

Struktur DLC-Leitfaden (LP)

1. Einleitung

1.1 Projektbeschreibung DLC

Hervorheben des Alleinstellungsmerkmals von DLC (Was macht DLC zu etwas Besonderem)

Tool zur Erstellung digitaler Bibliotheken/Sammlungen

Mit DLC ist ein Open Source tool verfügbar, welches Max-Planck-Institute und anderem wissenschaftlichen Einrichtungen und Organisationen die Möglichkeit bietet, digitale Sammlungen zu erstellen, und diese zu verwalten. Bearbeitende können jederzeit Daten ändern, z.B. Images löschen, ersetzen oder neu einstellen, bibliographische, administrative oder deskriptive Metadaten verändern, löschen und neu einstellen.

Integration von TEI-XML-Volltexten

Optional bietet DLC die Funktion Volltexte gemäß TEI-P5 aufzunehmen, so dass z.B. alte Drucke durch Zugabe der Volltexte durchsuchbar gemacht werden können.

Autonomie von einem zentralen Infrastrukturbetreiber

Akteure, die DLC nutzen, können autonom sämtlichen Daten online stellen, die zur Erzeugung digitaler Sammlungen erforderlich sind. Images, Volltexte oder bibliographische Metadaten müssen nicht bei einer zentralen Stelle gemeldet werden, damit sie in das System kommen. Dadurch werden sie unabhängig von der Freigabe und/oder Administration zusätzlicher Akteure. Durch manuellen upload oder durch die Angabe von urls für einen ftp Transfer können Sammlungen je nach Bedarf angelegt werden.

Erstellung des Inhaltsverzeichnisses mittels online-editor

Eine weitere Besonderheit von DLC ist der Online-Editor, der zur Erstellung einer navigierbaren Struktur (z.B. eines Inhaltsverzeichnisses) dient. Der Online-Editor stellt eine Vielzahl von Instrumenten für die Paginierung von Seiten, für die rechts/links Anzeige oder für die Typisierung von Überschriften zur Verfügung (Inhaltsverzeichnis, Titelblatt, Abschnitt etc.).

Flexibilität in der Form der bereitzustellenden Daten

Images können als png, jpg oder tiff nach DLC transferiert werden. Bibliographische Metadaten können als MAB2xml transferiert oder manuell eingegeben werden. TEI-Volltexte können transferiert oder in begrenztem Umfang im Online-Editor erstellt werden.

Mehr Sichtbarkeit der eigenen Digitalisate durch die Funktion einer sammlungsübergreifenden Suche

Neben den genannten Besonderheiten, die sich vor allem durch die autonome Bearbeitung digitaler Sammlungen im Browser auszeichnen, besteht ein weiterer Vorteil von DLC in der Genierung einer wachsenden Sammlung heterogener Bestände. Institute und Partner, die ihre DLC-Anwendung unter der eSciDoc Plattform der MPDL integrieren stellen ihre Sammlungen dadurch zugleich in einen sammlungsübergreifenden Raum, d.h. eine übergreifende Suche in allen Sammlungen wird dadurch generiert.

- Schreibende: Ingo Caesar, Andrea Kulas

1.2 Anliegen des Leitfadens

- Zielgruppe
 - Infospezialisten/Bibliothekare der MPG u. darüber hinaus UND
 - WissenschaftlerInnen
- Zweck: Wegweiser
- Schreibende: Leitende

1.3 Digitalisierungsprojekt: Voraussetzungen

- Vorüberlegungen zum eigenen durchzuführenden Digitalisierungsprojekt (i. S. einer Selbstklärung für den Leser) --> Hier sollen die W-Fragen gestellt werden.
- zu jedem Punkt folgt ein Verweis auf das zu behandelnde Kapitel im Leitfaden oder auf Referenz aus anderen Leitfäden/Handbücher, die diese Punkte bereits ausführlich behandeln
- Schreibende: Anna Klug (Erstellung von Vorlagen, weiteres Vorgehen: kollaborativ)

W-Fragen:

Was?

