

dhmuc site visit

SIP, AIP, DIP

**– Theorie und Datenmodelle
hinter digitalen Archiven**

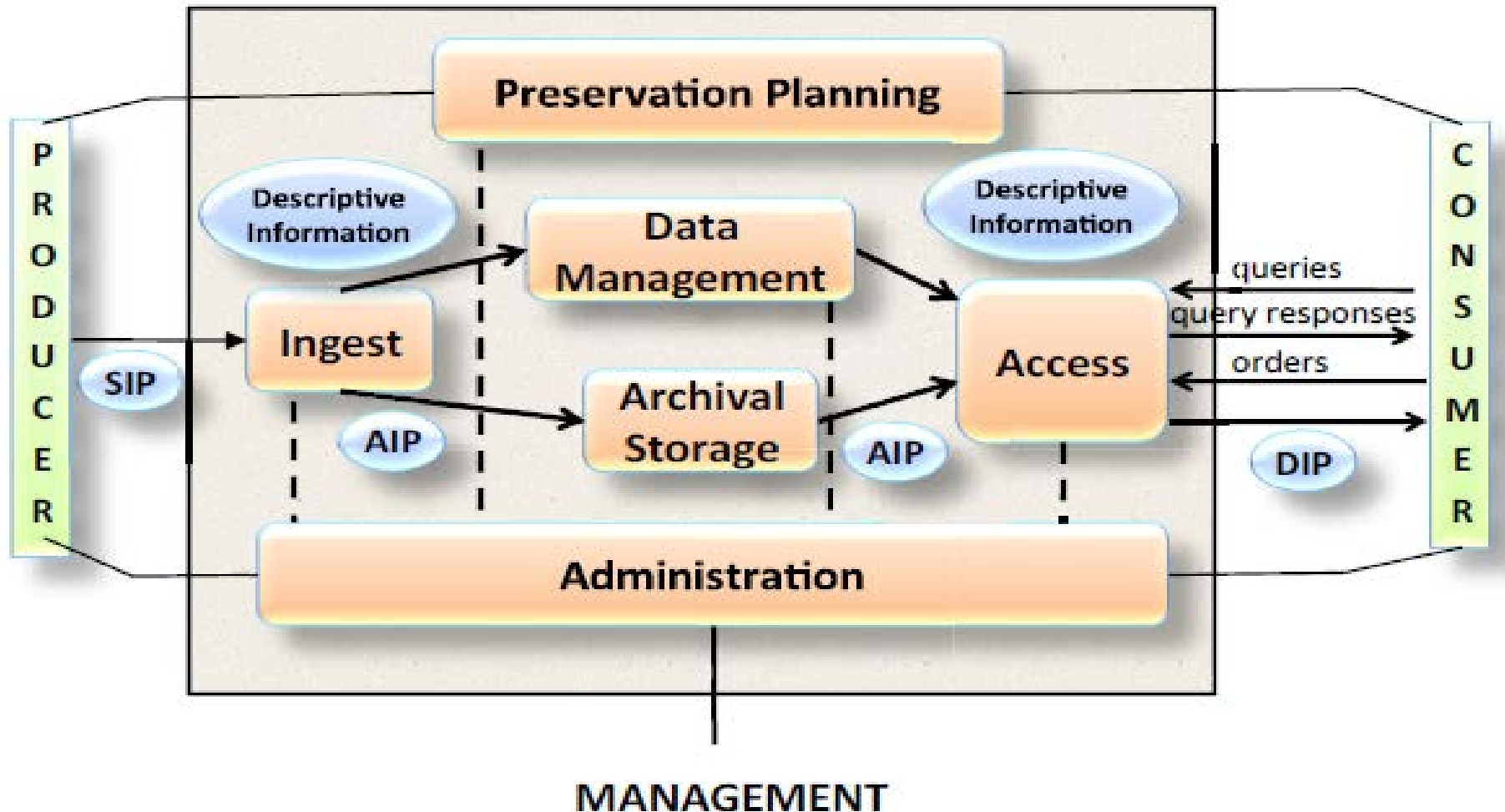
**Dr. Michael Puchta M.A.
Generaldirektion der Staatlichen Archive Bayerns
E-Mail: michael.puchta@gda.bayern.de
Telefon: 089/28638-2484**

