



# Kooperative Angebote von GBV und GDZ im Bereich Digitalisierung

Ralf Stockmann <[stockmann@sub.uni-goettingen.de](mailto:stockmann@sub.uni-goettingen.de)>

NIEDERSÄCHSISCHE STAATS- UND  
UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN 

 **GDZ**  
Göttinger Digitalisierungszentrum

**GBV**  
Gesamtkatalog 

# GDZ: Laufende Projekte (Auszug)

- SSG Digitalisierung (Mathematik)
  - VD17 / VD18 (seit heute 6:00: Vollasttest)
  - EZOOLO 2
  - Rezensionenzeitschriften
  - DigiWunschbuch
- 
- > 7 Mio. Seiten
  - Unterschiedliche Erschließung



# Anforderungen an Massendigitalisierung



- Workflow-Steuerung
- Computing-Kapazität
  - Suchindex, Rendering, ...
- Storage
- Verlässlichkeit
- Ansprechpartner die „unsere Sprache“ verstehen

**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>Imageneachbearbeitung</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	Imageneachbearbeitung		5	Struktur- und Metadaten		6	Import DMS		7	Archivierung			DigiWunschBuch
Nr.	Titel	Status																								
1	Bibliographische Aufnahme																									
2	Scannen																									
3	Qualitätskontrolle																									
4	Imageneachbearbeitung																									
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																								

# Strategische Ziele

- Kein eigenes „Blech“ warten (Outsourcing des Betriebs in Rechenzentren)
- Angebot „von der Stange“ das an die Projektbedürfnisse angepasst werden kann (Design, spezialisiertes Retrieval)
- Garantierte Kompatibilität zu DFG-Richtlinien (METS, TIFF, PDF, DFG-Viewer, TEI-Volltext)
- Skalierbare Lösungen, auch für kleine Standorte sinnvoll nutzbar
- Partner der „unsere Sprache“ versteht

# Drei Schritte:



# 1. Schritt

## Volltexterkennung (OCR)

- Grundlage: ABBYY Finereader Recognition Server
- 16 Kerne
- 80 Seiten pro Minute
- Konvertierung des Finereader-XML Outputs in TEI P5
- Wortkoordinaten für Image-Highlighting
- Derzeit nicht: Fraktur
  
- In Arbeit: Fehlerrobuste Übergabe aus Goobi heraus, für Anwender transparent

# OCR

1. Zielsetzung
2. Auswahl des Materials
3. Benchmarking / Standards
4. Produktionsorganisation / Logistik
5. Arbeitsvorbereitungen (AV)
6. Digitalisierung
7. Qualitätskontrolle (QK)
8. Bildoptimierung / Verarbeitung
9. Metadatengenerierung
10. Präsentation / Verwertung
11. Zugang / Evaluation
12. (Langzeit) Archivierung



# Seitenpreis: ca. 1 Cent





# 2. Schritt

## Storage

- Produktionsstorage für die Wochenproduktion
- Hintergelagerter Archivstorage für die Master-Files
- Schneller Storage für die Präsentationsexemplare
  
- Im Einzelfall zu klären:
  - Datentransport
  - Sogar Scannen über das Netz möglich?
  
- Preis: im Einzelfall zu errechnen

# 3. Schritt

## Goobi-Hosting

- Apache, MySQL, PHP, JAVA, Tomcat, Lucéne, TYPO3, OAI, RSS, WebDAV, PDF-Renderer, ...
- Server-Bereitstellung sowohl der Produktions- als auch der Präsentationsumgebung
- Wartung und Updates durch verschiedene Dienstleister (eventuell auch GBV selbst)
- Vorgefertigte Layouts und Workflow-Vorlagen
- Verfügbarkeit und Preis noch zu klären

# Drei Schritte: Massendigitalisierung für jede Einrichtung



1. Volltexterkennung (OCR)
2. Storage-Hosting
3. Hosting der Produktions- und Präsentationsumgebung (Goobi)

# Vielen Dank...

# ...Fragen?

Ralf Stockmann ([stockmann@sub.uni-goettingen.de](mailto:stockmann@sub.uni-goettingen.de))

NIEDERSÄCHSISCHE STAATS- UND  
UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN 

 **GDZ**  
Göttinger Digitalisierungszentrum