



MARTIN-LUTHER-UNIVERSITÄT HALLE-WITTENBERG
Universitäts- und Landesbibliothek Sachsen-Anhalt

URN Granular: Uniform Resource Names zur Identifizierung und Adressierung digitalisierter Drucke

Praxisbericht zur fachlichen Konzeption und technischen Umsetzung. Deutscher Bibliothekartag Erfurt 2009

Dr. Dorothea Sommer

dorothea.sommer@bibliothek.uni-halle.de

Kay Heiligenhaus M.A.

kay.heiligenhaus@semantics.de

Übersicht

- Projektkontext
- Persistente Adressierbarkeit von Netzressourcen: Anspruch
- Persistente Adressierbarkeit von Netzressourcen: Realität
- Zielstellung des Verfahrens URN Granular
- Projekt Ponickau: Workflow und Qualitätsmanagement
- URN-Konzept der DNB: Anwendung im Projekt Ponickau
- Das Prinzip der Granularität
- Das Harvestingverfahren
- Ergebnisse und Ausblick



Sammlung Ponickau

Sammlung Ponickau - Mozilla Firefox

Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe

http://bibliothek.uni-halle.de/dbib/digital/sammlung_ponickau/

Meistbesuchte Seiten Erste Schritte Aktuelle Nachrichten

Martin-Luther-Universität Halle-Wittenberg
**UNIVERSITÄTS- UND LANDESBIBLIOTHEK
SACHSEN-ANHALT**

Kataloge Image-Kataloge GBV KVK Datenbanken Zweigbibliotheken A-Z OPAC Suche im ULB-Katalog

Die Universität

- Universitäts- und Landesbibliothek (ULB)
- Digitale Bibliothek
- Digitale Sammlungen
- Sammlung Ponickau**
- Historische Drucke
- Hochschulschriften»
- Pflichtexemplare
- Vorderer Orient
- Lepsius Digital
- Historische Zeitungen
- Kartensammlung

Fakultäten

Zentrale Einrichtungen

Wissenschaftliche Zentren

Sammlung Ponickau

Die Sammlung



Die Bibliotheca Ponickaviana ist eine bedeutende Sondersammlung der **Universitäts- und Landesbibliothek Sachsen-Anhalt**. Sie stammt aus den Beständen der alten Wittenberger Universitätsbibliothek, die in großen Teilen nach der Vereinigung der Universitäten Wittenberg und Halle im Jahr 1817 an die hallesche Universitätsbibliothek kam. Zusammengetragen von dem Geheimen Kriegsrat Johann August von Ponickau (1718-1802), der sie der Wittenberger Universität im Jahr 1789 schenkte, und später kontinuierlich erweitert, ist sie eine einzigartige Sammlung von literarischen und anderen Zeugnissen zur Geschichte und Geographie der ehemaligen preußischen Provinz Sachsen und der anhaltinischen Gebiete, des Königreichs Sachsen und Thüringens.

Links

- Projektseite
- Detailsuche
- Neuzugänge
- Impressum
- Katalogrecherche
- DFG-Viewer

Weitere Einstellungen

- Schrift: größer + kleiner -
- Erweiterte Suche

Login für Redakteure

Benutzername:

Passwort:

Anmelden

Fertig



Ausgangssituation

- Mengengerüst:
 - 10.000 Drucke mit ca. 600.000 Seiten
- Projektlaufzeit:
 - 2 Jahre
- Physische Bestandsanalyse:
 - Heterogener Bestand im Quart-, Oktav-, Folioformat, eingebundene Materialien in variierenden Formaten, Kleinschrifttum, Sammelbände
- Speicherkapazität:
 - ca. 15 Terabyte
- Umsetzung der Aufgabenstellung in Kooperation der ULB mit zwei Dienstleistungsfirmen:
 - Mikro Univers Berlin (Scannen, Vorstrukturieren)
 - semantics Aachen (Verarbeitung, Webpräsentation)



Persistente Adressierbarkeit von Netzressourcen: Anspruch

DFG-Praxisregeln UA „Kulturelle Überlieferung“:

