Automatischer Ingest aus Goobi
Ein Praxiseinblick

Frankfurt, 01. Juni 2017

Was war Goobi nochmal?

- Workflowtool für Digitalisierungsprojekte
- Mittlerweile auch Präsentationstool
- Entwickelt in Java als Open Source
- Verbreitung zunächst vor allem in Deutschland
- Seit 2009 außerhalb Deutschlands in 56 Einrichtungen in 12 Ländern
- Jährliche User Meetings in Göttingen und London

Goobi - Workflowsteuerung

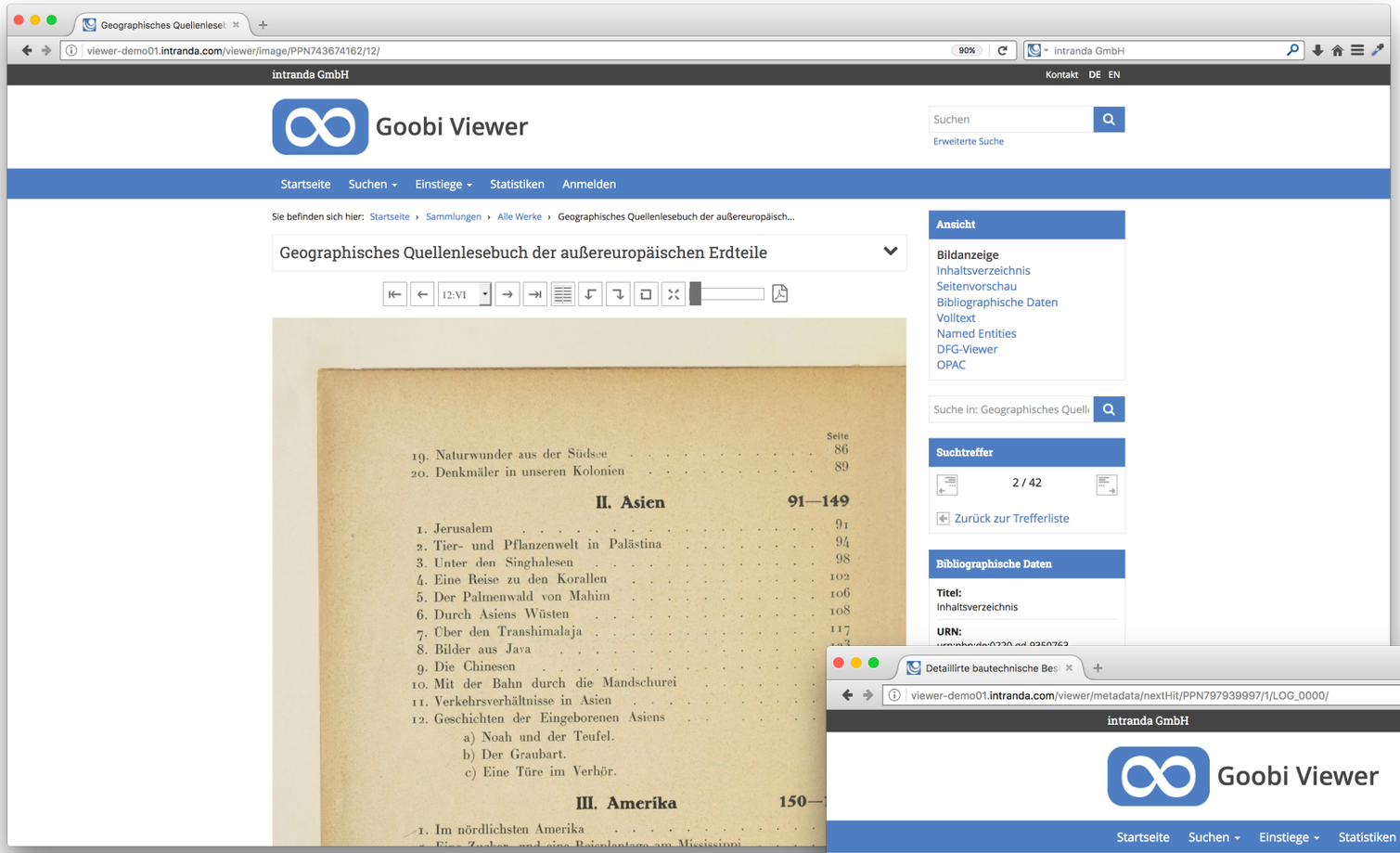
Workflows organisiert in Arbeitsschritten

