

***DO's and DONT's***  
***Entscheidungsfindung für***  
***eine Softwarelösung***  
***E-Infrastructure, Wien 21.4.2016***  
***Bettina Kann***

## Guideline to Digital Preservation

Diese Guideline soll Hilfestellung für die Langzeitarchivierung von Daten und Objekten im Kontext des Publikations- und Forschungswesens bieten. Sie ist ausdrücklich nicht für den Kontext der Compliance anwendbar. Sie soll dazu ermächtigen, die richtigen Fragen zur Auswahl der für die eigene Institution geeigneten Langzeitarchivierungs-Lösung zu stellen und bei der Entscheidung für eine System-Lösung behilflich sein.

Langzeitarchivierungssysteme werden hier als Systeme verstanden, die im Workflow hinter einem Repositorium stehen, in dem digitale Objekte und ihre Metadaten gespeichert und angezeigt werden sowie recherchierbar sind.

Allfällige Begriffserklärungen finden Sie im Glossar des Clusters C.

### INHALT

Warum will ich Langzeitarchivierung (LZA) betreiben?.....	2
Datenerhebung / Was will ich langzeitarchivieren?.....	2
Gibt es an meiner Institution bereits ein Institutional Repository (IR)?.....	6
Grundsätzliche Anforderungen an ein Langzeitarchivierungs-System.....	6
Software-Modelle / Welches Implementierungsmodell kommt für mich in Frage?.....	7
Organisatorisches.....	11
Workflow.....	13
Rechtliche Fragen.....	13

### Warum will ich Langzeitarchivierung (LZA) betreiben?

- Gibt es externe (z.B. gesetzliche/rechtliche/vertragliche) oder Vorgaben zur Archivierung (Pflichtexemplare, evtl. Verwaltung Verträge mit Fördergebern)?
- Herrscht bei allen Stakeholdern übereinstimmendes Bewusstsein für LZA?

Die Langzeitarchivierung ist gekennzeichnet durch einen Zeitrahmen, der in einigen Fällen Jahre, in anderen Jahrzehnte oder Jahrhunderte lang sein kann. In diesem Zeitrahmen verändern sich nicht nur die Dateiformate, sondern auch die Anwendungsprogramme, Betriebssysteme und die Speichermedien.

### Datenerhebung / Was will ich langzeitarchivieren?

#### Welche Daten habe ich?

- Publikationen
- Hochschulschriften (wissenschaftliche Abschlussarbeiten)
- Digitalisate alter Handschriften, Bücher, Zeitschriften, Urkunden
- künstlerische/audiovisuelle Werke
- Digitale Materialien der Architektur und des Industrial Design
- angekaufte elektronische Ressourcen: E-Books, E-Journals
- Web-Seiten
- Forschungsdaten

Forschungsdaten sind Daten, "die im Zuge wissenschaftlicher Untersuchungen, Digitalisierung, Quellenforschungen, Experimente, Messungen, Erhebungen, Befragungen entstehen."<sup>12</sup>

<sup>12</sup> <http://www.ibwissen.info/definition/lexikon/langzeitarchivierung-long-term-digital-preservation>



**Bundesrecht konsolidiert: Gesamte Rechtsvorschrift für Mediengesetz, Fassung vom 19.04.2016**

 [Druckansicht](#)

Andere Formate: 

**Langtitel**

Bundesgesetz vom 12. Juni 1981 über die Presse und andere publizistische Medien (Mediengesetz - MedienG)  
StF: [BGBl. Nr. 314/1981](#) (NR: GP XV [RV 2](#) [AB 743](#) [S. 79](#), BR: [2350](#) [AB 2351](#) [S. 412](#).)

