (LTO 8, 12 TB/Band)
 - 8 Laufwerke, 4 Roboter
 - Gesamtkapazität: 100.000 slots/ **120 PB** (nach aktueller LTO-Technologie)
 - aktuelle Belegung: **15 PB** (1419 Tapes) in der Library und **4,7 PB** (392 Tapes) im Tresor (mit Redundanz)

- Software
 - derzeit nur *systeminterne, proprietäre* Software (Oracle HSM) mit rudimentären Funktionalitäten (nichthierarchische Fileablage), *keine* relationale Datenbankanwendung
 - Metadatenverwaltung erfolgt derzeit ausschließlich in AFISsen
 - *geplant*: eigene Datenbankanwendung „BASYS-3-Magazin“ zur Metadatenverwaltung und Unterstützung der Arbeitsprozesse

Das Digitale Magazin in der Anwendungsarchitektur des Bundesarchivs



Ausgangslage und Rahmenbedingungen

Bedarfslage und besondere Herausforderungen

Besonderheiten im Bundesarchiv

Das „Digitale Magazin“ des Bundesarchivs verwahrt bisher:

- fast ausschließlich: *Schutz-* und *Ersatzdigitalisate* von analogem Archivgut
- kaum dagegen: *genuin digitale* Unterlagen
 - Ausnahme: Filme aus der Filmförderung
 - so gut wie keine Unterlagen aus Verwaltungskontext

Gründe:

- Digitalisierungsstrategie/ Massendigitalisierung/ Filmdigitalisierung
- vergleichsweise geringeres Übernahmevermögen an genuin digitalen Verwaltungsunterlagen/ vorläufige Archivierungsinfrastrukturen/ DZAB als „Puffer“

Speicherbedarfe/ Speichermengen

- Gesamtmenge: 5.681 TB = **5,7 PB**
- Bereich Akten/ Töne/ Großformate: **737 TB**
- Bereich Film: 4.943 TB = **4,9 PB**
- täglicher Zuwachs: 12 TB

→ ohne Berücksichtigung der 4fach redundanten Speicherung

→ ohne Berücksichtigung der Speicherinfrastrukturen außerhalb des Digitalen Magazins

Unter Berücksichtigung der 4fach redundanten Speicherung liegen derzeit **19,7 PB** im DM.

Der Ingest in das Digitale Magazin des Bundesarchivs

Allgemeine Charakteristik – Voraussetzungen – Ablauf

- „*Ingest*“ = strukturierter Prozess zur Verbringung digitaler Unterlagen auf den Langzeitspeicher, in eigener Terminologie: „Bandsicherung“
- *nicht softwaregestützt* bzw. „*automatisiert*“, sondern *skriptbasiert* über die Kommandozeile

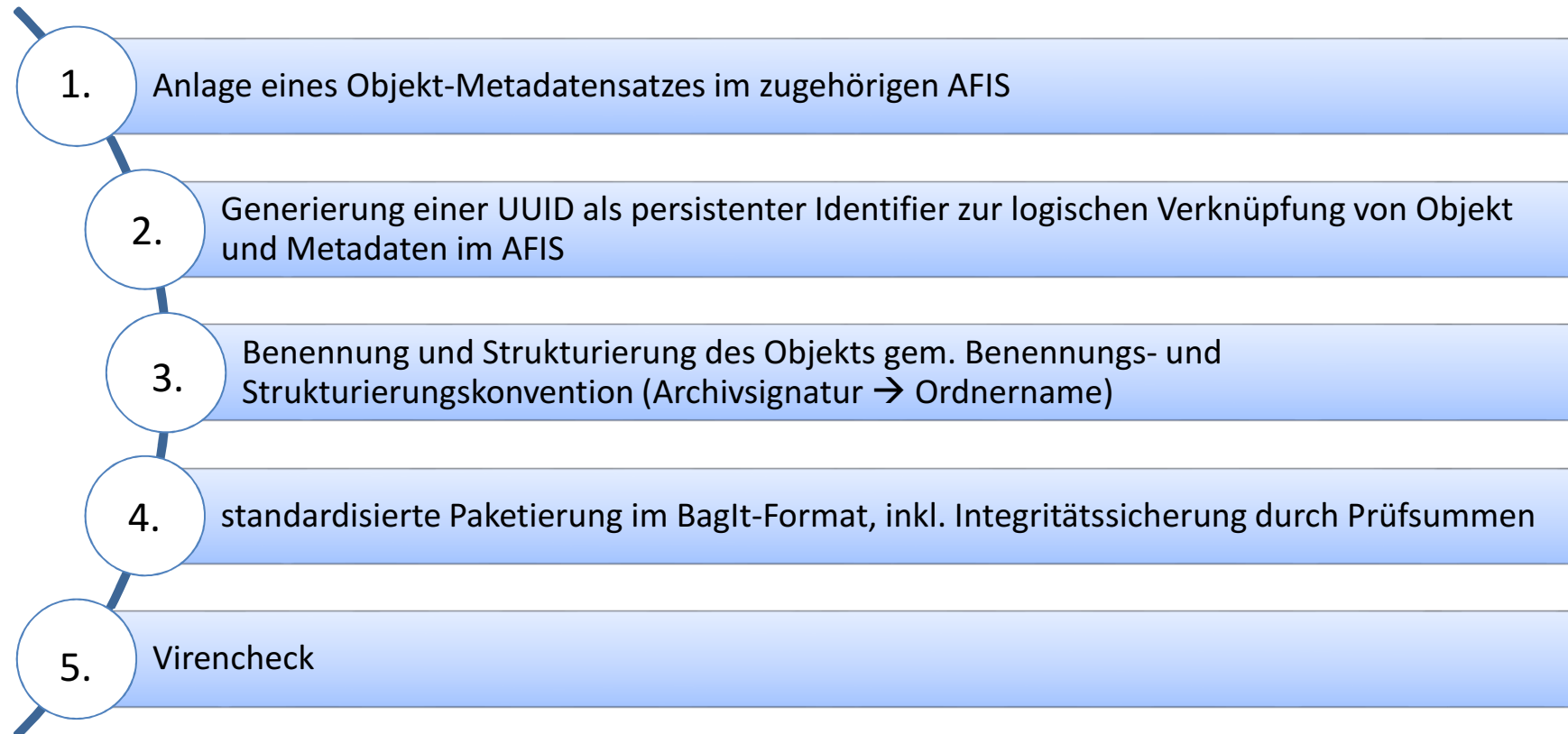
Folgen:

