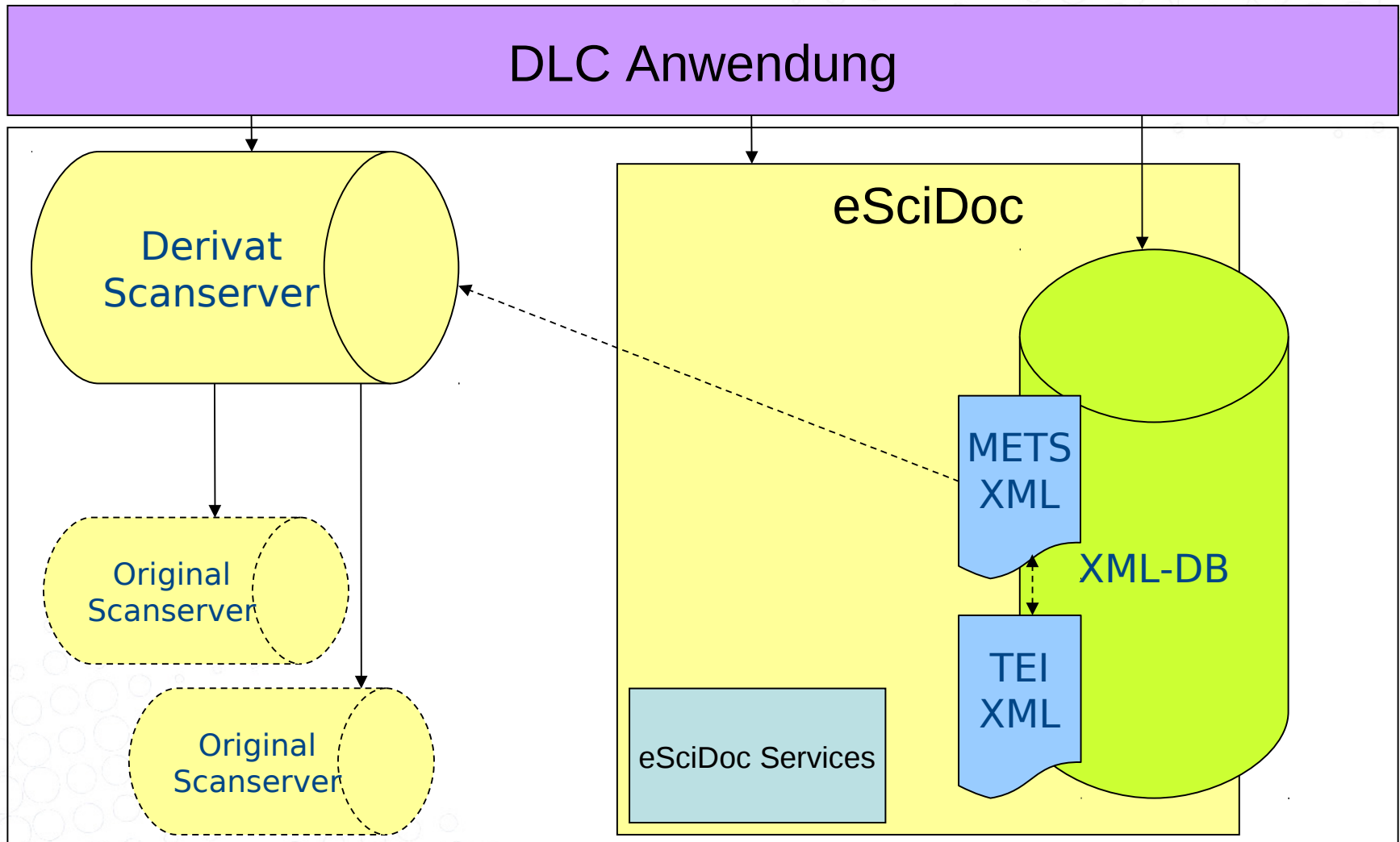




Übersicht über technische Architektur & interne Datenformate

Digitization Lifecycle Project

Einfache Sicht auf die Systemarchitektur



Scanserver

- Derivat Scanserver
 - Enthält Bildderivate (JPEGs) in verschiedenen Auflösungen
 - Keine Original-TIFFs
 - Evtl. Einbindung eines Imaging Server Systems (z.B. Digilib, IIPImage) - Weitere Evaluation nötig
- Original Scanserver
 - Enthalten Original Tiffs, kein Zugriff von DLC
 - Idee: Zentraler Bildserver darf auf Dateisystem zugreifen – dann könnten Original Scans über Imaging Server System on demand zur Verfügung gestellt werden
 - Systemadministrator muss dies einrichten
- Zwei Scan-Ingestszenarien:
 - Bilder werden vom lokalen PC hochgeladen (Größenbeschränkung), Derivate werden automatisch erstellt.
 - Bilder werden über einen Dateipfad auf Originalserver zur Verfügung gestellt
 - Wie funktioniert der Ingest?

METS – Von losen Scans zu einem Buch

- Containerformat
- Automatische Erstellung beim Ingest
- Enthält:
 - Referenzen zu den Bilddateien und deren Auflösungen
 - Physische Struktur des Buches
 - Bibliographische Metadaten (MODS)
 - Referenz zu kodikologischen Metadaten
 - Referenz zum TEI
 - Später nach Erstellung: Paginierung

METS: Referenzen zu Bilddateien

- Abgelegt in File-Section von METS
- Verschiedene File-Groups für verschiedene Auflösungen

```
<mets:fileSec>
  <mets:fileGrp USE="DEFAULT">