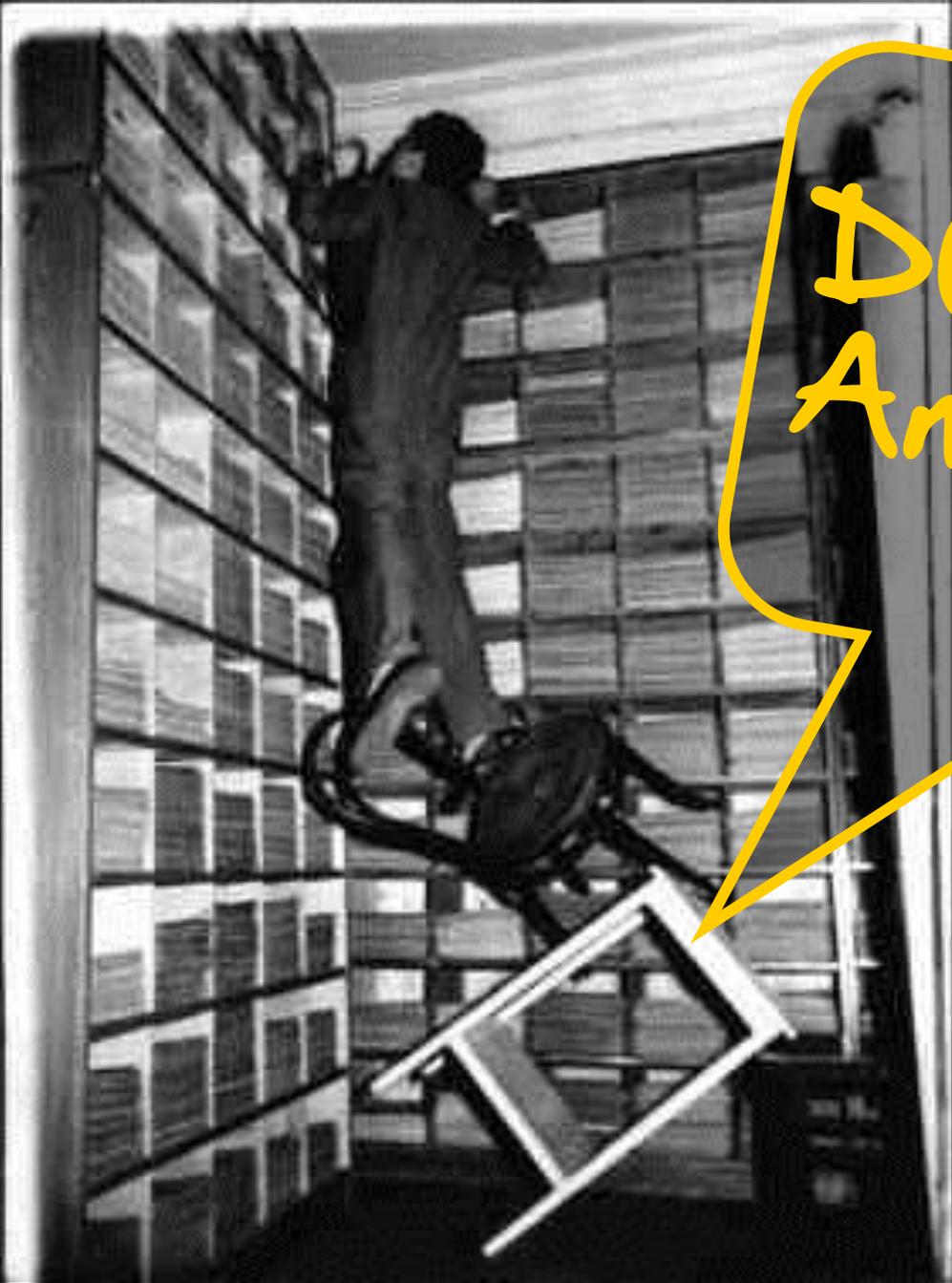
**Änderung**

- [BGBl. Nr. 211/1987](#) (VfGH)
- [BGBl. Nr. 233/1988](#) (NR: GP XVII [RV 451](#) [AB 532](#) [S. 57](#), BR: [AB 3461](#) [S. 500](#).)
- [BGBl. Nr. 20/1993](#) (NR: GP XVIII [RV 503](#) [AB 851](#) [S. 95](#), BR: [4401](#) [AB 4411](#) [S. 563](#).)
- [BGBl. Nr. 91/1993](#) (NR: GP XVIII [RV 715](#) [AB 775](#) [S. 101](#), BR: [4477](#) [AB 4467](#) [S. 564](#).)
- [BGBl. I Nr. 105/1997](#) (NR: GP XX [RV 49](#) [AB 812](#) [S. 82](#), BR: [5491](#) [AB 5506](#) [S. 629](#).)
- [BGBl. I Nr. 75/2000](#) (NR: GP XXI [RV 98](#) [AB 262](#) [S. 34](#), BR: [AB 6172](#) [S. 667](#).)
- [BGBl. I Nr. 130/2001](#) (NR: GP XXI [RV 754](#) [AB 787](#) [S. 81](#), BR: [6457](#) [AB 6481](#) [S. 681](#).)
- [BGBl. I Nr. 136/2001](#) (NR: GP XXI [RV 742](#) [AB 824](#) [S. 81](#), BR: [6458](#) [AB 6459](#) [S. 681](#).)
- [BGBl. I Nr. 49/2005](#) (NR: GP XXII [RV 784](#) [AB 874](#) [S. 110](#), BR: [AB 7292](#) [S. 722](#).)
- [BGBl. I Nr. 151/2005](#) (NR: GP XXII [RV 994](#) [AB 1077](#) [S. 122](#), BR: [AB 7387](#) [S. 725](#).)
- [CELEX-Nr. [32003L0006](#)]
- [BGBl. I Nr. 112/2007](#) (NR: GP XXIII [RV 299](#) [AB 335](#) [S. 41](#), BR: [7802](#) [AB 7851](#) [S. 751](#).)