„In den Anfangszeiten der Digitalisierung hat man das Problem der Zitierfähigkeit digitaler Ressourcen meist unterschätzt. Dabei ist es gerade die Zitierbarkeit, die die wissenschaftliche Nutzung einer digitalisierten Quelle im Internet auch erlaubt. Im Unterschied zu vormaligen Sekundärformen wie Mikrofilm oder Papierausdruck ist die Internetressource nicht nur eine Kopie des Originals, die wie das Original behandelt und ergo zitiert wird, sondern ein selbständiges Objekt in einem sich dynamisch entwickelnden integralen Forschungsraum....Wird ...eine Kopie ins Internet gestellt, muss sie, um von anderen Dokumenten oder Datenbanken aus erreichbar zu sein, eindeutig angesprochen werden können.“

http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/aktuelles/download/praxisregeln_kulturelle_ueberlieferung_0208.pdf



- **DFG-Regeln guter wissenschaftlicher Praxis; Empfehlung 7/1998;**
„Primärdaten als Grundlagen für die Veröffentlichung sollen auf haltbaren und gesicherten Trägern in der Institution, wo sie entstanden sind, für 10 Jahre aufbewahrt werden.“
- **Allianz der Wissenschaftsorganisationen: Initiative Digitale Information
11.6.2008, Aktionsfeld Nr. 4**
„Auch hier gilt es, in den jeweiligen Disziplinen angemessene Strukturen und Prozesse zu etablieren, um Forschungsprimärdaten zu sichern, zu archivieren und für eine Nachnutzung bereit zu stellen“.
- **European Science Foundation (ESF)/ European Heads of Research Councils (EUROHORCS), Juni 2008: Vision European Research Area (ERA) mit dauerhaftem Zugriff zu qualitätsgesicherten Primärdaten (Nr. 8)**



Persistente Adressierbarkeit von Netzressourcen: Realität

- **27 Monate nach Veröffentlichung sind bis zu 13% der Online Ressourcen nicht mehr retrievalfähig**

Dellavalle, Robert P., Hester, Eric J., Heilig, Lauren F. u.a.: Information Science: Going, going, gone: Lost internet references. Science 302, no. 5646 (Oct. 31,2003), 787-788;

- **Online Zusatzmaterial in 6 Top-Zeitschriften aus dem STM-Bereich war nach zwei Jahren zu 4,7% und nach 5 Jahren zu 9,6% nicht mehr auffindbar**

Evangelou, Evangelos; Trikalinos, Thomas A.; Ionnadis, John P.A.: Unavailability of online supplementary scientific information from articles published in major journals. FASEB Journal 19 (2005) 14, S. 1943-44

- **Kein ausreichender Speicherplatz mehr: 2007 ist das Jahr, in dem erstmals die Menge der generierten digitalen Information die Speicherkapazitäten überschritten hat.**

(Gantz, J.F. The diverse and exploding digital universe: An updated forecast to worldwide information growth through 2011. International Data Corporation 2008)



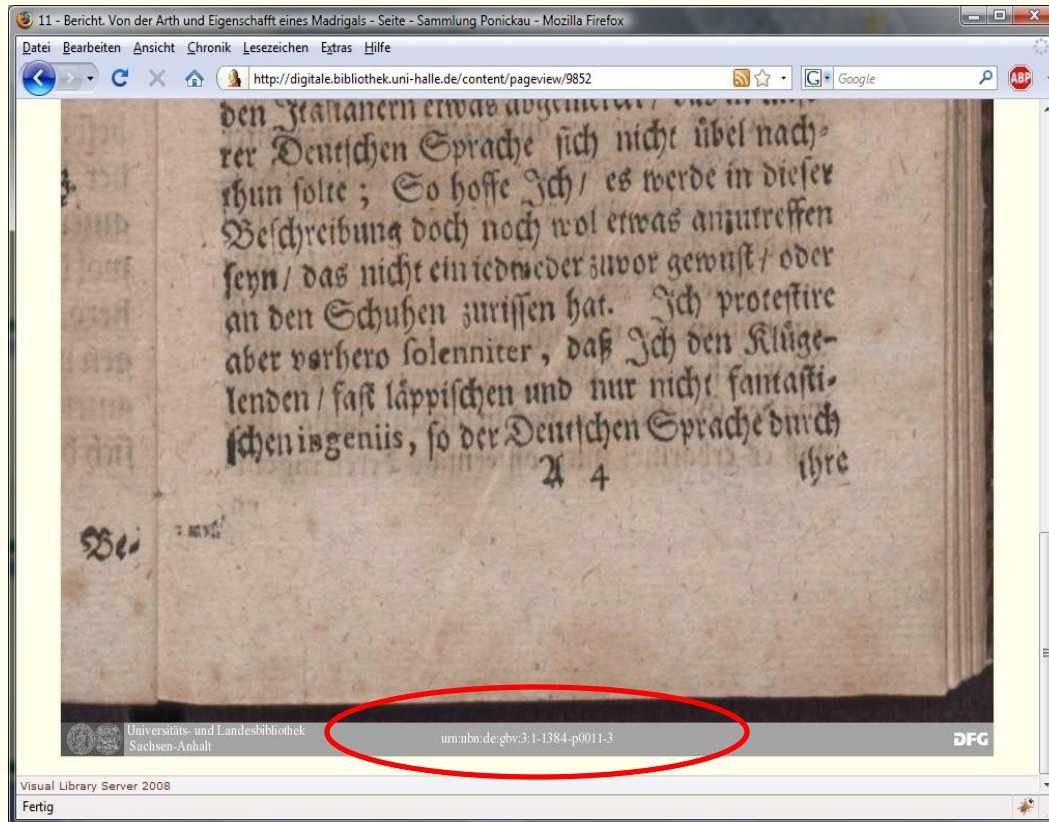
Lösungen:

- Datenselektion und -sicherung
- Orte: safe places, domains of trust
Nationalbibliotheken?
Pflichtexemplarbibliotheken?
- Metadaten (METS, MODS, etc.)
- Tools: Persistent Identifiers (URN, DOI, PUR, ARK, etc.)
- Konzepte und Verfahren: z. B. Persistentes Zitieren

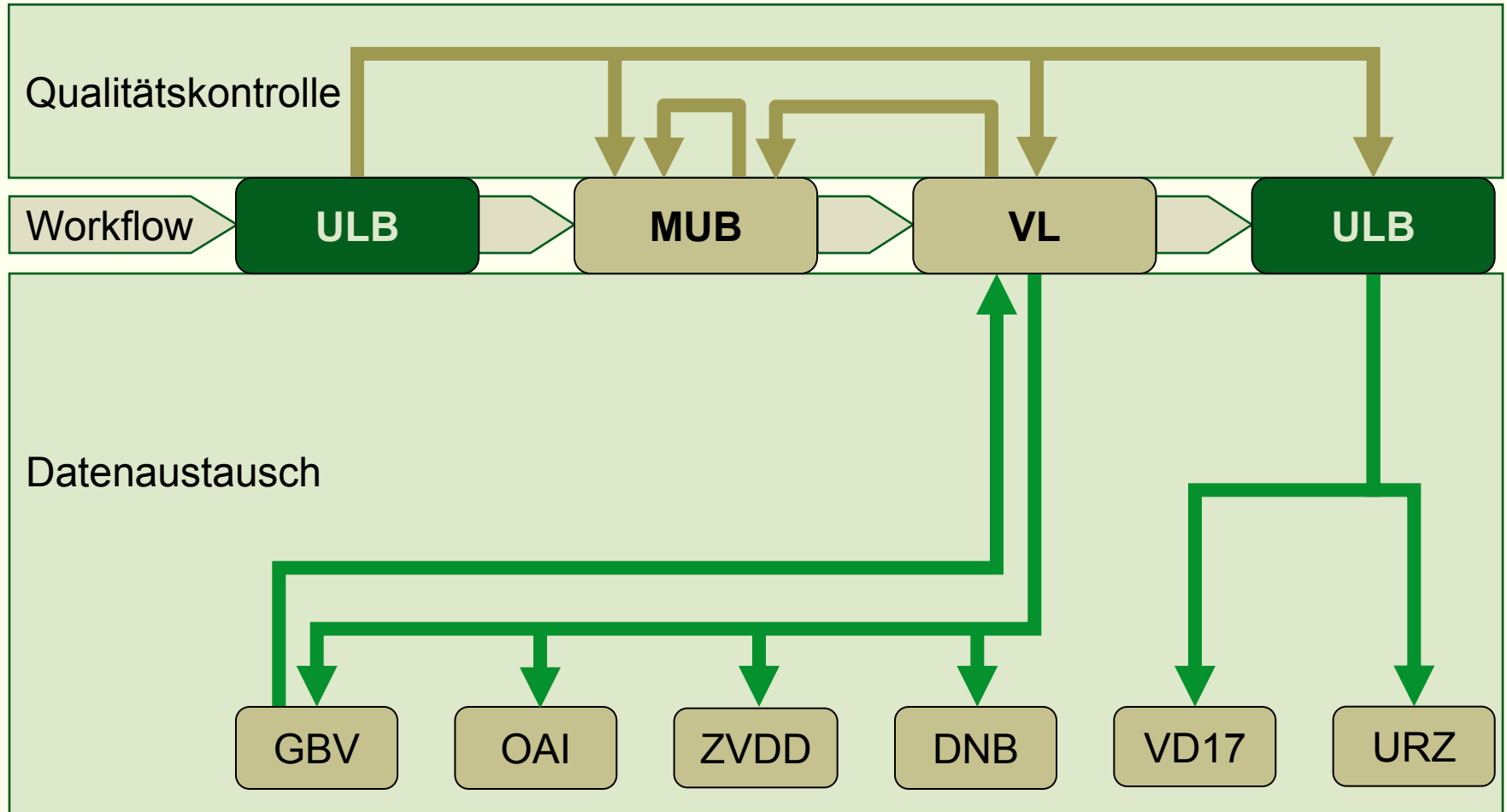


URN Granular: Zielstellung des Verfahrens

Langfristiges, verlässliches, nachhaltiges Zitieren eines digitalen Werkes **bis auf die Seitenebene.**



Projekt Ponickau: Qualitätskontrolle, Workflow und Datenaustausch



Qualitätssicherung: Mögliche Fehlerquellen

- Bei der Anordnung der Images durch
 - falsch angehängte Images in den Drucken
 - überblätterte Seiten beim Scannen
 - fehlerhafte Seitenzählung
 - überzählige Bilder
- Beim Farbmanagement durch
 - verfärbte Images
 - verzerrte Seitenwiedergabe beim Scannen von umfangreichen Konvoluten
 - fehlende Farbprofile
 - fehlende Colorchecker
- Bei der Auszeichnung von Strukturdaten durch
 - falsche Auszeichnung der Strukturen