Institutionenübergreifende Grundlagen der Langzeitspeicherung

OAIS und PREMIS

- 1. Internationaler Standard OAIS (Open Archival Information System – ISO 14721:2012): Aus der Weltraumforschung stammendes, rein theoretisches Modell zum Aufbau und Betrieb eines digitalen Archivs**
- 2. PREMIS 3.0: Im Bibliotheksbereich entwickelte Metadatenstandard für die Langzeiterhaltung**

Funktionale Bereiche eines OAIS



Langzeitarchivierung **vs.** digitale Archivierung im archivischem Sinne

Die zeitliche Dimension der digitalen Archivierung

Die Staatlichen Archive Bayerns verwahren Unterlagen aus über 1.200 Jahren

- ⇒ Der Erhaltungsauftrag geht von der zeitlichen Dimension deutlich über das IT-Verständnis von „Langzeitspeicherung“ oder „Langzeitarchivierung“ hinaus.**
- ⇒ archivwürdige digitale Informationen sind für alle Zeiten vorzuhalten (= ggf. deutlich länger als Aufbewahrungsbestimmungen in den Abgabebehörden im Sinne von „dauernd“).**
- ⇒ archivierte digitale Informationen „veralten“ nicht, d.h. es findet im Normalfall keine Entfernung mehr aus dem digitalen Archiv statt.**

Die rechtliche Dimension der digitalen Archivierung I

Digitale Archive im Sinne des Legalbegriffs von „Archivierung“ dienen nicht nur Forschung und Wissenschaft (Archive als „Gedächtnis“ ihres Trägers)

Digitale Archive haben bis heute auch **grundlegende Funktionen im Rechtsstaat:**

- **spätere Rückgriffe von Verwaltung und Justiz (z.B. Rechtsterrorismus, Staatsverträge)**
- **Rechtssicherung der Bürgerinnen und Bürger (z.B. Wegerechte, Erbrecht, Entschädigungsansprüche)**

Die rechtliche Dimension der digitalen Archivierung II

Digitale Archive **unterscheiden sich aufgrund ihrer rechtlichen Funktion auch funktional** grundlegend von Repositorien und Langzeitspeicherlösungen im Wissenschaftsbereich:

- Beweiswerterhalt
- Nachvollziehbarkeit jeder Änderung an den digitalen Informationen
- Differenzierte Lösungen zur Einsicht in die Inhalte aufgrund gesetzlicher Schutz- und Geheimhaltungsvorschriften

=> Die Modelle der digitalen Archivierung sind nicht nur technisch bedingt, sondern v.a. rechtlich determiniert!

Praktische Folgen der rechtlichen Determinierung der digitalen Archivierung I

- **Nachvollziehbarkeit / Evidenz des ursprünglichen Zustands sollte dokumentiert werden (Sicherung des urspr. Look-and-Feel über Screenshots, Sicherung der Anwender- und Betriebshandbücher)**
- **Die Formatkonvertierung muss dokumentiert werden (was, wann, wie, warum und mit welchen Folgen)**
- **Erhaltung der qeS bis zur Abgabe (?) an das Archiv**

Praktische Folgen der rechtlichen Determinierung der digitalen Archivierung II

- **Bewusstsein, dass es nicht genügt, die Daten irgendwie zu speichern, recherchierbar zu machen, lesbar zu halten und anzuzeigen, sondern dass es im Kern stets auch um die **Sicherung des Rechtsstaatsprinzips** geht (Art. 20 Abs. 3 GG)**

=> Nachvollziehbarkeit des Verwaltungshandelns, vollständige und wahrheitsgemäße Datenhaltung, Sicherung von Authentizität und Integrität der Daten usw.

