


# Wege ins Archiv

Der nestor Leitfaden für die  
Informationsübernahme

St. Gallen, 27. April 2009

Jens Ludwig

[ludwig@sub.uni-goettingen.de](mailto:ludwig@sub.uni-goettingen.de)

 NIEDERSÄCHSISCHE STAATS- UND  
UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN



- Hintergrund des Leitfadens
- Aufbau des Leitfadens
  - Objekte
  - Prozesse
  - Management
- Erfahrungen und Weiterentwicklungen

- **Hintergrund des Leitfadens**
- Aufbau des Leitfadens
  - Objekte
  - Prozesse
  - Management
- Erfahrungen und Weiterentwicklungen

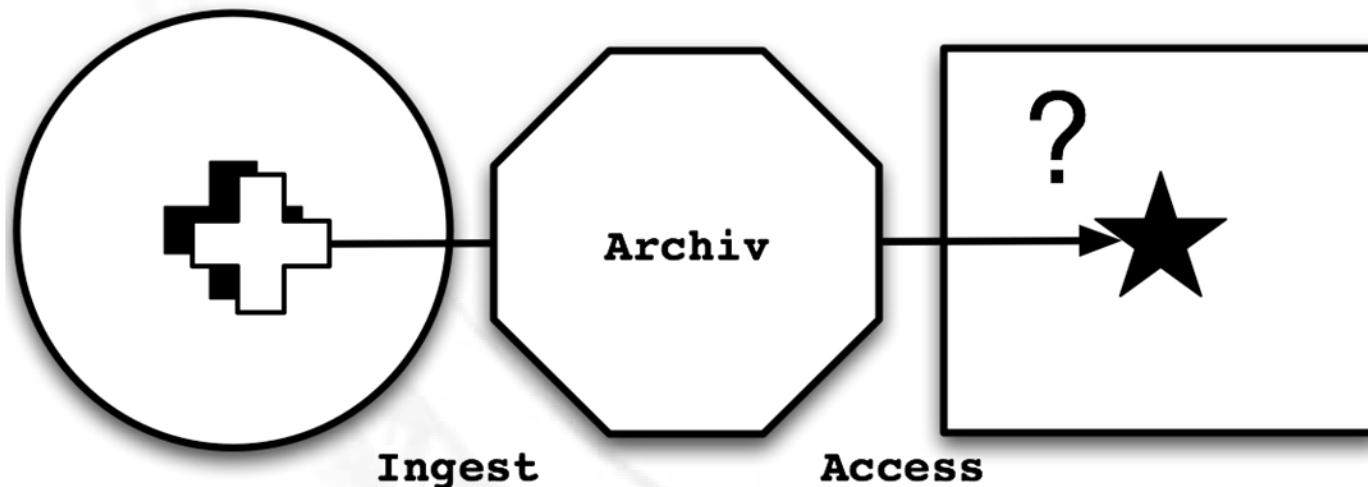
- nestor-Arbeitsgruppe „Standards für Metadaten, Transfer von Objekten in digitale Langzeitarchive und Objektzugriff“
  - primär Bibliotheken und Archive
  - Leitung zusammen mit Herrn Huth
  - diverse Mitglieder im Publikum...
- „Wege ins Archiv - Ein Leitfaden für die Informationsübernahme in das digitale Langzeitarchiv“
  - Entwurf zur öffentlichen Kommentierung
  - Unterstützung durch die Initiative INS des BMWI und DIN

# Wieso ist es wichtig?

- Beagrie et al 2008: Alle UK Datenzentren betrachten den Ingest als größten Kostenfaktor. Z.B. UK Data Archive:
  - Acquisition and Ingest: 42%
  - Archival Storage and Preservation: 23%
  - Access 35%
- Aber: Teile der Ingest-Aufgaben (z.B. Metadaten) später durchzuführen ist noch teurer.
- Ingest ist der erste Schritt: Er sollte leicht fallen, um Masse und Schwung für die nächsten zu gewinnen.
- „Garbage in – Garbage out“: Das dLZA kann Anfangsfehler selten verbessern.

# Wieso ist es schwierig?

- Objekte kommen aus verschiedenen tech. und orga. Kontexten und Prozessen. Und sollen später in anderen verwendbar sein.



- Die späteren Verwendungen müssen schon am Anfang berücksichtigt werden. Welche Eigenschaften sollen erhalten werden? Welche Metadaten sind notwendig?

- Häufige Frage aus der Praxis: Wie kann ich meine Objekte für die LZA „abgeben“?
- Akteure beim Ingest:
  - Gedächtnisinstitutionen oder allgemeiner:  
„digitale Langzeitarchive“
  - Produzenten und Lieferanten von Informationen
- Wichtig, schwierig und nachgefragt: Deshalb ein Leitfaden für effizientere Praxis und als gemeinsame Arbeitsbasis von dLZA und Produzent.