    <mets:file MIMETYPE="image/jpeg" ID="ptr_03">
      <mets:FLocat LOCTYPE="URL" xlink:href=
        "http://r-coreservice.mpd.l.mpg.de:80/ir/item/escidoc:326843/components/component/escidoc:326841/content"
        xmlns:xlink="http://www.w3.org/1999/xlink"/>
    </mets:file>
    <mets:file MIMETYPE="image/jpeg" ID="ptr_13">
      <mets:FLocat LOCTYPE="URL" xlink:href=
        "http://r-coreservice.mpd.l.mpg.de:80/ir/item/escidoc:326847/components/component/escidoc:326845/content"
        xmlns:xlink="http://www.w3.org/1999/xlink"/>
    </mets:file>
  </mets:fileGrp>
</mets:fileSec>
```

METS: Physische Struktur

- Abgelegt in structMap von METS
- Enthält Referenzen zu verschiedenen Auflösungen
- Enthält Paginierung

```
<mets:structMap TYPE="PHYSICAL">
  <mets:div ID="physstruct0" ADMID="amd1" DMDID="dmd0" TYPE="physSequence">
    <mets:div ID="physical_page_0" ORDER="0" ORDERLABEL="Einband" TYPE="page">
      <mets:fptr FILEID="ptr_04"/>
      <mets:fptr FILEID="ptr_03"/>
      <mets:fptr FILEID="ptr_02"/>
    </mets:div>
    <mets:div ID="physical_page_1" ORDER="1" ORDERLABEL="Einband" TYPE="page">
      <mets:fptr FILEID="ptr_14"/>
      <mets:fptr FILEID="ptr_13"/>
      <mets:fptr FILEID="ptr_12"/>
    </mets:div>
    <mets:div ID="physical_page_2" ORDER="2" ORDERLABEL="Vorsatz" TYPE="page">
      <mets:fptr FILEID="ptr_24"/>
      <mets:fptr FILEID="ptr_23"/>
      <mets:fptr FILEID="ptr_22"/>
    </mets:div>
  </mets:div>
</mets:structMap>
```