- viel „Handarbeit“, erhöhter zeitlicher Aufwand
- erhöhte Anforderungen an die Qualifikation des Personals (Fachinformatiker)
- gespeichert wird *nur das Objekt*, Metadatenverwaltung erfolgt über die AFISse, ausschließlich *logische* Verknüpfung per UUID
- Grundset an Metadaten wird ins das Bag des Archivierungspakets geschrieben
- bislang *zwei* Ingestwege in Abhängigkeit von AFIS-Anbindung
 - Akten/ Großformate/ Töne (→ AFIS: BASYS-Akte)
 - Film (→ AFIS: BASYS-Film)
- *keine* allgemeine Formatbegrenzung, Begrenzung jedoch durch Format-Validator

Der Ingest in das DM besteht aus 3 / optional 4 Phasen:



Phase 1: Formale und technische Vorbereitung des Ingest



Phase 2 und 3: Preingest und Ingest „Bandsicherung“ (skriptbasiert)

- Phase 2: *Preingest*: Vorbereitung des Ingest bzw. der Bandsicherung (*dm-archive-prepare*)
 - automatisierte und händische Metadatenanreicherung, *keine* Extraktion
 - Kurzvalidierung der Bagstruktur (gebaggte Lieferungen)
 - Erzeugung der Bagstruktur (ungebaggte Lieferungen)

- Phase 3: *Ingest*: Durchführung des Ingest bzw. der Bandsicherung (*dm-archive-store*)
 - Validierung der Bagstruktur, inkl. Integritätsprüfung (aller Prüfsummen)
 - Validierung der Dateiformate
 - Erzeugung eines TAR-Containers (on the fly)
 - Übertragung der Objekte auf den Bandspeicher

Phase 4 (optional): DIP-Erzeugung/ DIP-Ingest (skriptbasiert)

- Erzeugung von Benutzungsrepräsentationen und Übertragung ins Präsentationssystem (*dm-ingest*)
 - Erzeugung von niedrig aufgelösten Benutzungsderivaten (TIFF → JPEG, WAVE → MP3)
 - Transfer ins Präsentationssystem liXe
 - (bisher nur: Akten/ Töne/ Großformate)

Fehlerdetektion und Fehlerbehebung

- Fehler/Probleme: können in jeder Phase des Prozesses oder schon zuvor während der Digitalisierung auftreten

- Fehler- bzw. Problemtypen:
 - Virenbefall
 - fehlerhafte Dateien/ Digitalisate
 - Abweichungen in der Lieferungsintegrität
 - Prüfsummenfehler
 - Fehlschlagen der Formatvalidierung
 - File-Exist-Fehler

- Phase 1: Virencheck, Qualitätssicherung Digitalisierung
- Phase 2 und 3: *automatisierte* Detektion von Standardfehlern
- bislang *keine* automatisierte Fehlerbehandlung, alle Fehler müssen *händisch* (skriptbasiert) bearbeitet werden

Ausblick

Kurz- und mittelfristige Ziele

- Automatisierung der Prozesse → „Knopfdruckingest“/ „Knopfdruckrecall“
- Entwicklung einer eigenen Datenbankanwendung für das Digitale Magazin bzw. die digitale Langzeitarchivierung „BASYS-3-Magazin“ zur Verwaltung der spezifischen Metadaten und Unterstützung der Prozesse
- Modulation weiterer Ingestworkflows, insbesondere für verschiedene Formen genuin digitaler Unterlagen



Präsentation:

Titel: Der Ingest in das Digitale Magazin des Bundesarchivs.-
Bestandsaufnahme und Ausblick

vorgetragen von: Dr. Beatrix Dietel

vorgetragen am: 15. März 2022

Kontaktdaten:

Ansprechpartner/-in: Dr. Beatrix Dietel

Telefon: 0761/47817-845

Email: b.dietel@bundesarchiv.de

Anschrift: Bundesarchiv
Referat IT 4
Wiesentalstr.10
79115 Freiburg