- Nicht OAIS-Terminologie
  - da abstrakt, voraussetzungsreich und z.T. zu granular
  - aber kompatibel und Ausgangspunkt (PAIMAS)
- Statt „Ingest“ „Informationsübernahme“
  - Nicht nur technischer Transfer, sondern primär der Prozess der Verantwortungsübernahme.
- Kein fester Ablauf und Kriterien
  - wegen komplexer Abhängigkeiten
  - stattdessen prägnante Ziele und Erläuterungen mit Beispielen



- Hintergrund des Leitfadens
- **Aufbau des Leitfadens**
  - Objekte
  - Prozesse
  - Management
- Erfahrungen und Weiterentwicklungen

- Objekte
  - Auswahl der zu archivierenden Information
  - Auswahl der Metadaten
  - Wesentliche Eigenschaften
- Prozesse
  - Transferpakete
  - Validierung
  - Transfer von Daten
- Management
  - Recht und Vertrag
  - Übernahmevereinbarung und Dokumentation
  - Teilbereiche des Managements

- Hintergrund des Leitfadens
- **Aufbau des Leitfadens**
  - **Objekte**
  - Prozesse
  - Management
- Erfahrungen und Weiterentwicklungen

# Auswahl der zu archivierenden Informationen

- Intellektuelle Entitäten auswählen: Was soll archiviert werden? (Üblicherweise nicht Dateien, sondern was die Dateien repräsentieren: Dokumente, Fotos, Messergebnisse...)
- Exportform analysieren: In welcher Form liefert der Produzent die Objekte? Ist das Format akzeptabel?

- Definition der benötigten Metadaten: Was muss bekannt sein, um die Objekte in einem anderen Kontext wieder zu benutzen? Z.B. nach OAIS
  - deskriptive Metadaten (zum Auffinden der Objekte)
  - representation Information (technische und inhaltliche Hintergrundinformationen zum Verständnis der Objekte)
  - reference information (Identifikatoren, etc)
  - context information (Beziehung der Objekte zu anderen)
  - provenance information (Was ist mit den Objekten geschehen?)
  - fixity information (Checksummen zum Test der Integrität)
- Verantwortung für Bereitstellung der Metadaten klären

# Identifikation der wesentlichen Eigenschaften

- Definition der wesentlichen Eigenschaften: Welche Aspekte sollen unbedingt erhalten werden? Z.B. nach InSPECT-Projekt:
  - content, eg. text, image, slides, etc.
  - context, eg. who, when, why.
  - appearance, eg. font and size, colour, layout, etc.
  - structure, eg. embedded files, pagination, headings, etc.
  - behaviour, eg. hypertext links, updating calculations, active links, etc.
- Erhebung der wesentlichen Eigenschaften

- Hintergrund des Leitfadens
- **Aufbau des Leitfadens**
  - Objekte
  - **Prozesse**
  - Management
- Erfahrungen und Weiterentwicklungen

# Definition von Transferpaketen

- Verhältnis zwischen Informationsobjekt und Transferpaketen: Wie werden die Objekte ans dLZA geschickt? 1 Objekt = 1 Paket?  $n$  Objekte =  $m$  Pakete?
- Rekonstruierbarkeit der Informationsobjekte: Wie gehören die Bestandteile eines „verpackten“ Objektes zusammen? Z.B. welche Metadaten zu welcher Datei? (Transferpakete sind die gemeinsame Sprache des dLZAs und Produzenten)
- Identifikation der Transferpakete: Kann anders als die Objektidentifikation sein.



- Definition der Validierungen: Welche Test sind durchzuführen? Z.B.
  - Vollständigkeit (Sind es alle Objekte?)
  - Integrität (Sind alle Objekte unverändert angekommen?)
  - Validität der Daten (Sind die Objekte inhaltlich/tech. korrekt?)
- Erfüllungsgrad und Konsequenzen bei Nichterfüllung
- Beteiligte Personen und Hilfsmittel
- Zeitliche Abfolge

- Gesetzliche/vertragliche Rahmen: Wie soll der Transfer ablaufen? Sind besondere Sicherheitsmaßnahmen notwendig? Z.B.
  - offline Übertragung
  - Verschlüsselung
  - digitale Signaturen
- Technische und organisatorische Möglichkeiten
- Definition und Test der Transferarbeitsschritte

- Hintergrund des Leitfadens
- **Aufbau des Leitfadens**
  - Objekte
  - Prozesse
  - **Management**
- Erfahrungen und Weiterentwicklungen