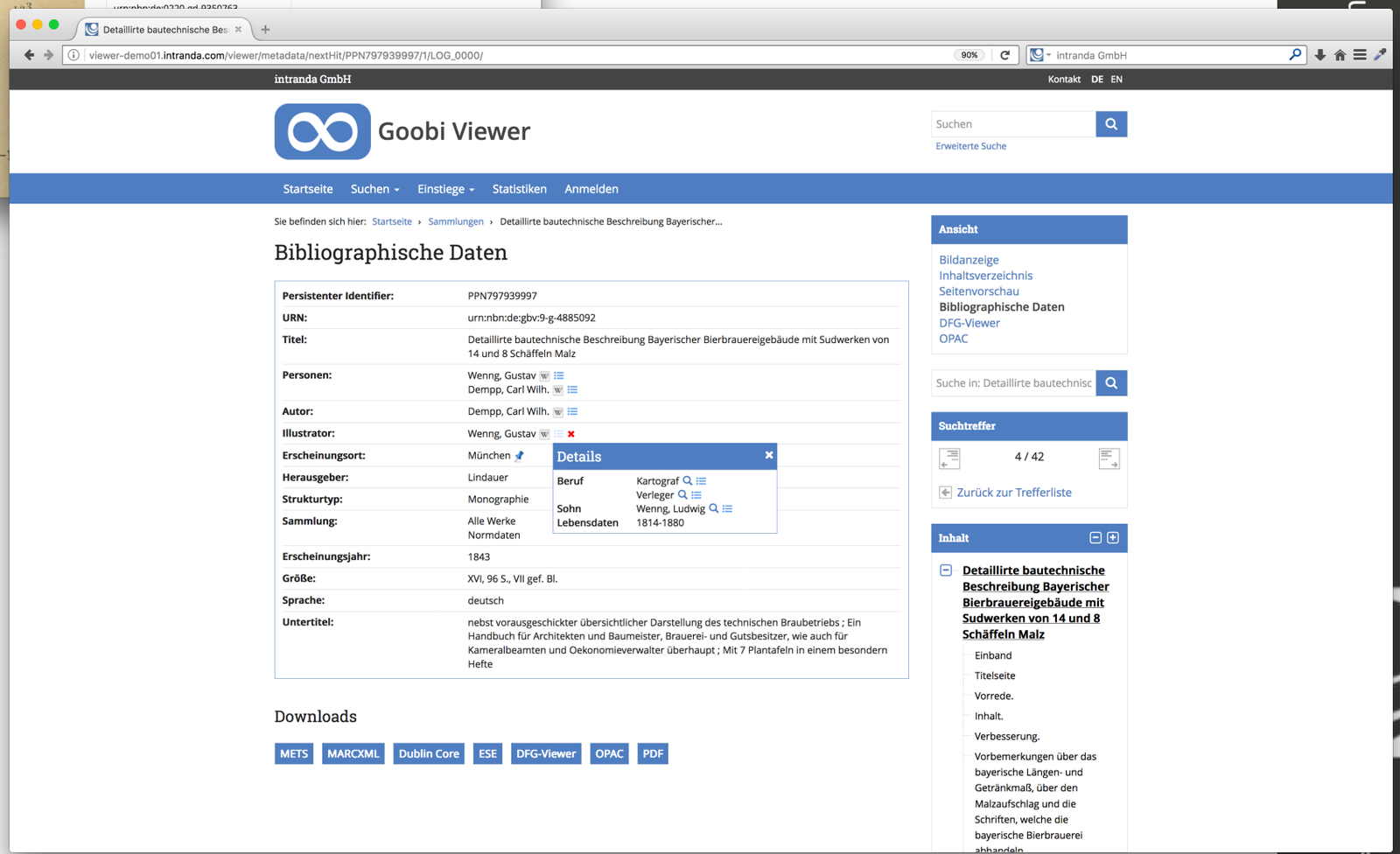
Editor für Paginierung, Strukturdaten und Metadaten