Archivrechtliche Grundlagen der digitalen Archivierung

Die theoretischen Grundlagen für das Digitale Archiv fußen auf den weiterhin gültigen Rechtsnormen, die für die Arbeit der Staatlichen Archive Bayerns konstitutiv sind (insbes. BayArchivG, BayUIG, BayDSG, BayStatG, SGB, BArchG, BStatG, Aussond-Bek, Aussond-Bek-FG, Aussonderungs-bekanntmachung Justiz; ArchivBO; Verordnung über die Gliederung der Staatlichen Archive Bayerns => Mandanten sind getrennt)

Archivfachliche Grundlagen der digitalen Archivierung

- 1. Fachkonzept für das Digitale Archiv der Staatlichen Archive Bayerns: 2013 bis 2016 erarbeitet unter Mitarbeit von Archivarinnen und Archivaren sowie Informatikern der GDA, des BAYHStA und verschiedener Staatsarchive (StAAm, StAA, StABa, StAM, StAN, StAWü)**
- 2. Internationale, nationale und verwaltungsinterne Erschließungsrichtlinien für die Metadatenaufnahme im Data Management**

Emulation vs. Formatmigration - Ein Beispiel für die archivische Positionierung im Diskurs um den informationellen Langzeiterhalt

Die Suche nach einem Weg zum Langzeiterhalt des digitalen Archivgutes I

1. Es gibt kein langzeitfähiges Speichermedium im archivischen Sinne. Aber die rein physikalische Erhaltung der Informationen (Bitstream Preservation) ist dennoch durch Refreshing, Spiegelung / Backup, Wechsel der Speichermedien, Hashwertbildung etc. grundsätzlich beherrschbar.
2. Problematisch ist v.a. die Lesbarhaltung der digitalen Archivalien (Content Preservation) aufgrund des schnellen Wechsels der Hard- und Software

Die Suche nach einem Weg zum Langzeiterhalt des digitalen Archivgutes II

3. **Allein an die Staatlichen Archive Bayerns sind potentiell über 3.000 verschiedene IT-Verfahren abgabepflichtig => Unzählige, meist hoch proprietäre Formate werden angeboten**
4. **Jede Erhaltungsmaßnahme an den Daten muss nachvollziehbar und ggf. rollback-fähig sein (Rechtsstaatsprinzip)**

Die Suche nach einem Weg zum Langzeiterhalt des digitalen Archivgutes II

5. Jede Lösung muss aufgrund der zeitlichen (Unendlichkeits)-Perspektive unabhängig von aktuellen Implementierungen vergänglicher Hersteller und Institutionen sein.

Die Lesbarhaltung elektronischer Unterlagen - Lösungsansätze

1. Computermuseum
2. Ausbelichtung auf Mikrofilm
3. Ausdruck auf Papier
4. Emulation
5. Formatmigration

Lösungsansätze im Preservation Planning I

1. Lösung des Problems der Lesbarhaltung durch ein eigenes Computer- / Technikmuseum:

- Ansätze waren teilw. bereits vorhanden durch die Übernahme alter Abspielgeräte für Magnetbänder mit Audio- und Videoinformationen von den Abgabebehörden.
- Scheiterte am schnellen Wechsel und der schieren Menge der Hard- und Software und den fehlenden Möglichkeiten diese wiederum zu erhalten.

Lösungsansätze im Preservation Planning II

2. Ausdruck der digitalen Informationen auf Papier:

- **Allein die Staatlichen Archive Bayerns haben bereits über 200 km an papiernen Archivgut**
- **Auch Papier schafft Erhaltungsprobleme (Papierzerfall) und v.a. können viele digitale Informationen nicht sinnvoll auf Papier abgebildet und ausgewertet werden (GIS- und CAD-Daten, Datenbankinhalte)**

Lösungsansätze im Preservation Planning III

3. Ausbelichtung der digitalen Informationen auf Mikrofilm:

- **Mikrofilm wird in den Archiven als Schutz- und Sicherungsmedium bereits seit Jahrzehnten eingesetzt und garantiert Erhaltung über mehrere hundert Jahre.**
- **Aber auch hier: Viele digitale Informationen können nicht sinnvoll auf Film abgebildet und ausgewertet werden (GIS- und CAD-Daten, Datenbankinhalte)**

Lösungsansätze im Preservation Planning IV

4. Emulation nach Jeff Rothenberg:

- Galt in der archivfachlichen Diskussion bis in das erste Jahrzehnt des neuen Jahrtausends als ernsthaft verfolgte Alternative zur Formatmigration (teilw. eigene Evaluationsprojekte)
- Vorteile: Keine Inhalts- und Funktionsverluste: => zentral für die Forschung, die auf den „intrinsischen Wert“ des Archivguts großen Wert legt. Zudem Hoffnung auf Kostenreduktion und Lösung des „Roll-Back-Problems“

