

## **Anforderungen an die Ingest-Routine von Digitization Lifecycle (MPI für Rechtsgeschichte, 28.4.2011)**

### **Vorbemerkung:**

Das Verfahren entspricht im Großen und Ganzen dem, das in unseren bisherigen Projekten zum Einsatz kam und das sich wegen seiner Unkompliziertheit und Schnelligkeit bewährt hat.

### **Ziel:**

Die Ingestprozedur wird von Anfang an so gebaut, dass

- ein Ingest sowohl von Einzelbänden als auch ein Ingest von „Lieferungen“ (d.h. mehrere/viele Bücher gleichzeitig) möglich ist;
- die Zuordnung verschiedener Datentypen (bibliographische Metadaten, Strukturdaten, Volltexte, Images, ...) in einem automatisierten Prozess erfolgt;
- die Umwandlung der Rohdaten in Lifecycle-Datenformate in einem automatisierten Prozess erfolgt;
- gleichzeitig mit dem Ingest neuer Daten eine Aktualisierung von Metadaten stattfinden kann;
- nötige Prüfroutinen, Protokollierungen und Exporte stattfinden.

### **Ausgangspunkt:**

Der Ingest erwartet mindestens Images und Bibliographische Metadaten. Es ist möglich, NUR Images oder bibliographische Metadaten „auf Vorrat“ an die definierten Stellen zu kopieren, findet das System aber nicht beide vor, wird kein Ingest durchgeführt.

### **Schritte:**

#### **1. Auswahl der Digitalen Sammlung**

#### **2. Kopieren von Bildern in ein definiertes Verzeichnis**

- 1. Lieferung: x Verzeichnisse mit Bildern zu x Büchern
- 2. Lieferung: x+n Verzeichnisse mit Bildern zu x+n Büchern
- ...

Variante 1 (Standard – bei guter Vorbereitung und Prüfung der Daten):

Es kommen mit einer neuen Lieferung jeweils nur neue Bilder hinzu, an den alten Bildlieferungen/-verzeichnissen ändert sich nichts.

Variante 2 (möglichst zu vermeiden, aber nicht auszuschließen):

Es kommen mit einer neuen Lieferung jeweils neue Bilder hinzu UND in den alten Bildlieferungen/-verzeichnissen werden einzelne Bilder ausgetauscht/überschrieben (neuer PID?, Versionierung?).

Variante 3 (möglichst zu vermeiden, aber nicht auszuschließen):

Es kommen mit einer neuen Lieferung jeweils neue Bilder hinzu UND Verzeichnisse aus älteren Bildlieferungen werden komplett ausgetauscht (neuer PID?, Versionierung?).

### **3. Kopieren von bibliographischen Metadaten in ein definiertes Verzeichnis**

- 1. Lieferung: EINE Datei mit den Bibliographischen Daten von Lieferung 1. Die Datei enthält alle bibliographischen Daten zu allen Bänden inkl. die Datensätze von Überordnungen (Gesamtaufnahme mehrbändiges Werk, Serie, ...) und von Unterordnungen (beigefügte/enthaltene Werke, ...)

- 2. Lieferung: EINE Datei mit den Bibliographischen Daten von Lieferung 1 und 2

...

➔ mit jeder Lieferung werden ALLE jeweils aktuellen bibliographischen Daten dieser Sammlung eingespielt und damit Aktualisierungen/Korrekturen vorgenommen.

#### Besonderheit fortlaufende Sammelwerke / Zeitschriften:

Hier gibt es nicht für jeden Band einen eigenen Datensatz, deshalb werden die bibliographischen Angaben mit Angaben zu den Bänden ergänzt:

- Bandangabe/Erscheinungsdaten des Einzelbandes in zusammenfassender Form;  
Bsp.: "Jg. 1846, Bd. 3 (1846)"
- Sortierfeld 1 (5stellig, mit führenden Nullen)
- Sortierfeld 2 (3stellig, mit führenden Nullen)
- Fußnote
- Besitzer der Digitalisierungsvorlage
- [zu ergänzen]

### **4a. Kopieren von offline erzeugten Strukturdaten in ein definiertes Verzeichnis**

- 1. Lieferung: x MODS-Dateien zu x Büchern

- 2. Lieferung: x+n MODS-Dateien zu x+n Büchern

...