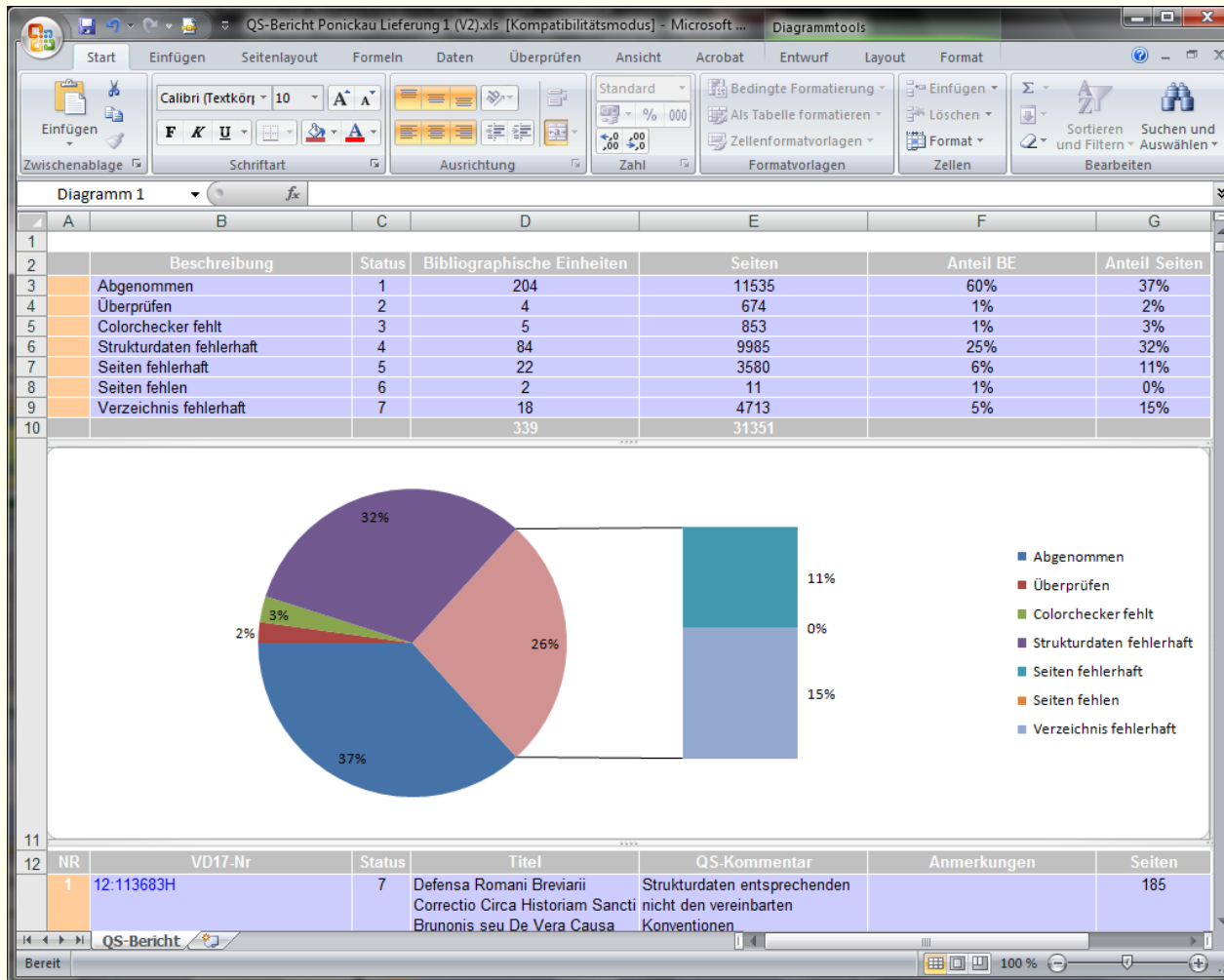


Verfahren der Qualitätssicherung

1. Vorgelagerte Arbeiten und Festlegungen
2. Laufende Maßnahmen der Qualitätssicherung
3. Kontrolle an den einzelnen Stationen des Digitalisierungsworkflows
4. ggf. Feedback der Nutzer



Qualitätssicherungsprotokoll



URN-Konzept der DNB: „URN-Objekte“

▪ Definition „URN-Objekt“

Im Rahmen des URN-Managements wird unter dem Begriff „digitales Objekt“ eine Einheit verstanden, für die eine URN vergeben werden kann. Eine Einheit bezieht sich im derzeitigen Stand auf eine statische Publikation wie z.B. monografische Online-Ressourcen. [...] Die kleinste Einheit des digitalen Objektes ist über einen im Internet allgemein verwendbaren Zugriffsmechanismus wie z.B. URL adressierbar.

(<http://www.persistent-identifizier.de/?link=3352>)



Präsentation: Werk-Adressierung

The screenshot shows a Mozilla Firefox browser window displaying a digital library record. The address bar contains the URL <http://digitale.bibliothek.uni-halle.de/content/titleinfo/9834>, which is circled in red. The page title is 'Von den Madrigalen - Titel - Sammlung Ponickau - Mozilla Firefox'. The browser's menu bar includes 'Datei', 'Bearbeiten', 'Ansicht', 'Chronik', 'Lesezeichen', 'Extras', and 'Hilfe'. The page header features the logos of the Universitäts- und Landesbibliothek Sachsen-Anhalt and the 'Sammlung Ponickau' logo. A navigation bar includes 'finden', 'Detailsuche', 'Home', 'Titel', 'Impressum', and 