Was soll digitalisiert werden?

Was sind die Gründe für das anstehende Digitalisierungsprojekt?

Was für Arbeitsschritte / Projektphasen resultieren daraus?

1. Bestandssicherung

= erste Phase: Erstellung von Scans und Gewährleistung ihrer Langzeitarchivierung; Erstellung einer technisch kompatiblen Ordnerstruktur und Dateibenennung

(Verweise auf Kap. Standardformatfragen, bezügl. Bildformate; Links zum Thema Langzeitarchivierung)

2. Ortsungebundene Bereitstellung des Materials übers Internet

= zweite Phase: Aufbau einer Digitalen Bibliothek und Gewährleistung der Auffindbarkeit der Digitalisate und der Benutzerfreundlichkeit (Anreicherung mit bibliographischen Metadaten und Strukturmetadaten)

3. Detaillierte Suche: Volltextsuche

= als Erweiterung der Digitalen Bibliothek (Zweite Phase): Maßgebliche Erleichterung des wiss. Arbeitens durch die Digitale Bibliothek

4. Neue Methoden wiss. Arbeitens

= dritte Phase: Erweiterung zur Virtuellen Forschungsumgebung: Annotationstools sollen auch das collaborative Arbeiten gewährleisten.

Implementation und wechselseitige Verlinkung von Datenbanken und anderen Internetressourcen

Was soll digitalisiert werden? Welche besonderen Hindernisse können auftreten?

(z. B. Notwendigkeit von Leihgaben, z. B. Notwendigkeit der Klärung der Rechte)

Wer?

Welche Kompetenzen sind nötig?

Wer leistet was im Projekt?

[Hier können Anforderungen an die unterschiedlichen Personenprofile gesammelt werden.]

Entscheidungsträger (= bei uns: Der "Lenkungskreis"):

Experten:

Projektkoordination:

Entwickler:

Wissenschaftliche Mitarbeiter:

Bibliothekare:

Wie?

Generelles:

[bitte sammeln!]

Zur Frage, wie das Projekt "Digitalisierung" anzupacken ist, wird im Wesentlichen auf die kommenden Kapitel verwiesen:

- Thema: Verständigung (wiss. & technische Basis) -> Verweis aufs Glossar
- Klärung der Formatfragen (Klärungszeitpunkt, Lösungsansätze)-> Verweise auf folgende Kapitel: Standardformatfragen; Volltext; TOC-Auszeichnung
- Zeitplan und Workflow -> Verweis aufs Kapitel Workflow; Verweis aufs Kapitel Virtuelle Forschungsumgebung (mit Hinweis auf die notwendige Einteilung in Projektphasen)
- Möglichkeit der Nachnutzung des DLC-OpenSource-Pakets > Verweis auf den praktischen Teil "Beschreibung der DLC-Anwendung"

Relevante Literatur:

[Bitte beim Schreiben der einzelnen Kapitel darauf achten, welche Literatur aus welchen Gründen verwendet wurde und im Anschluss hier sammeln!]

Fragenkatalog Digitalisierung (kew):