- [BGBl. I Nr. 8/2009](#) (NR: GP XXIV [RV 20](#) [AB 41](#) [S. 11](#), BR: [AB 8044](#) [S. 765](#).)
- [BGBl. I Nr. 131/2011](#) (NR: GP XXIV [AB 1608](#) [S. 137](#), BR: [AB 8636](#) [S. 803](#).)
- [BGBl. I Nr. 50/2012](#) (NR: GP XXIV [RV 1726](#) [AB 1757](#) [S. 153](#), BR: [AB 8715](#) [S. 808](#).)
- [BGBl. I Nr. 101/2014](#) (NR: GP XXV [IA 718/A](#) [AB 439](#) [S. 53](#), BR: [AB 9279](#) [S. 837](#).)



**Sonstige Textteile**

Der Nationalrat hat beschlossen:



DO: IST-  
Analyse

Datenerhebung  
und -analyse

→ Datenerhebung: welche Daten habe ich aktuell und in Zukunft → Alles LZA?

→ Datenformate und Struktur: aktuell und im zukünftigen LZA-System.

# Datenerhebung und -analyse

Audio/Video  
(Wav,mpeg  
...)

Applikationen

Textdokumente  
(doc,odf,pdf,txt  
...)

Tabellen/Statistik  
(ods, xls,sas,spss  
...)

Strukturierter Text  
(html,json,xml,  
tex...)

Forschungsdaten

Grafiken,  
Bilder

Datenbanken  
(Access,MySql,  
Oracle ...)

2-D/3-D-  
Entwürfe  
(CAD)

Source Code

Wenige (offene)  
Formate → wenige(r)  
Probleme  
DO: Policy, welche  
Formate erlaubt sind.

Datenerhebung  
und -analyse

## SPEICHER

- Zentral auf Speichersysteme der Institution
- Lokal auf PCs/externen Festplatten
- Cloud
- CDs/DVDs
- ...