'Hilfe'. The main content area is titled 'Sammlung Ponickau' and includes a 'Titel' tab, 'Inhalt', 'Übersicht', 'Vollbild', and 'Titel zitieren' options. A left sidebar contains sections for 'Download' (Von den Madrigalen [10,94 mb]), 'Archiv' (METS (OAI-PMH)), 'Verfügbarkeit' (In meiner Bibliothek), and 'Nachweis' (VD17 - Universitäts- und Landesbibliothek Sachsen-Anhalt). The central image shows the book cover of 'Von den Madrigalen' with a small label 'Hd 372'. The right sidebar displays the 'Titelaufnahme' (Title Record) with the following details:

Titel	Caspar Ziegler bequemesten ... Deutschen Spra Anitzo vermehr Register	ur Musik ... unserer mpeln/ fügen
Verfasser	Ziegler, Kaspar W	
Beiträger	Schütz, Heinrich W	
Verleger	Hartmann, Andreas <Witwe> W Brüning, August <Witwe> W	
Erschienen	Wittenberg : Hartmann ; Wittenberg : Brüningische Schriften, 1685	
Online-Ausg.	[Online-Ausg.] Halle, Saale : Universitäts- und Landesbibliothek Sachsen-Anhalt, 2008	
Umfang	[3] Bl., 51 S., [3] Bl ; 8°	
Sprache	Deutsch	
VD17	23:285066U	
Gattungsbegriff	Musikbuch / Poetik / Gedichtsammlung	

A 'Persistenten Link kopieren' dialog box is open, showing the persistent link <http://n-resolving.de/urn:nbn:de:gbv:3:1-1384>, which is also circled in red. The dialog box text reads: 'Die persistente Auflösung dieses Links erfolgt durch den URN-Resolver der Deutschen Nationalbibliothek'. The footer of the browser window shows 'Visual Library Server 2008' and 'Fertig'.