Lösungsansätze im Preservation Planning V

4. Emulation nach Jeff Rothenberg:

Derzeit gilt der Ansatz aber als (vorerst) nicht praktikabel:

- Lizenzprobleme
- Nicht die Quellsoftware, sondern die darin gehaltenen und gepflegten Daten besitzen Archivwürdigkeit
- Zu viele zu emulierende Systeme (im Freistaat Bayern waren 2012 rund 3.000 im Betrieb)
- Technisch teilw. komplexer als erhofft

Lösungsansätze im Preservation Planning VI

5. Modifizierte Formatmigration:

In der archivfachlichen Diskussion zur digitalen Bestandserhaltung (Preservation Management gem. OAIS) wird unter Formatmigration nicht das Umkopieren / Refreshing verstanden, sondern die Konvertierung eines Formats in ein anderes Format.

Heute setzen die Archive auf die Formatmigration und verbinden diese mit den Konzepten des

Repräsentationenmodells und der **Signifikanten Eigenschaften**, beides Ansätze die aus dem PREMIS und damit dem Bibliotheksumfeld stammen.

Die Lesbarhaltung elektronischer Unterlagen - Das Repräsentationenmodell I

Jedes digitale Archivale besteht aus zwei Ebenen:

- 1. Informationsobjekt (IO), das als Beschreibung der Intellectual Entity, der intellektuellen und konzeptionellen Einheit des Archivales verstanden werden muss und für die archivseitig definierte Menge des dauerhaft zu erhaltenden und die Archivwürdigkeit in erster Linie konstituierende Gehalt der digitalen Archivalieneinheit steht. Jedes IO besteht aus 1-n Repräsentationen (R)**

Die Lesbarhaltung elektronischer Unterlagen - Das Repräsentationenmodell II

Jedes digitale Archivalie besteht aus zwei Ebenen:

- 2. Unterhalb der Ebene des Informationsobjekts befinden sich eine oder mehrere Repräsentationen. Jede Repräsentation enthält dabei für sich die zur Wiedergabe des Informationsobjekts erforderlichen Dateien. => Für jede Archivalie gibt es 1-n technische Ausprägungen**

Die Lesbarhaltung elektronischer Unterlagen - Das Repräsentationenmodell III

Vorteile des Repräsentationenmodells:

- **Nachvollziehbarkeit des archivischen Handelns und damit Gewährleistung der Vertrauenswürdigkeit des Digitalen Archivs**
- **Langfristige Unterstützung des Preservation Planning durch Rückgriffmöglichkeit auf ältere Migrationsstufen**

Die Formatmigratoron in Verbindung

mit dem Repräsentationenmodell I

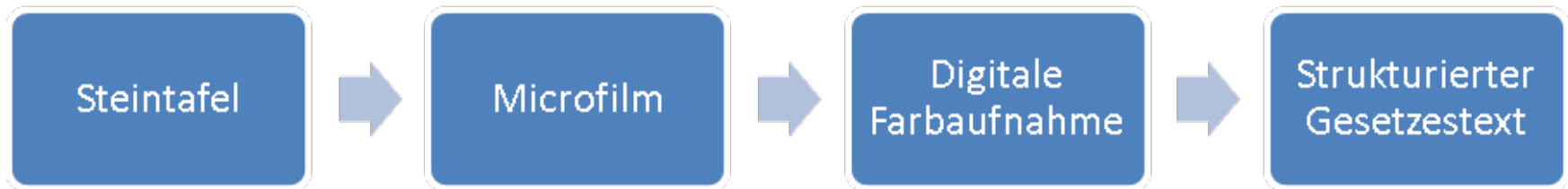
Beispiel

Die 10 Gebote - von Jehova auf Steintafeln direkt an Moses übergeben: Zweifelsohne archivwürdig nach Herkunft und Bedeutung für die Rechtsicherheit der Menschen.

Die Gebote selbst sind die Intellektuelle Entität, ihr Trägermaterial (Steintafeln) ist die zugehörige Repräsentation.

Die Formatmigraton in Verbindung mit dem Repräsentationenmodell II

Beispiel



Die Intellektuelle Entität bleibt konstant, die Erscheinungsform (Repräsentation) wandelt sich im Laufe der Jahrhunderte und nimmt zu.