Datenerhebung  
und -analyse

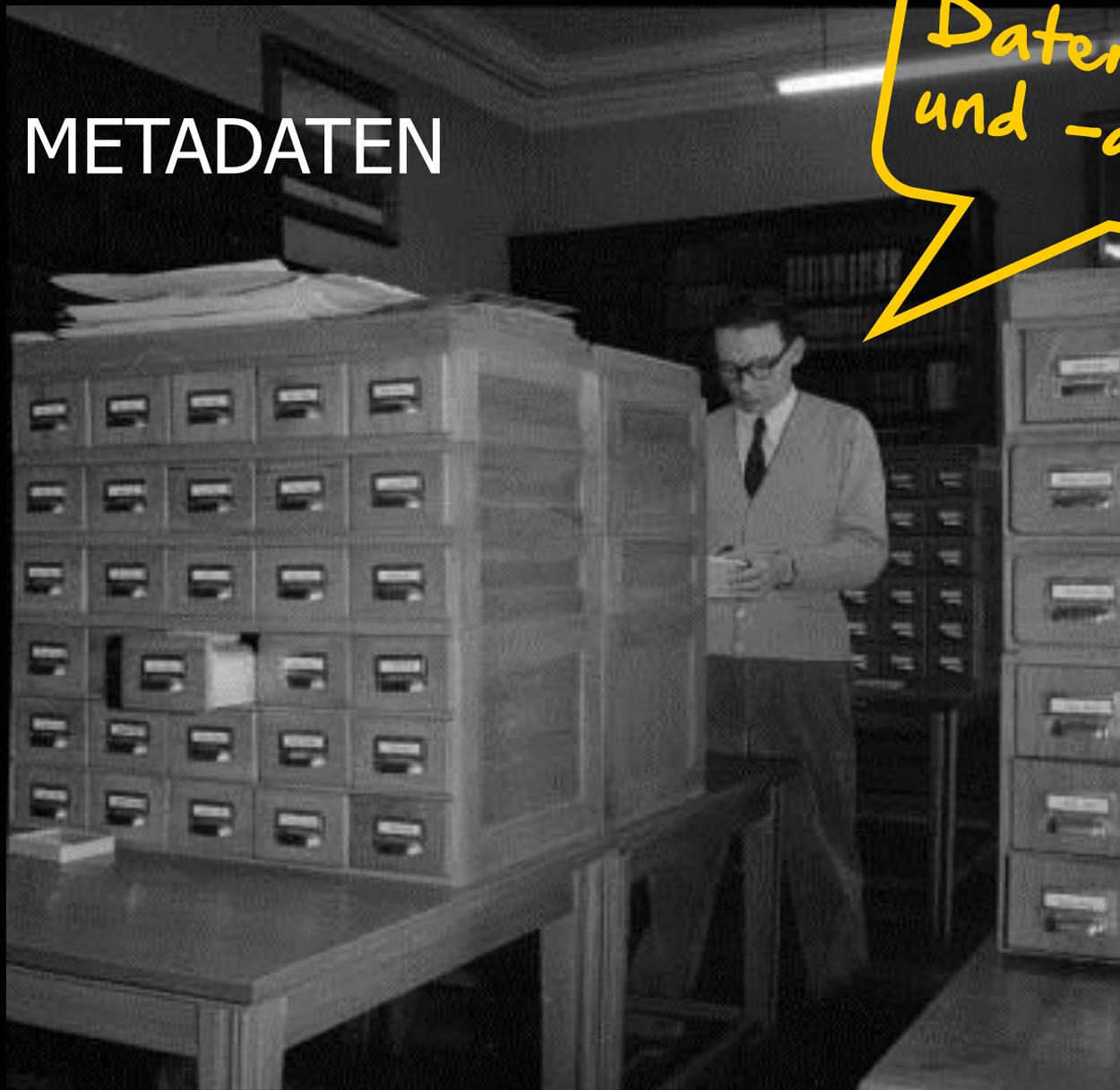
# SPEICHER

---

- Belegter Speicherplatz
- Zuwachs?
- Benötigter Speicherplatz

# METADATEN

Datenerhebung  
und -analyse





Datenerhebung  
und -analyse

WO werden digitale Daten derzeit  
archiviert?

→ Repository vorhanden?