Das Prinzip der „Granularität“ / „Teil-Objekte“

Granularität und Ausprägung der Objekte

Eine URN verweist auf mindestens eine URL des Gesamtobjektes, über die ein Objekt adressiert wird. Eine URN kann auch mehrere Kopien desselben Objektes, die über verschiedene URLs zugänglich sind, verwalten sowie unterschiedliche Präsentationsformate der Objekte einschliessen. [...] URNs können für adressierbare Teilobjekte vergeben werden.

(<http://www.persistent-identifier.de/?link=3352>)



Präsentation: Einzelseiten-Adressierung

11 - Bericht. Von der Arth und Eigenschafft eines Madrigals - Seite - Sammlung Ponickau - Mozilla Firefox

Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe

<http://digitale.bibliothek.uni-halle.de/content/pageview/9852>

Google

den ...
rer Deutschen Sprache sich nicht übel nach-
thun solte; So hoffe Ich/ es werde in dieser
Beschreibung doch noch wol etwas anzutreffen
seyn/ das nicht ein tedweder zuvor gewunst/ oder
an den Schuhen zurißten hat. Ich protestire
aber vorhero solenniter, daß Ich den Klüge-
lenden / fast läppischen und nur nicht fantasti-
schen isgeniis, so der Deutschen Sprache durch
A 4 ihre

Bei

Universitäts- und Landesbibliothek
Sachsen-Anhalt

urn:nbn:de:gbv:3:1-1384-p0011-3

DFG

Visual Library Server 2008

Fertig



Syntax: URN mit Seitenreferenzierung

Basis: X-Epicur-Format

urn:nbn:de:gbv:3:1-xxxx-p0001-3

<urn des Werkes>-p<physische Seitenzählung, vierstellig, mit führenden Nullen>-<Prüfziffer>



Datenaustausch DNB: xepicur

OAI - Sammlung Ponickau - Mozilla Firefox

http://digitale.bibliothek.uni-halle.de/oai/?verb=GetRecord&metadataPrefix=epicur&i

```
<epicur:epicur>
  <administrative_data>
    <delivery>
      <update_status type="urn_new"><update_status>
    </delivery>
  </administrative_data>
  <record>
    <identifier scheme="urn:nbn:de" >urn:nbn:de:gbv:3:1-1384<identifier>
    <resource>
      <identifier scheme="url" role="primary" type="frontpage" >http://digitale.bibliothek.uni-halle.de/urn/urn:nbn:de:gbv:3:1-1384<identifier>
      <format scheme="text/html" </format>
    </resource>
    <isPartOf>
      <identifier scheme="urn:nbn:de" >urn:nbn:de:gbv:3:1-1384-p0001-7<identifier>
      <resource>
        <identifier scheme="url" role="primary" >http://digitale.bibliothek.uni-halle.de/pum/urn:nbn:de:gbv:3:1-1384-p0001-7<identifier>
        <format scheme="text/html" </format>
      </resource>
    </isPartOf>
    <isPartOf>
      <identifier scheme="urn:nbn:de" >urn:nbn:de:gbv:3:1-1384-p0002-3<identifier>
      <resource>
        <identifier scheme="url" role="primary" >http://digitale.bibliothek.uni-halle.de/pum/urn:nbn:de:gbv:3:1-1384-p0002-3<identifier>
        <format scheme="text/html" </format>
      </resource>
    </isPartOf>
    <isPartOf>
      <identifier scheme="urn:nbn:de" >urn:nbn:de:gbv:3:1-1384-p0003-8<identifier>
      <resource>
        <identifier scheme="url" role="primary" >http://digitale.bibliothek.uni-halle.de/pum/urn:nbn:de:gbv:3:1-1384-p0003-8<identifier>
        <format scheme="text/html" </format>
      </resource>
    </isPartOf>
    <isPartOf>
      <identifier scheme="urn:nbn:de" >urn:nbn:de:gbv:3:1-1384-p0004-3<identifier>
      <resource>
        <identifier scheme="url" role="primary" >http://digitale.bibliothek.uni-halle.de/pum/urn:nbn:de:gbv:3:1-1384-p0004-3<identifier>
        <format scheme="text/html" </format>
      </resource>
    </isPartOf>
  </record>
</epicur:epicur>
```

