



Einführung und praktischer Umgang mit PIs

für Digitalisate innerhalb der Produktionsumgebung Goobi

Ralf Stockmann <stockmann@sub.uni-goettingen.de>

Jochen Kothe <kothe@sub.uni-goettingen.de>

NIEDERSÄCHSISCHE STAATS- UND
UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN 

 **GDZ**
Göttinger Digitalisierungszentrum

Gliederung



- Rolle von PIs im Kontext der Massendigitalisierung
- Evaluation von PI Varianten
- Konkrete Implementierung von URNs
- Problemfelder

Anforderungen an PIs in der Massendigitalisierung



- Ab heute: VD18 Test
 - 40 Bände pro Tag durch alle Stationen
 - Bedeutet: 40 fehlerfreie URN-Vermittlungen durch DNB, fehlerrobust
- Workflow-Tools
- Automatisierte Schritte (OCR, Imaging, PI-Vergabe?)
- Für Anwender transparent

Aktuelle Vorgänge

Treffer: 852

Vorgangstitel	Status	Projekt																								
rabegott_PPN538327944		DigiWunschBuch																								
VD17Berlin_Monographie		DigiWunschBuch																								
VD17Berlin_Zeitschrift		DigiWunschBuch																								
bucegerm_PPN599575697_0004		DigiWunschBuch																								
grunervox_PPN561207860		DigiWunschBuch																								
friefrie_PPN576755788		DigiWunschBuch																								
schljouj_PPN599511834																										
<table border="1"><thead><tr><th>Nr.</th><th>Titel</th><th>Status</th></tr></thead><tbody><tr><td>1</td><td>Bibliographische Aufnahme</td><td></td></tr><tr><td>2</td><td>Scannen</td><td></td></tr><tr><td>3</td><td>Qualitätskontrolle</td><td></td></tr><tr><td>4</td><td>Imagenachbearbeitung</td><td></td></tr><tr><td>5</td><td>Struktur- und Metadaten</td><td></td></tr><tr><td>6</td><td>Import DMS</td><td></td></tr><tr><td>7</td><td>Archivierung</td><td></td></tr></tbody></table>	Nr.	Titel	Status	1	Bibliographische Aufnahme		2	Scannen		3	Qualitätskontrolle		4	Imagenachbearbeitung		5	Struktur- und Metadaten		6	Import DMS		7	Archivierung			DigiWunschBuch
Nr.	Titel	Status																								
1	Bibliographische Aufnahme																									
2	Scannen																									
3	Qualitätskontrolle																									
4	Imagenachbearbeitung																									
5	Struktur- und Metadaten																									
6	Import DMS																									
7	Archivierung																									
steffen_test_ackesyst_8734655245		DigiWunschBuch																								
klenneue_PPN599032928		DigiWunschBuch																								
schlentw_PPN599517972		DigiWunschBuch																								
schlnest_PPN599524510_0003		DigiWunschBuch																								
schlneuj_PPN599542888_0001		DigiWunschBuch																								

1. Zielsetzung
2. Auswahl des Materials
3. Benchmarking / Standards
4. Produktionsorganisation / Logistik

5. Arbeitsvorbereitungen (AV)
6. Digitalisierung
7. Qualitätskontrolle (QK)
8. Bildoptimierung / Verarbeitung
9. OCR / Volltextgenerierung
10. Metadatengenerierung

PRODUCTION

11. Präsentation / Verwertung
12. Zugang / Evaluation
13. (Langzeit) Archivierung

PRESENTATION

1. Zielsetzung
2. Auswahl des Materials
3. Benchmarking / Standards
4. Produktionsorganisation / Logistik

5. Arbeitsvorbereitungen (AV)
6. Digitalisierung
7. Qualitätskontrolle (QK)
8. Bildoptimierung / Verarbeitung
9. OCR / Volltextgenerierung
10. Metadatengenerierung

PRODUCTION

11. Präsentation / Verwertung
12. Zugang / Evaluation
13. (Langzeit) Archivierung

PRESENTATION



Evaluation



- Studie Hilse/Kothe 2007
- Open Acces: PDF auf www.cerl.org
- Vergleich der verschiedenen PI Systeme
 - Technik
 - Einsatzgebiete
 - Kosten
- Mittelmäßige Bewertung für URN, Grund: zentrales Resolving fehleranfällig

Entscheidung der SUB für URN (granular)



- Alle machen es
- Kommt aus der Bibliothekszene (Nationalbibliotheken)
- DNB als kompetenter und engagierter Partner
- Kooperative Weiterentwicklung möglich (granular)

Generierung

- NOID-Server als Webservice
 - frei konfigurierbares Präfix
 - ID aus beliebigen alphanumerischen Zeichen

```
Präfix          - ID  
urn:nbn:de:gbv-7-goobi-12945
```

- frei definierbare „Key – Values“

```
myUrl = http://resolver.sub...
```


Integration

- Primärer METS-Export löst PI Generierung aus
 - Strukturebene

```
- <MODS:mods>  
  - <MODS:identifier type="purl">  
    http://resolver.sub.uni-goettingen.de/purl?PPN348459866  
  </MODS:identifier>  
  <MODS:identifier type="urn">urn:nbn:de:gbv-7-goobi-12944</MODS:identifier>  
- <MODS:recordInfo>
```

- Seitenebene

```
- <METS:structMap TYPE="PHYSICAL">  
  - <METS:div ID="phys1" ADMID="amd1" DMDID="dmdlog1" TYPE="physSequence">  
    - <METS:div ID="phys2" CONTENTIDS="urn:nbn:de:gbv-7-goobi-12946" ORDER="1">  
      <METS:fptr FILEID="fi1_800"/>  
      <METS:fptr FILEID="fi1_500"/>
```

- Verschiedene Download-Formate (PDF, ...)