Definition Signifikanter Eigenschaften

Jede Formatmigration kann potentiell zu inhaltlichen und funktionalen Verlusten der digitalen Überlieferung führen => Bereits im zeitlichen Kontext der Aussonderung werden von den Aussonderungsreferentinnen und –referenten diejenigen „signifikanten“ Eigenschaften einer digitalen Abgabe im Rahmen der Erschließung des SIPs erfasst, die dauerhaft, über alle Formatmigrationen erhalten bleiben müssen.

Das Konzept der Signifikanten Eigenschaften I

- **Kein neuer Ansatz, sondern im Bereich der analogen Restaurierung und Bestandserhaltung längst etablierte Praxis – das war nur den Archiven selbst nie so bewusst!**
- **Neu ist allerdings die Dimension und die Beschaffenheit des betroffenen Archivguts**

Das Konzept der Signifikanten Eigenschaften II

- Es werden von Anfang an mit den **Signifikanten Eigenschaften** Informationen und Funktionalitäten der digitalen Archivalie definiert, die über alle Migrationsmaßnahmen hinaus zwingend erhalten bleiben müssen.
- Das Idealziel ist es das Vorhandensein dieser **Signifikanten Eigenschaften** nach Migrationszyklen automatisiert überprüfen zu können.

Schwächen des Konzepts der Signifikanten Eigenschaften I

- Das Konzept der **Signifikanten Eigenschaften** steht im Widerspruch zum archivfachlichen Diskussionsstand: Die Archive sind bei analogen Archivalieneinheiten stets bestrebt, sämtliche daran haftenden Informationen zu erhalten, da nicht vorausgesehen werden kann, welche gesellschaftlichen, wirtschaftlichen, rechtlichen und politischen Entwicklungen die Zukunft bringt und welche Forschungsinteressen künftige Generationen haben werden.

Schwächen des Konzepts der Signifikanten Eigenschaften II

- **In der Papierwelt gilt es nicht nur die Primärinformation, sondern auch den indirekte, "intrinsischen" Informationswert (Menne-Haritz) der Archivalie zu erhalten. Nur in absoluten Ausnahmefällen, wenn das (analoge Trägermaterial) aus konservatorischen oder lagerungstechnischen Gründen nicht erhalten werden kann, findet im Archiv eine Übertragung der Archivalie auf ein Ersatzmedium und damit ein Verlust intrinsischer Informationen statt.**

Schwächen des Konzepts der Signifikanten Eigenschaften III

- Sowohl in der analogen Welt als auch in der elektronischen Umgebung hängt die Archivwürdigkeit im Regelfall an der gesamten Archivalie – auch, wenn primär der lesbare Inhalt die Bewertungsentscheidung über die Archivwürdigkeit bestimmt.

Schwächen des Konzepts der Signifikanten Eigenschaften IV

- **Auch der intrinsische Wert eines Dokuments, also zum Beispiel die Beschaffenheit des Trägermaterials, die verwendete Tinte, die Bindung der Akte usw. können weitreichende Aussagen über den Entstehungszusammenhang und die Überlieferungsgeschichte ermöglichen und sind für die hilfswissenschaftliche und historisch-kritische Forschung (z.B. Echtheitsfragen) von großer Bedeutung.**

Schwächen des Konzepts der Signifikanten Eigenschaften V

- Die Erhaltung der ersten R einer IE ist hierfür ein Ansatz, der aber mit Blick auf die sinkende Wahrscheinlichkeit der Lesbarmachung alter Repräsentationen mit zunehmendem Zeitablauf durch die Übernahme von Dokumentationsmaterial zum Quellsystem, Screenshots zum ursprünglichen „Look-and-Feel“ sowie die selektive Erfassung ausgewählter ergänzender Metadaten (z. B. Abspielzeit des „Originals“) flankiert werden muss.

Folgen der Formatmigration als Erhaltensstrategie für die archivwissenschaftliche Diskussion

- Entwicklung der (umstrittenen) Theorie der **Archivfähigkeit** => Sind nicht mehr lesbare Formate automatisch nicht archivwürdig?
- Wiederbelebung der längst abgeschlossenen Diskussion über die **Prognosemöglichkeit künftiger Nutzungsszenarien** anhand der Signifikanten Eigenschaften



Danke für die Aufmerksamkeit!