Fertig



Datenaustausch DFG-Viewer: METS

OAI - Sammlung Ponickau - Mozilla Firefox

Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe

http://digitale.bibliothek.uni-halle.de/oai/?verb=GetRecord&metadataPrefix=mets&id=

```
<mets:structMap TYPE="PHYSICAL">
  <mets:div ID="phys9834" TYPE="physSequence">
    <mets:fptr FILEID="PDF_9834" ></mets:fptr>
    <mets:div ID="phys9839" TYPE="page" LABEL="[Seite 1]" CONTENTIDS="urn:nbn:de:gbv:3:1-1384-p0001-7" ORDER="1">
      <mets:fptr FILEID="IMG_DEFAULT_9839" ></mets:fptr>
      <mets:fptr FILEID="IMG_THUMBS_9839" ></mets:fptr>
      <mets:fptr FILEID="IMG_MIN_9839" ></mets:fptr>
      <mets:fptr FILEID="IMG_MAX_9839" ></mets:fptr>
      <mets:fptr FILEID="PDF_9839" ></mets:fptr>
    </mets:div>
    <mets:div ID="phys9840" TYPE="page" LABEL="[Leerseite]" CONTENTIDS="urn:nbn:de:gbv:3:1-1384-p0002-3" ORDER="2">
      <mets:fptr FILEID="IMG_DEFAULT_9840" ></mets:fptr>
      <mets:fptr FILEID="IMG_THUMBS_9840" ></mets:fptr>
      <mets:fptr FILEID="IMG_MIN_9840" ></mets:fptr>
      <mets:fptr FILEID="IMG_MAX_9840" ></mets:fptr>
      <mets:fptr FILEID="PDF_9840" ></mets:fptr>
    </mets:div>
    <mets:div ID="phys9841" TYPE="page" LABEL="[Leerseite]" CONTENTIDS="urn:nbn:de:gbv:3:1-1384-p0003-8" ORDER="3">
      <mets:fptr FILEID="IMG_DEFAULT_9841" ></mets:fptr>
      <mets:fptr FILEID="IMG_THUMBS_9841" ></mets:fptr>
      <mets:fptr FILEID="IMG_MIN_9841" ></mets:fptr>
      <mets:fptr FILEID="IMG_MAX_9841" ></mets:fptr>
      <mets:fptr FILEID="PDF_9841" ></mets:fptr>
    </mets:div>
    <mets:div ID="phys9842" TYPE="page" LABEL="[Leerseite]" CONTENTIDS="urn:nbn:de:gbv:3:1-1384-p0004-3" ORDER="4">
      <mets:fptr FILEID="IMG_DEFAULT_9842" ></mets:fptr>
      <mets:fptr FILEID="IMG_THUMBS_9842" ></mets:fptr>
      <mets:fptr FILEID="IMG_MIN_9842" ></mets:fptr>
      <mets:fptr FILEID="IMG_MAX_9842" ></mets:fptr>
      <mets:fptr FILEID="PDF_9842" ></mets:fptr>
    </mets:div>
    <mets:div ID="phys9844" TYPE="page" LABEL="[Seite 5]" CONTENTIDS="urn:nbn:de:gbv:3:1-1384-p0005-9" ORDER="5">
      <mets:fptr FILEID="IMG_DEFAULT_9844" ></mets:fptr>
      <mets:fptr FILEID="IMG_THUMBS_9844" ></mets:fptr>
      <mets:fptr FILEID="IMG_MIN_9844" ></mets:fptr>
      <mets:fptr FILEID="IMG_MAX_9844" ></mets:fptr>
      <mets:fptr FILEID="PDF_9844" ></mets:fptr>
    </mets:div>
    <mets:div ID="phys9845" TYPE="page" LABEL="[Leerseite]" CONTENTIDS="urn:nbn:de:gbv:3:1-1384-p0006-4" ORDER="6">
      <mets:fptr FILEID="IMG_DEFAULT_9845" ></mets:fptr>
      <mets:fptr FILEID="IMG_THUMBS_9845" ></mets:fptr>
      <mets:fptr FILEID="IMG_MIN_9845" ></mets:fptr>
      <mets:fptr FILEID="IMG_MAX_9845" ></mets:fptr>
      <mets:fptr FILEID="PDF_9845" ></mets:fptr>
    </mets:div>
  </mets:structMap>
```