DMS - Import

- Indexierung von Metadaten und Volltext
- Versionierung der METS-Dateien
- Bereitstellung der Master TIFs

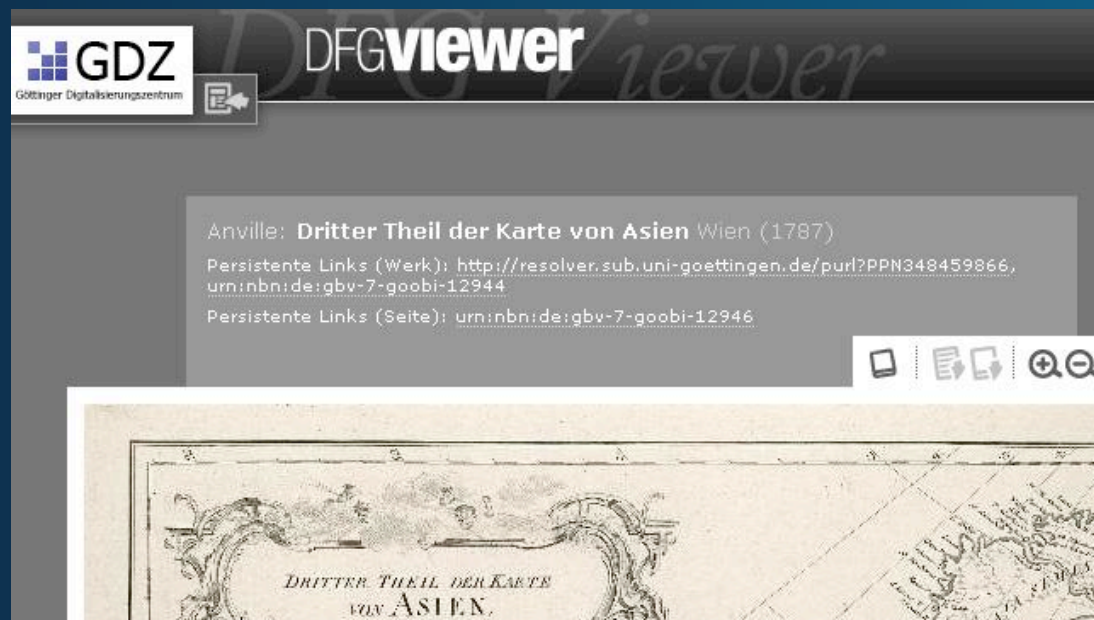
**LETZTE MÖGLICHKEIT
PI - RELEVANTE
DATEN ZU KORRIGIEREN**

Nutzung

- Persistente Links

PURL: <http://resolver.sub.uni-goettingen.de/purl?PPN348459866>
URN: <urn:nbn:de:gbv-7-goobi-12944>
PPN (digital): PPN348459866
PPN (original): PPN000214817
Links:   

- DFG - Viewer



Fallstricke: Fehlerkorrekturen

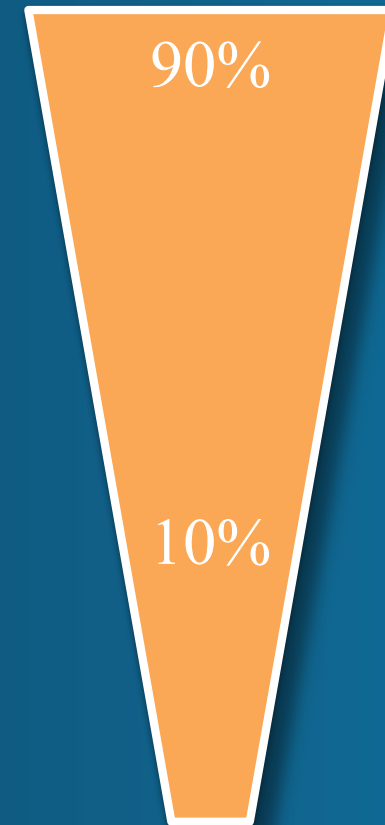
- Ca. 1% der Werke des GDZ „kommen zurück“
- 8.000 Bände in 10 Jahren, = 8 pro Jahr
- Aber: bei mind. 40 Bänden pro Tag (siehe VD 18): alle drei Tage eine Korrektur



Fallstricke

Jede Korrektur ein neuer PI?

- Tipp- / OCR - Fehler in Metadaten
 - Imageset ist in Ordnung, aber Fehler in
 - Bibliographischen
 - Technischen
 - Strukturellen
- Korrekturen am Imageset
 - Fehlende Seiten ergänzen
 - Seitendreher korrigieren
 - Optimierung bestehender Images (Flecken entfernen, Wasserzeichen sichtbar machen)
- Grundfrage: welche Bereiche müssen zitierpersistent gehalten werden? Neue „Auflagen“ des digitalen Werkes?



Fallstricke

- „Temporäre Dokumente“ - eigener PI?
 - Generierte PDFs
 - Skalierte Images
- Metadaten in mehreren Repositories?
 - Lokale Präsentation
 - ZVDD
 - DFG-Viewer
- ...

Integration in Goobi 1.6: September 09

Vielen Dank...

...Fragen?

Ralf Stockmann <stockmann@sub.uni-goettingen.de>

Jochen Kothe <kothe@sub.uni-goettingen.de>

NIEDERSÄCHSISCHE STAATS- UND
UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN 

 **GDZ**
Göttinger Digitalisierungszentrum