- Identifikation juristischer Personen und Akteure
- Verpflichtungen bzgl Archivmaterial sind bekannt
- Urheberrecht ist festgestellt
- Regelung bzgl Urheberrecht
- Gewährleistung und Haftung

# Übernahmevereinbarung und Dokumentation

- Dokumentation
  - der Ergebnisse aller Planungen (Übernahmevereinbarung)
  - der Durchführung (Protokoll, insbesondere Abweichungen)
- Ermöglicht Nachvollziehbarkeit
- Trägt zur Authentizität und Integrität der Objekte bei
- Trägt zur Vertrauenswürdigkeit der Institutionen bei
- Muss selbst auch erhalten werden (z.B. als Provenienzmetadaten)

# Teilbereiche des Managements

- Gesamtblick, Angemessenheit der Ziele und der Mittel
- Bereitstellung der Mittel
- Kontrolle der Umsetzung und Wirksamkeit
- Teilbereiche:
  - Qualität und Sicherheit
  - Vorgaben und Standards für alle Aufgaben
  - Prozesse: Integration in Arbeitsabläufe (Wann und an welcher Stelle erfolgt LZA?)
  - Kosten und Risiken: Konsequenzen der (Nicht-)Durchführung?

- Hintergrund des Leitfadens
- Aufbau des Leitfadens
  - Objekte
  - Prozesse
  - Management
- **Erfahrungen und Weiterentwicklungen**

Vorstellung auf Workshop beim Bundesarchiv:

- Viele Teilnehmer (primär Bib. und Archive)
- Bedarf in der Praxis vorhanden
- Kürze des Leitfadens gut
- Perspektive anderer Communities reinbringen (Museen, aber evtl auch Industrie, Wissenschaft?)
- Mehr Beispiele und Sammlung von Umsetzungen/Case-Studies gewünscht
- Einige Vorschläge zu Verbesserungen im Detail...



Vorstellung im Rahmen der nestor Spring School 2009:

- Bedarf in der Praxis vorhanden
- Übung als Rollenspiel: dLZA und Produzent müssen trotz verschiedener Interessen in unterschiedlichen Szenarien den Ingest organisieren
- Ergebnis:
  - Die Teams auf beiden Seiten müssen Bewusstsein der Aufgabe haben, Schulung ggf sinnvoll.
  - Technische und rechtliche Fachkenntnisse in Teams wären hilfreich.

Mitarbeit an DRIVER Report:

- „Enhanced Publications“ (EP): komponierte Objekte aus Publikation und anderen Daten
- Transfer einer EP aus verschiedenen Repositorien in verschiedene dLZA
- Einige nestor-Aspekte wurden gerne übernommen
- In Niederlanden werden die Komponenten der EPs nach Zuständigkeit getrennt archiviert:
  - Ingest wird eher zum Registrieren von Identifiern der Komponenten
  - Schwierigkeit: dauerhaftes Synchronisieren der dLZA-Policys

# Geplante Weiterentwicklung

- Veröffentlichung englischer Übersetzung Juni
- Weitere Anwendungen
- Weiterentwicklung im Rahmen des DIN NABD 15
- Bereiche mit m.E. Bearbeitungsbedarf:
  - Komplex „Preservation Level“: wesentliche Eigenschaften, designated Community, Authentizität, Recht und Vertrag
  - „Teilbereiche des Managements“ angleichen
  - Übernahmevereinbarungen ausdifferenzieren

# Mögliche Weiterentwicklungen?

- Wären Community spezifische Versionen sinnvoll?
  - Für Bib., Archive, Museen, Industrie, Wissenschaft, ...
  - mit spezifischen Anforderungen, Begründungen, Beispielen?
- Case-Studies durchführen
- Idee aus DRIVER: Vorlagen für Übernahmevereinbarungen entwerfen?

**Vielen Dank für Ihre  
Aufmerksamkeit!**

**Fragen?**

- Beagrie, Neil; Chruszcz, Julia; Lavoie, Brian:  
Keeping Research Data Safe – A Cost Model and  
Guidance for UK Universities. JISC 2008.  
<http://www.jisc.ac.uk/media/documents/publications/keepingresearchdatasafe0408.pdf>
- Consultative Committee for Space Data Systems:  
Reference Model for an Open Archival Information  
System (OAIS). CCSDS 2002.  
<http://public.ccsds.org/publications/archive/650x0b1.pdf>

- nestor AG Standards: Wege ins Archiv. Ein Leitfaden für die Informationsübernahme in das digitale Langzeitarchiv. nestor 2008. URN: urn:nbn:de:0008-2008103009
- Wilson, Andrew: Significant Properties Report. 2007. [http://www.significantproperties.org.uk/documents/wp2\\_2\\_significant\\_properties.pdf](http://www.significantproperties.org.uk/documents/wp2_2_significant_properties.pdf)