Fertig



Verwaltung: Einzelseiten-Informationen

The screenshot displays the 'Portal - Library Manager 2008' interface. The main window shows a manuscript page with a decorative header and the title 'Bericht Von der Arth und Eigenschafft eines Madrigals.' The left sidebar contains a tree view of the document's structure, with 'Seite 1' highlighted. The right sidebar shows the 'Eigenschaften' (Properties) panel, which is divided into several sections:

- Angaben zur Seite:** Includes 'Typ' (Seite), 'Graphie' (<kein Eintrag>), and 'Beschreibung' (Seite 1).
- Angaben zur Packierung:** Includes 'Typ' (Arabisch) and 'Wert' (1).
- Identifier:** Includes 'URN' (urn:nbn:de:gbv:3:1-1384-p0011-3) and 'VLID' (9853).
- Status:** Includes 'Freigegeben' (checked).
- Bildinformationen:** Includes 'Software' (MUB - Zeutschel OS 11.8, Scanner), 'Zeitstempel' (2007-08-27 06:56:48), 'Seitenname' (23-285066U_0006_T_6\23-285066), 'Breite' (1243 Pixel), 'Höhe' (1971 Pixel), 'X-Auflösung' (300 dpi), 'Y-Auflösung' (300 dpi), 'Farbtiefe' (24 bpp), and 'Größe' (7.65 mb).
- Historie:** Includes 'Angelegt' (2007-09-29 22:32:54.934000) and 'Geändert' (2009-03-15 12:36:43.308000).

Red circles highlight the 'Seite' entry in the 'Angaben zur Seite' section, the 'Arabisch' entry in the 'Angaben zur Packierung' section, the 'URN' and 'VLID' entries in the 'Identifier' section, the 'Software' and 'Zeitstempel' entries in the 'Bildinformationen' section, and the 'Angelegt' and 'Geändert' entries in the 'Historie' section.



Konsistenzprüfung: Resolver der DNB

Mozilla Firefox

http://nbn-resolving.de/urn:nbn:de:gbv:3:1-1384-p0011-3?n2ls&xml

Mit dieser XML-Datei sind anscheinend keine Style-Informationen verknüpft. Nachfolgend wird die Baum-Ansicht des Dokuments angezeigt.

```

<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
<epicur>
  <header>
    <request>urn:nbn:de:gbv:3:1-1384-p0011-3&n2ls&xml</request>
    <responseDate>2009-03-15 13:20:38</responseDate>
  </header>
  <record>
    <identifier scheme="urn:nbn:de" urn:nbn:de:gbv:3:1-1384-p0011-3</identifier>
    <resource>
      <identifier scheme="url" role="primary">
        http://digitale.bibliothek.uni-halle.de/purn/urn:nbn:de:gbv:3:1-1384-p0011-3
      </identifier>
      <format scheme="imt" text/html</format>
    </resource>
  </record>
</epicur>
  
```

Fertig

Zusammenfassung URNs

Datentyp	Gesamt	Gelöscht	Freigegeben	Nicht freigegeben	URN	Keine URN	Geprüft (DNB)	Ungeprüft (DNB)	Registriert (DNB)	Nicht registriert	Nicht registriert (Frei)
bibliogr. Einheiten	<u>10651</u>	0	<u>9448</u>	<u>1203</u>	<u>10143</u>	<u>508</u>	<u>10143</u>	0	<u>8881</u>	<u>1251</u>	<u>558</u>
Seiten	<u>582026</u>	0	<u>536175</u>	<u>45851</u>	<u>560844</u>	<u>21182</u>	560844	<u>0</u>	478869	51087	30888



Ergebnisse (Stand 6/2009)

- Verfahren befindet sich seit Frühjahr 2008 im Regelbetrieb und kann nachgenutzt werden;
- 10.147 Drucke persistent auf Werkebene zitierbar;
- 581.347 Digitalisate auf Seitenebene zitierbar;
- Insgesamt: 591.494 URNs vergeben (Werk- und Seitenebene)
- Stabile Harvestingprozesse.



Wie weiter?

Siehe:

Dorothea Sommer, Christa Schöning-Walter, Kay Heiligenhaus: URN granular: Persistente Identifizierung und Adressierung von Einzelseiten digitalisierter Drucke. Ein Projekt der Deutschen Nationalbibliothek und der Universitäts- und Landesbibliothek Sachsen-Anhalt. *ABI-Technik*. 28 (2008), Heft 2, S. 97-106

Für Rückfragen:

dorothea.sommer@bibliothek.uni-halle.de

k.heiligenhaus@semantics.de