➔ mit jeder Lieferung werden ALLE jeweils aktuellen Strukturdaten dieser Sammlung eingespielt und damit Aktualisierungen/Korrekturen vorgenommen.

Auch beim Vorhandensein eines online Strukturdateneditors relevant (von dem wir ausgehen), um Aktualisierungen / Ergänzungen ggf. offline durchführen zu können. Strukturdaten online erfasst ➔ Export ➔ Änderung ➔ Ingest (Überschreiben)

#### **Und/oder**

### **4b. Kopieren von offline erzeugten TEI-Volltexte in ein definiertes Verzeichnis**

- 1. Lieferung: x TEI-Dateien x Büchern

- 2. Lieferung: x+n TEI-Dateien zu x+n Büchern

...

➔ mit jeder Lieferung werden ALLE jeweils aktuellen TEI-Volltexte dieser Sammlung eingespielt und damit Aktualisierungen/Korrekturen vorgenommen.

[4c. zu ergänzen: weitere, z.B. kodikologische, Metadaten]

### **5. Anstoß des Ingests, bewirkt:**

**5a. Automatisches Errechnen von Image-Derivaten (jpeg, PDF), wenn technisch erforderlich: Imageimport in Digilib, eSciDoc**

**Danach: Automatisches Verschieben der Image-Rohdaten in ein „Zwischenarchiv“** (also aus dem Ingest-Verzeichnis raus)

### 5b. Automatische Zuordnung der verschiedenen Datentypen anhand von Identifiern, Verzeichnis- und Dateinamen

Verschiedene denkbare Varianten (können sich von Sammlung zu Sammlung unterscheiden, innerhalb einer Sammlung aber nicht gemischt werden):

Bibliographische Daten	Images / Verzeichnisname	MODS / Dateiname	TEI / Dateiname
<b>Monographie, Band e. Mehrbd. Werks:</b>			
Katalognr. (lokal)	Katalognr.	Katalognr.	Katalognr.
Signatur (lokal) MAB 544	Signatur	Signatur	Signatur
Verbund-ID (z.B. PPN)	Verbund-ID	Verbund-ID	Verbund-ID
Standardnr. (z.B. VD16) MAB 580_	Standardnr.	Standardnr.	Standardnr.
<b>Band einer Zeitschrift:</b>			
ZDB-ID MAB 025z	ZDB-ID als Präfix_Band...	ZDB-ID als Präfix_Band...	ZDB-ID als Präfix_Band...

### 5c. Automatisches Erstellung von METS-xml

### 5d. Automatisches Erstellen von TEI

### 5e. Verknüpfung mit Normdaten

### 5d. Automatische Vergabe von Persistent Identifiern (PID)

Handle-PID für das gesamte Buch (Digitale Faksimiles) (verschiedene PIDs für verschiedene Auflösungen?)

Handle-PID für jede einzelne Seite (Digitale Faksimiles) (verschiedene PIDs für verschiedene Auflösungen?)

Handle-PID für Struktur- und Volltextdaten?

[zu ergänzen]

### 6. Prüfroutinen

- Existieren mindestens Images und bibliographische Metadaten, die einander zugeordnet werden können?

- Sind die gelieferten Strukturdaten valide?

- Sind die TEI-Daten valide?

[- zu ergänzen]

## **7. Protokollierung**

- Anzahl der enthaltenen Bände (gesamt / neu)
  - Anzahl der enthaltenen Images (gesamt / neu)
- [- zu ergänzen]

## **8. Export**

- Export von URL/PID zum Einspielen in den lokalen Bibliotheks-OPAC
- [- zu ergänzen]
- 

## **Ingest-Manager**

Mittels eines Ingest-tools ließe sich der Ingest weniger statisch gestalten. Es ist zu überlegen, ob so etwas Sinn macht.

Funktionen

- die Definition von Identifiern, Verzeichnis und Dateinamen, die zur Verknüpfung der Daten herangezogen werden (s. 5b);
- die Angabe von lokalen- oder Webverzeichnissen, auf denen Daten zum Ingest bereitliegen;
- die Auswahl, welche Art von Daten ingestet werden sollen.
- Definition: was soll mit den Daten gemacht werden?
  - o Ersetzen von Bibliographische Metadaten
  - o Hinzufügen/ersetzen von Imageverzeichnissen
  - o Hinzufügen/ersetzen von Strukturdaten
  - o Hinzufügen/ersetzen von Volltexten
  - o [zu ergänzen]