- Materialgattung: Welche Materialien sollen digitalisiert werden? Kriterien koennen etwa sein: Erhaltungszustand, Nutzungshaeufigkeit, Nachfrage, Alter, Seltenheit. Dabei ist abzuwaegen zwischen den Belangen der Bibliothek (Konservierung), den Wuenschen der Nutzer (Zugang) und dem Aufwand einer Digitalisierung.
- Grad der Digitalisierung: Wie intensiv soll die Digitalisierung sein? Reicht ein Scan der Seiten aus, soll ein simples TOC vorhanden sein, soll ein Volltext angeboten werden zur Suche im Text oder sogar ein Volltext mit semantischer Auszeichnung zur Benutzung in wissenschaftlichen Umgebungen?
- Auszeichnungsgrade: welche Funktion und welchen relevanten Nutzen haben diverse Auszeichnungsgrade? Wie kann, fuer ein spezifisches Projekt, die Intensitaet der Auszeichnung und der dafuer erforderliche Zeit- und Personalaufwand abgeschaezt werden, abgesehen vom erhoekten finanziellen Aufwand?
- Ausfuehrung: soll das Projekt so angelegt sein, dass es in-house mit eigenen Kraeften betrieben werden kann? Ist von vornherein ein Auslagern auf externe Dienstleister vorgesehen? Ist eine Zusammenarbeit mit Google Books in Erwaegung gezogen worden?
- Publikation: sollen die Ergebnisse nur innerhalb der eigenen Arbeitsgruppe zur Verfuegung stehen? Ist eine Publikation auf eigener Website vorgesehen? Oder auf externen Portalen, die zumeist bestimmte Formate vorschreiben?
- Formate: Welche Bildformate kommen in Betracht, und was sind ihre grundsaeztlichen Vor- und Nachteile (TIFF? JPEG? Raw? DNG?). Bei einer Volltexterstellung kommt noch die Frage nach den verwendeten Textformaten hinzu (AsciiDoc? DocBook? DITA? TEI?). Wo kann hier abgewogen werden zwischen Aufwand in der Erstellung, Nutzen fuer das Digitalisierungsprojekt und eventueller Nach-Nutzung in anderen Projekten?

- Applikationen: Hierbei liegt das Augenmerk auf der wichtigen Unterscheidung zwischen dem Produkt (Digitalisat in einem bestimmten Text/Bildformat, entweder mit einfachem Zugang ueber Web oder Einbindung in Forschungsumfeld) und der Produktion selbst. Beides muss unabhaengig voneinander definiert werden: das Produktionsumfeld wird sich - und muss sich, um aktuell zu bleiben - alle 2-3 Jahre aendern, waehrend die definierten Bild- und Textformate langfristig brauchbar bleiben muessen.

1.4 Glossar

- Wichtig, dass dieses am Anfang steht, damit dem Leser u. auch dem Schreibenden klar ist, welche Begriffe Verwendung finden u. wie diese im DLC-Kontext zu verstehen sind.
- Schreibende: Ingo Caesar (Koordinator)

2. Rahmenvoraussetzungen/Hintergrund-informationen (theoretischer Teil)

- Probleme, was muss berücksichtigt werden

2.1 Standardformatfragen

(Scans/Bildformate, Datenformate, Textauszeichnung)

- Hier soll es um die gängigen Standards für Digitalisierung und Textauszeichnung gehen.
- Zum Punkt Textauszeichnung: Hier geht es um die Behandlung von TEI und xml allgemein. Genaueres zu Volltextfassung (derzeit bildet dies noch die Ausnahme bei Digitalisierungsprojekten) und Mindestauszeichnung (TOC plus ggf. weitere Auszeichnung) in den 2 folgenden Kapiteln
- Schreibende: BHertz, MPIeR (auf Titanpad)
- Scans/Bildformate (kew: tiff, jpeg, png, RAW; aufloesung, farbprofil, 12bit/14bit/16bit, compressed-uncompressed-lossy)
- Datenformate (METS/MODS, MAB, MARC (+xml) Formate)
- Textauszeichnung (kew: TEI vs. Others ALTO XML, ASCII Text; tools, - vorhandene: online [ROMA, DLC] und local [Oxygen])
- Standards folgen u./o. eigene Entwicklung verfolgen (kew: excursus public/private TEI)
- Verknüpfung mit Standardnormdaten (Getty-Normdaten, GND etc., linked open data-wikipedia dbpedia, VIAF)
- Verlinkung zu Normdaten anderer Fachrichtungen sind von Interesse (georef - getty places)

2.2 Volltext

(OCR, double keying, Vergleich OCR/double keying, Tagging-Anleitung)

- Review/Stand der Dinge: Kapitel ist eine Mischung aus Verweisen u. einer ausführlichen Darlegung zum Thema Volltextfassung/Umgang mit Volltexten (Erfassung, Verarbeitung, Recherche)
- Erfahrungen aus dem eigenen Projekt einfließen lassen (i. S. einer lesson learned)
- Schreibende: MPIB (auf Titanpad)

2.3 TOC-Auszeichnung

- Schreibende: Klaus Werner, Anna Klug, Ingo Caesar (auf Titanpad)

2.4 Workflow

- DLC soll Spektrum aufzeigen, was möglich ist
- Workflow ist abhängig von den Anforderungen ans Endprodukt
- Was will ich mit den Digitalisaten machen?

