



THÜRINGER UNIVERSITÄTS- UND LANDESBIBLIOTHEK JENA

OCR in MyCoRe

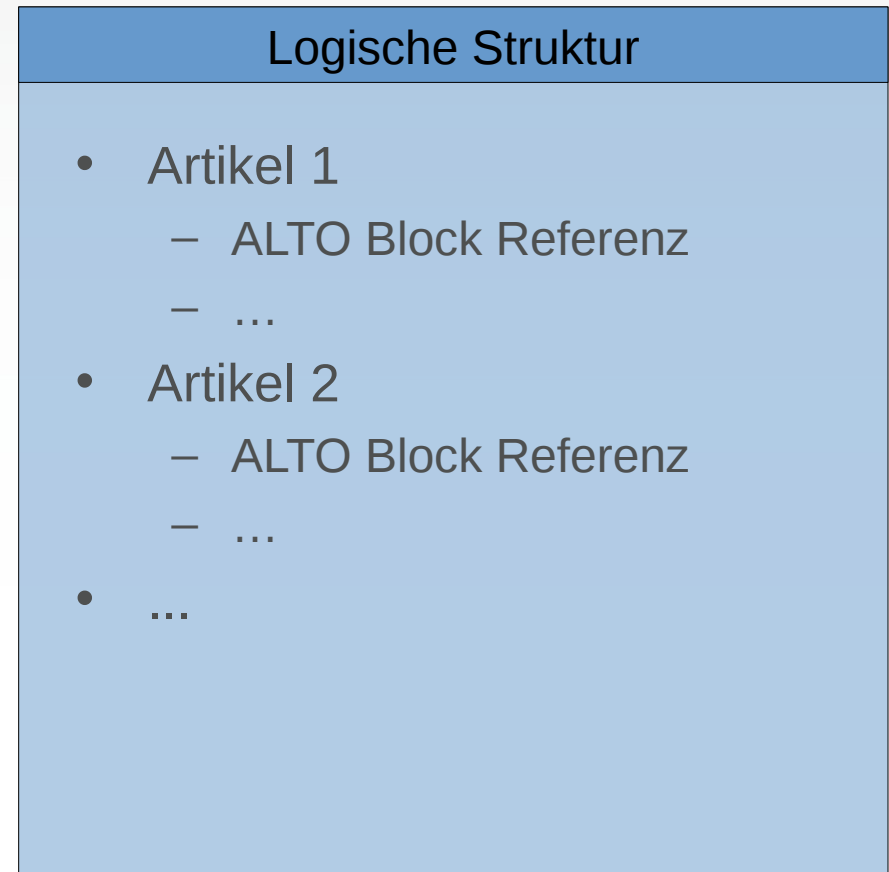
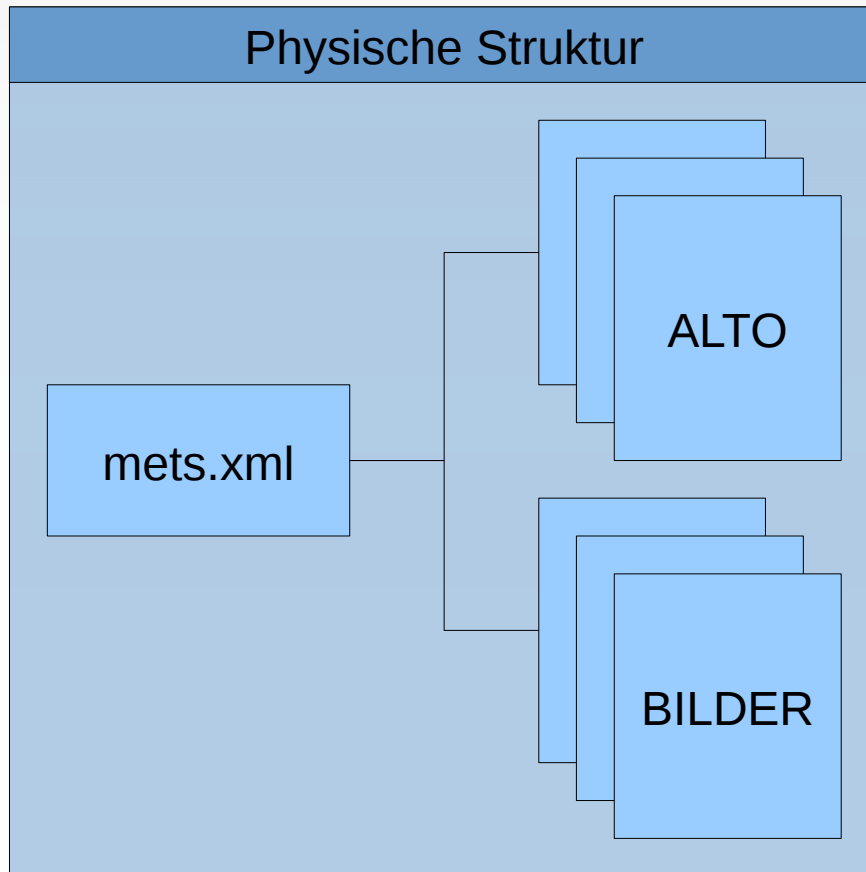
- Präsentation ([link](#))
- Formate
 - ALTO
 - METS
 - MCR Derivate
- METS Validator
- Projekt Workflow am Beispiel des „Jenaer Volksblatt“

Formate - ALTO

- Library of Congress
- XML Format
- Eine Datei pro Seite
- Block → Zeile → Wort



- Library of Congress
- Containerformat



- METS Profile
- MyCoRe METS Format
- <http://141.35.20.245:8081/mets-validator/>
- Java Library
 - <https://mycore.de/svn/maven/mets-validator/trunk/>
- Webapp Code
 - <https://mycore.de/svn/maven/mets-validator-webapp/trunk/>

- von April 1890 bis Mai 1941 erschienen
- „liberale“ Zeitung Jenas
- DFG Projekt der ThULB und Stadtarchiv Jena
- Zurzeit zehn Jahrgänge OCR erkannt

- Digitalisierungszentrum in der ThULB
 - Graustufenscan mit 400dpi
 - Automatisiertes Ausschneiden und Geraderücken
 - Manuelle Kontrolle mit Lightroom
 - Dateinamenbenennung nach Jahrgang, Ausgabe etc.
- Keine Abbyy Lizenz für OCR
 - Abbyy Marktführer für OCR
- Ausschreibung
 - Vorgabe von METS/ALTO nach MyCoRe Profil
 - Automatisierte Artikelerkennung
 - Programm für die Nachstrukturierung
- → UIBK Innsbruck

- Bände scannen und bearbeiten
- Digitalisate nach Innsbruck senden
- Auf die OCR Ergebnisse warten (METS/ALTO)
- Nachstrukturierung mit Structify
 - Artikel Separierung
 - Überschriften Korrektur
- Automatisiertes Hochladen in den Zeitschriftenserver
 - Bände (Jahrgänge, Monate) und Artikel werden angelegt
 - Derivate werden erstellt
- Kontrolle und Metadaten Anreicherung durch Bibliothekare

Vielen Dank für Ihre Aufmerksamkeit!