METS: Bibliographische Metadaten

- MODS Format (transformiert aus MAB-Record)
- abgelegt in DMD-Section von METS

```
<mets:dmdSec ID="dmd0">
  <mets:mdWrap MIMETYPE="text/xml" MDTYPE="MODS">
    <mods xmlns="http://www.loc.gov/mods/v3" version="3.3">
      <language>
        <languageTerm type="text">ger</languageTerm>
      </language>
      <name ID="name100b" type="personal" authority="pnd">
        <namePart>Ferdinand &lt;Römisches Reich, Kaiser, III.></namePart>
        <role>
          <roleTerm type="code" authority="marcrelator">asn</roleTerm>
        </role>
      </name>
      <titleInfo ID="title331">
        <title>Capitulation, Oder Capitul, Puncten und Articul, Darauff ein Römischer König/ zu einem Kayser
          zuerheben/ Von den Churfürsten deß H. Röm. Reichs erwöhlet und angenommen pflegt zu werden</title>
      </titleInfo>
      <titleInfo ID="title335">
        <subTitle>Mit Marginalien und einem Register</subTitle>
      </titleInfo>
      <note type="statement of responsibility">[Ferdinand &lt;Römisches Reich, Kaiser, III.>]</note>
      <originInfo>
        <place>
          <placeTerm type="text">Franckfurt am Mayn</placeTerm>
        </place>
      </originInfo>
    </mods>
  </mets:mdWrap>
</mets:dmdSec>
```


TEI: Zusätzliche Daten

- Enthält:
 - Logische Struktur des Buches
 - Volltext
 - Annotationen ?
- Automatische Erstellung beim Ingest...
 - Falls kein eigenes TEI vorhanden
 - Struktur kann im Webeditor bearbeitet werden
- ..oder Upload eines fertigen TEIs beim Ingest
 - Schema und Regeln müssen erfüllt sein (z.B. Anzahl pb = Anzahl hochgeladener Bilder)
 - Keine Bearbeitung im Struktur-Webeditor möglich
- XInclude nicht möglich: 1 TEI pro Werk
 - Externes Tool zum Zusammensetzen

TEI: Logische Struktur

```
<text>
<body>

<!-- cover - pages 1,2 -->
<pb facs="mpdl_mets_new.xml #physical_page_0"/>
<div type="cover">
  <pb facs="mpdl_mets_new.xml #physical_page_1"/>
</div>

<!-- fly-leaf - pages 3-6 -->
<pb facs="mpdl_mets_new.xml #physical_page_2"/>
<div type="fly-leaf">
  <pb facs="mpdl_mets_new.xml #physical_page_3"/>
  <pb facs="mpdl_mets_new.xml #physical_page_4"/>
  <pb facs="mpdl_mets_new.xml #physical_page_5"/>
</div>

<!-- title page - pages 7-8 -->
<pb facs="mpdl_mets_new.xml #physical_page_6"/>
<div type="title_page">
  <pb facs="mpdl_mets_new.xml #physical_page_7"/>
</div>

<!-- dedication - pages 9-10 -->
<pb facs="mpdl_mets_new.xml #physical_page_8"/>
  <div type="dedication">
    <pb facs="mpdl_mets_new.xml #physical_page_9"/>
  </div>

<!-- other - pages 11-14 -->
<pb facs="mpdl_mets_new.xml #physical_page_10"/>
<div type="other">
  <head>
    Hochgebohrner Reichs=Graff, Gnädigster Graff und Herr
  </head>
</div>

```

XML Datenbank

- Berkeley XML Database
 - Performancevorteile gegenüber eXist XMLDatenbank
- Enthält METS und TEI Dokumente
- Erlaubt Verwaltung zusätzlicher Metadaten pro Dokument (Ersteller, Bearbeitungsdatum etc.)
- Erlaubt
 - Schnelles Empfangen von Dokumenten und Dokumentteilen
 - Suche über XQuery
 - Updates von XML



eSciDoc Infrastructure

- Persistenzschicht
- Langzeitarchivierung
- Authorization&Authentication Instanz
- Nutzermanagement
- Übernimmt die Daten aus der XML-DB „on demand“