- Schreibende: Lisa Pegelow, Klaus Werner (auf Titanpad)
- Workflows siehe <https://docs.google.com/#home> (Anmeldung erforderlich, Freigabe erforderlich)

2.5 Erfordernisse an eine virtuelle Forschungsumgebung

- Navigation / Präsentation der Ergebnisse / Viewing (ggf. Einbinden der Dokumentation der AG Viewing)
- Einbinden u. a. von Landkarten, Lexika, Wörterbüchern
- Verknüpfung mit/Verweis auf: relevante/n Schnittstellen, Portale/n zum Posten, andere/n Datenbanken
- Prozedere zum Bekanntmachen der neuen virtuelle Forschungsumgebung
- Vision(en) aus DLC heraus entwickeln
- Schreibende: noch zu gründende AG virtuelle Forschungsumgebung/wiss. Arbeiten

3. Beschreibung der DLC-Anwendung (praktischer Teil)

- Mitmach-Teil: Anreize der Nachnutzung schaffen
- (Nach-)Nutzung durch
 - Bibliothekare: Mitmachanreize schaffen UND
 - WissenschaftlerInnen: direkte Ansprache an WissenschaftlerInnen --> Vermarktung von DLC unter WissenschaftlerInnen
- Anleitung zur Benutzung der Tools
- Schreibende: MPDL (1. Version), AG Digitalisierungsleitfaden (2. Version)

Gliederung für die Beschreibung von DLC

0. Einleitung

1. Datenformate & Architektur

a. Importformate (für Monographien, Mehrbandwerke, Bände)

- Scans
- Bibliographische Metadatenformate
- Volltext

b. Exportformate (siehe auch Kapitel 7)

- Zitierstile

c. Speicherformate und Mappings

d. DLC Architektur

2. Administration und Rechteverwaltung (Nutzer & Rechte, Adminbereich)

3. Vom Hochladen zum Publizieren

a. Manuelles Hochladen vom Desktop aus (Ingest-Seite, Reordering, Validierung)

- Manuelle Input-Maske für Bibl. Daten

b. Arbeitsbereich „Meine Datensätze/My Items“

- Release-Funktion
- Reingest/Edit (Update-Funktion)

c. Massenimport

- Arbeitsbereich für Massenimport

4. Browsing

a. Ansicht Einzelwerk

- Bildansicht
 - Bildbetrachtung mit Digilib
 - Paginator
- Volltextansicht
- TOC (Highlighting)
- Ansicht Thumbnails

d. Ansicht Institutionen / Kollektionen /Suchergebnisse

5. Suche

- a. Einfache Suche
- b. Erweiterte Suche (Highlighting im Volltext)
6. Anreichern der Digitalen Objekte/Editieren
 - a. Erstellung Basisdokument Volltext (TEI SD)
 - i. Inhaltsverzeichnis /TOC (Strukturelemente, Seiteneinstellungen wie Anzeigetyp und Seitentyp, Validierungsfehler, Veröffentlichen)
 1. Versionierung
 - ii. Paginieren
 1. Physische und logische Seitennummer
 2. Paginierungsarten- und muster
 - b. Bibliographische Metadaten
 - c. Update
 7. Export
 - a. Von einzelnen Daten
 - i. DLC Objekte allgemein (z.B. als METS XML)
 1. DFG-Viewer
 - ii. Volltext (-> HTML/PDF/XML)
 1. Extern erstellter Volltext
 2. Online erstellte Strukturdatei (TOC)
 - iii. Bilder (Scans) (-> PDF)
 - iv. Bibliographische Angaben (-> Cover page in PDF)
 - b. Schnittstellen
 - i. OAI-PMH Interface (->ZVDD METS + DC-Simple)
 8. Wissenschaftliche Basisfunktionen (Annotationen)
 9. DLC als Open-Source-Projekt (Lizenz etc.)
 10. Weitere Informationsquellen / Unterstützung / Hilfe für die Benutzung von DLC (siehe auch Kapitel 4.1)