DO:  
Workflowanalyse

- ▣ [Ink 3.B.14](#)
  - ▣ [Band 1](#)
    - ⊕ [Einband](#)
    - ⊕ [Davor](#)
    - ▣ [Altes Testament](#)
      - ⊕ [Prologus S. Hieronymi – Vorrede](#)
      - ⊕ [Liber Genesis – 1. Buch Mose](#)
      - ⊕ [Liber Exodi – 2. Buch Mose](#)
      - ⊕ [Liber Levitici – 3. Buch Mose](#)
      - ⊕ [Liber Numerorum – 4. Buch Mose](#)
      - ⊕ [Liber Deuteronomii – 5. Buch Mose](#)
      - ⊕ [Liber Iosue – Josua](#)
      - ⊕ [Liber Iudicum – Richter](#)
      - ⊕ [Liber Ruth – Ruth](#)
      - ⊕ [Liber primus Regum \(Samuhelis\) – 1. Buch...](#)
      - ⊕ [Liber secundus Regum \(Samuhelis\) – 2. Buch...](#)
      - ⊕ [Liber tertius Regum \(Malachim\) – 1. Buch...](#)
      - ⊕ [Liber quartus Regum \(Malachim\) – 2. Buch...](#)
      - ⊕ [Liber primus Paralipomenon \(Verba...\)](#)
      - ⊕ [Liber secundus Paralipomenon \(Verba...\)](#)
      - ⊕ [Oratio Manasse](#)
      - ⊕ [Liber primus Ezrae – 1. Buch Esra](#)
      - ⊕ [Liber secundus Ezrae – Nehemia \(2. Buch...](#)
      - ⊕ [Liber tertius Ezrae – 3. Buch Esra](#)
      - ⊕ [Liber quartus Ezrae – 4. Buch Esra](#)
      - ⊕ [Liber Tobiae – Tobit](#)
      - ⊕ [Liber Iudith – Judith](#)
      - ⊕ [Liber Hester – Esther](#)
      - ⊕ [Liber Iob – Hiob](#)
      - ⊕ [Liber Psalmorum – Psalmen](#)

Zeige: 

LZA only?  
Präsentation?



DIE neue LZA-Software

Grundsätzliche Auswahlkriterien  
für ein LZA-System

# Implementierungsmodell

- Proprietär vs. Open Source
- Onsite vs. SaaS
- Kauf vs. Eigenentwicklung
- Mischform

# Implementierungsmodell

- Referenzimplementierungen?
- Institutionelle Vorgaben?
- Mandantenfähigkeit?
- Know-How in der Institution?

