



Mögliche Funktionalitäten einer DLC Webumgebung & Technische Machbarkeit

Andrea Kulas

Digitization Lifecycle Meeting 06./07.06.2011



Themen

- Ingest – Was lade ich wie hoch?
- Editing – Was kann ich bearbeiten?
- Datenformate für Strukturdaten
- View & Search - Was sehe ich und nach was kann ich suchen?
 - Szenarien: Basisfunktionalitäten -> erweiterte Funktionalitäten +?
 - Was ist neu? Alleinstellungsmerkmale?
 - Offene Fragen für die MPDL
 - Klärung von Begriffen

1. Ingest - WAS lade ich WIE hoch?

- Szenario 1:
 - Institute laden Scans (auch gleichzeitig) + MAB file (bibl. Daten) zu einem Buch hoch.
(-> Buch wird von der MPDL im Container Format METS abgelegt)
TEI kann gleichzeitig oder später folgen.
 - Upload von kodikologischen Metadaten (Welches Format?)
 - Zusatzfeature: Maske zum manuellen Eingeben der bibliographischen Daten
- Szenario 2: aufwendiger
 - Austausch von Scans, Zufügen von Scans zu einem Buch
(-> GUI Anpassung)

1. Ingest - Fortsetzung

- Szenario 3: Hochladen von mehreren Büchern gleichzeitig
 - Automatische Zuordnung von Scans zu Buch (durch Dateinamen etc.)
-> ?
 - Hochladen von mehreren Büchern gleichzeitig -?
Eventuelle Alternativen:
 - Mehrere Bücher in einen ZIP gepackt (Scans, MAB, TEI; Festlegung auf bestimmtes Schema als Voraussetzung)
 - Upload Manager/ Ingest Tool, der offline funktioniert (Realisierung nach Projektlaufzeit?)
- Was ist neu?
 - Nutzer laden selbst ganze Bücher (Scans, MAB, TEI) in eine Webanwendung + machen sie nutzbar
 - Abbildung des Workflows für Digitalisierung im Web
 - Nutzer lädt Bücher hoch -> Anzeige der Bücher (automatische Anlage der Struktur, Transformationen etc.) -> Bearbeitungs-möglichkeiten, Suchmöglichkeiten -> Virtuelle Forschungsumgebung

2. Editieren - Was können wir bearbeiten?

1. Szenario

- Verschiedene Möglichkeiten des Editierens für selbst erstellte Daten (Struktur, Bibliographische Daten, Annotation)

- TEI Dokumente (offline erstellt) werden offline bearbeitet

▪ 2. Szenario: „Problem der Masse“

- Editieren von kodigologischen Metadaten (strukturiert) im Web

▪ 3. Szenario: ???

- Begrenzte Auswahl an Elementen (Strukturdaten?) im TEI /Volltext bearbeiten

- TEI dynamisch -> Wo steht was? Kein Verlass wo etwas steht.

- Es gibt keinen XML-Editor (für TEI) im Web

- Desktop vs. Webanwendung

▪ Was ist neu?

- Struktureditor (falls gewünscht)

3. Datenformate für Strukturdaten

- Datenformate - Varianten: METS und TEI
 - Wo legen wir die Struktur ab? In TEI oder in METS?
 - Präferenz für TEI; noch zu klärende Frage: Können beliebige Metadaten im TEI integriert werden?
 - Lösung Redundanz: TEI kommt -> alte METS oder TEI Struktur wird gelöscht/überschrieben
- Frage: Was für Strukturtypen und Strukturmetadaten wollen die Institute haben (-> siehe aktuelle Version in VIRR)?
 - <http://test-virr.mpdl.mpg.de:8080/virr/>
(<http://virr.mpdl.mpg.de/virr/>)
 - Dieser Punkt ist nur relevant für einen Struktureditor.
-> brauchen wir diesen?

Was lesen wir aus dem TEI heraus?

- TEI enthält: Strukturdaten, Metadaten, Volltext
 - Struktur herauslesen (Elemente div, pb, head, body ausreichend)
 - Anzeige der Volltexte für die einzelnen Seiten -> Voraussetzung: Anzahl der Seitenumbrüche <pb> im TEI entspricht Anzahl der Scans (= eine Seite)
 - Anzeige (+ Bearbeitung) der Metadaten
 - Für die Suche (im Volltext)
- Was ist neu?
 - Volltexte in Bezug zu Scans gesetzt

4. View & Search – Was sehen wir und nach was suchen wir?

- Search
 - Szenario 1: simple search
 - Szenario 2: advanced search (Menge der Felder bestimmt Zeitaufwand)
 - Szenario 3: Faceted Search
- View:
 - Szenario 1: Standard (siehe VIRR):
 - Werk durchblättern
 - Strukturbaum anzeigen
 - Kapitel aufrufen
 - Verschiedene Ansichten (einseitig, zweiseitig, Thumbnail-Übersicht, Struktur-Gliederung, Volltext)
 - Hervorheben/Markieren in HTML



4. View & Search - Fortsetzung

- Szenario 2: komplex
 - Desktop-Anwendungen im Web zu schaffen
 - Große Scans (e.g. Karten) anschauen (Digilib-Funktionen)
 - Interaktive Funktionen: Lupenfunktion, zoomen, vergrößern
 - Hervorheben/Markieren von Stellen im Bild
 - Bildmanipulation (Edit von Bildern): Drehen
- Was ist neu?
 - Markieren + Referenz erstellen (-> neue Forschungsaspekte im Rahmen einer virtuellen Forschungsumgebung schaffen)
 - Annotation, Referenz auf Teilbereiche (Pixel) eines Bildes

Was schaffen wir in der Projektlaufzeit?

Basis -> Ausbau

Schritt 1: Basisfunktionalitäten für eine !Web-Anwendung! in den Kategorien:

- Ingest
- View
- Search
- Edit
- Export
- Wissenschaftliche Nutzung

Was schaffen wir in der Projektlaufzeit?

**Schritt
2:
Was
ist das
Neue,
das
wir in
der
Zeit
schaffen
können?**

- Ingest eines Buches (Scan, MAB, TEI incl. Metadaten) durch Nutzer
- Webeditor für Strukturdaten
- Volltexte + Scans
- Markieren, Referenzen (bis hin zu Bildausschnitten, Pixeln?)
-

■ Voraussetzung für Entwicklungen in diesen Bereichen: Entwickler haben eine Vorstellung von der möglichen Realisierung



Fragen, Anregungen, Meinungen???