4. Schluss

4.1 Ausblick/Nachhaltigkeit des DLC-Projektes u. anderer Digitalisierungsprojekte

- Dieses Kapitel soll den Versuch darstellen, DLC u. die Arbeit des Projektes in eine längerfristige/nachhaltige Perspektive zu setzen (i. S. einer Einordnung vor einen längeren Horizont)
- Hier könnte auch über andere Projekte (u. U. auch nicht namentlich genannt) kürzest referiert werden, wie dort der Stand der Dinge ist, wie sich deren Erfolg ausmacht, v. a. hinsichtlich Effizienz, Nachhaltigkeit u. Nutzbarkeit (externe Tools kreieren bzw. externe Tools einbindbar in andere Anwendungen --> Frage der Kompatibilität)
- Schreibende: Leitende (inkl. Input von Seiten der MPDL zu DLC-Tools/Nachhaltigkeit von Softwareentwicklung)

5. Anhang

(Vorlage Tagging-Anleitung, (kommentierte) Liste anderer Leitfäden (auch als Zusammenfassung), weiterführende Literatur, (kommentierte) Liste anderer virtuellen Forschungsumgebungen; Portale/Schnittstellen, Auflistung der beteiligten MPIs, MPG, affiliated partners)

- Tagging-Anleitung soll als Bsp. dienen, wie Auszeichnung geschehen kann u. welche Fragen gestellt werden müssen
- Auflistung der anderen Leitfäden stellt u. a. eine Zusammenfassung der im Text genannten Leitfäden dar, ggf. plus weitere
- Liste der Forschungsumgebungen nach Fachrichtungen sortiert
- Entwurf eines Formulars für WissenschaftlerInnen: Digitalisierungsantrag an die Bibliothek
- Schreibende: AG Digitalisierungsleitfaden

Themen rund um die Digitalisierung, auf die es zu verweisen gilt (d. h.: keine eigene Ausarbeitung durch das DLC-Handbuch) (LP)

- Rechtsfragen (Urheberrecht bei gemeinfreien Werken): s. u. a. in: Kreuzer, Till (2011): Digitalisierung von gemeinfreien Werken durch Bibliotheken. Ein Leitfaden, hg. v. Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen
- Materialbeschaffung der Digitalisierungsvorlagen aus
 - eigenem Bestand oder via
 - Leihe
- Langzeitarchivierung
- Varianten für die Auszeichnung und Abbildung von Sonderzeichen + Probleme --> Unicode: z. B.: textGrid-Tool Unicodetabelle
- Auszeichnung Strukturdaten: s. dazu DFG-Strukturdatenset, zvdd-Strukturdaten
- Einsatz von Workflow-Software: z. B.: Agora <http://www.agora.de/>
- IT-Themen
 - Konzept zur Ordnerstruktur, Dateibenennung, Ablage der Daten (Bilder, Textdateien)
- Checkliste u. Good Practice Beispiele: s. u. a. in: Bestandsaufnahme zur Digitalisierung von Kulturgut und Handlungsfelder. Erstellt vom Fraunhofer Institut für Intelligente Analyse und Informationssysteme IAIS im Auftrag des BKM, 2007
- Metadatenvergabe
- Umgang mit Dienstleistern: s. DFG-Richtlinien
 - Musterverträge

PROBLEM: die Fußnoten werden nicht aus Word übernommen - Gibt es da eine praktikable Lösung?