OAIS

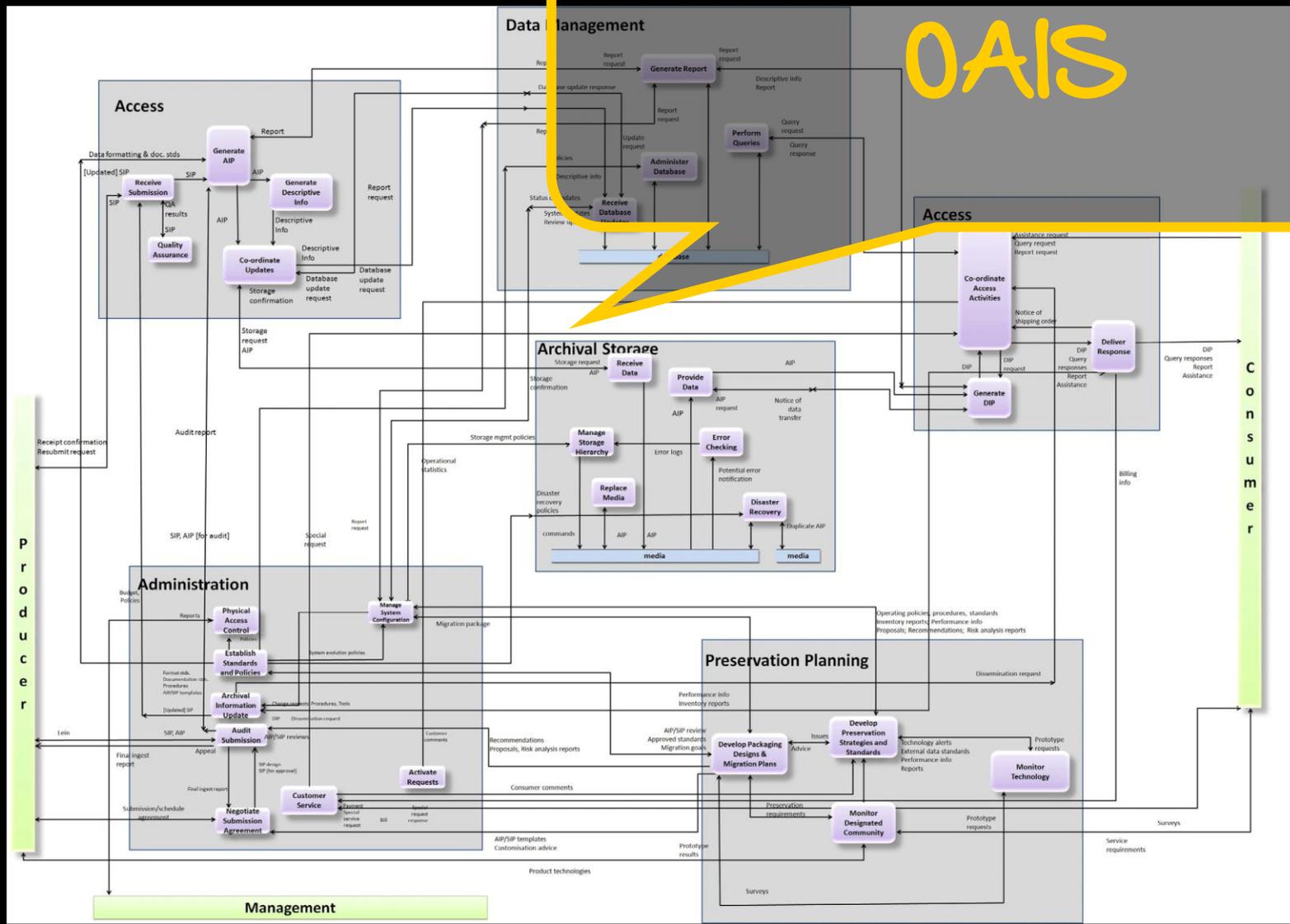


Figure A-1: Composite of Functional Entities

# Datensicherheit

- Datenintegrität
- Virencheck
- Hashvalidierung
- Backup
- Redundante Datenhaltung?
- Data Recovery

# Systemadmin

- Erforderliche Hardwareressourcen
- Skalierbarkeit
- Monitoring
- Datenbank und Betriebssystem (Policy?)
- Unterschiedliche Speicherklassen?
- Bandbreite?

# Metadaten

- Standards
- Technische
- Inhaltliche
- Strukturelle
- Administrative → DO: klären der rechtlichen Bedingungen für Zugriff auf Objekte. Verankerung in Metadaten

# Datenmodell

- Vorhanden?
- Einheitlich vs. flexibel
- Verständlich?

Wenn vorhanden und prinzipiell brauchbar →  
DON'T: neu entwickeln, adaptieren, "verbessern"

# Datenmodell

AIP und DIP:

DO: Trennen von Archivobjekt und  
Präsentationsobjekt

- Mehr Flexibilität bei Präsentation der Objekte
- Möglichkeit, Präsentation komplett von LZA zu trennen.

# Workflow

- Ingest Automatisiert vs. manuell → (Programmier)Aufwand, um SIPs zu erzeugen?
- Datenzugriff (Personen, Applikationen)
- Anbindung externer Applikationen

# LZA- Funktionalitäten

- Formatvalidierung
- Unterstützung von Formatmigrationen inkl. Versionsbeziehungen
- Unterstützung für Preservation Planning
- Persistent Identifier?
- Löschungen erlaubt/protokolliert?

Finanzierung und  
Personal

Vielen Dank für  
Ihre  
Aufmerksamkeit!