01.06.2017

intranda GmbH

Goobi viewer





IIIF kompatible Bildanzeigen

Metdatenanzeige inkl. Integration von Live-Daten aus der GND

Was war Goobi nochmal?

- Zumeist Retrodigitaliserungsprojekte in Bibliotheken
  - Arbeitsorganisation (Arbeitsschritte)
  - Metadatenerfassung
  - Automatisierung (Scripte etc.)
  - Gehostet oder selbst betrieben
  - Koordinierung der Datenmassen
  - Veröffentlichung der Ergebnisse
  - Datenübernahme aus anderen Systemen
  - Datenübergabe an andere Systeme

Was war Goobi nochmal?

Input / Import

Workflow step 1
Scannen

Workflow step 2
Qualitätskontrolle

Workflow step 3
Bildoptimierung

Workflow step 4
Metadatenerfassung

Workflow step 5
OCR

Workflow step 6
Katalogupdate

Output / Export

Daten kommen aus einer externen Quelle

Arbeitsschritte im Workflow:

- Schritte werden nacheinander ausgeführt
- Schritte werden ausgeführt als Aufgaben
- Aufgaben können manuell oder automatisch sein
- Aufgaben können Tools, Scripts o.ä. aufrufen
- Aufgaben können als Plugins über eigene Nutzeroberfläche verfügen
- Benutzer haben eine Aufgabenliste
- Abhängig von Workflow, Projekt und Nutzergruppe können Aufgaben ausgeführt werden
- Validierungsplugins vermeiden Fehler

Daten werden zu einem anderen System geliefert

Zusammenspiel mit MyCoRe bisher

- Beispiel Thulb Jena
  - im Wesentlichen nur Workflow zur Koordinierung einiger Arbeitsschritte in Goobi
    - Metadatenerfassung innerhalb von MyCoRe

- Beispiel Braunschweig TU
  - Workflow vollständig via Goobi
  - Manueller Upload von PDF-Dateien nach MyCoRe

Zusammenspiel mit MyCore bisher

- Beispiel Rostock UB
  - Vollständige Workflowsteuerung durch Goobi
  - Metadatenerfassung in Goobi
  - Automatischer Datenabzug der internen METS-Datei sowie der Bilder nach MyCoRe über Eigenimplementierung „von außen“
- Kommunikation mit Goobi via Web-API
- Besonderheit außerdem: Bereitstellung der Digitalisate via IIIF in MyCoRe und Anzeige in Landesportal auf Basis des Goobi viewers

Zusammenspiel mit MyCore bisher

Zweyter Theil: Aus mehr als 7

www.digitale-bibliothek-mv.de/viewer/image/83206579X/14/ 90% intranda GmbH

KONTAKT Suchen

Mecklenburg-Vorpommern

Mecklenburg Vorpommern MV tut gut.

DIGITALE BIBLIOTHEK

Über uns | Highlights | Sammlungen | Suchen | Anmelden

Startseite — Sammlungen — Universitätsbibliothek Rostock — Zweyter Theil: Aus mehr als 750. Speisen bestehend...

Ansicht

- Bildanzeige
- Download
- Seitenvorschau
- Bibliographische Daten
- DFG-Viewer
- Besitznachweis
- Objekt beim Datenlieferanten

Suche in: Zweyter Theil: Aus mehr als 750.

Suchtreffer

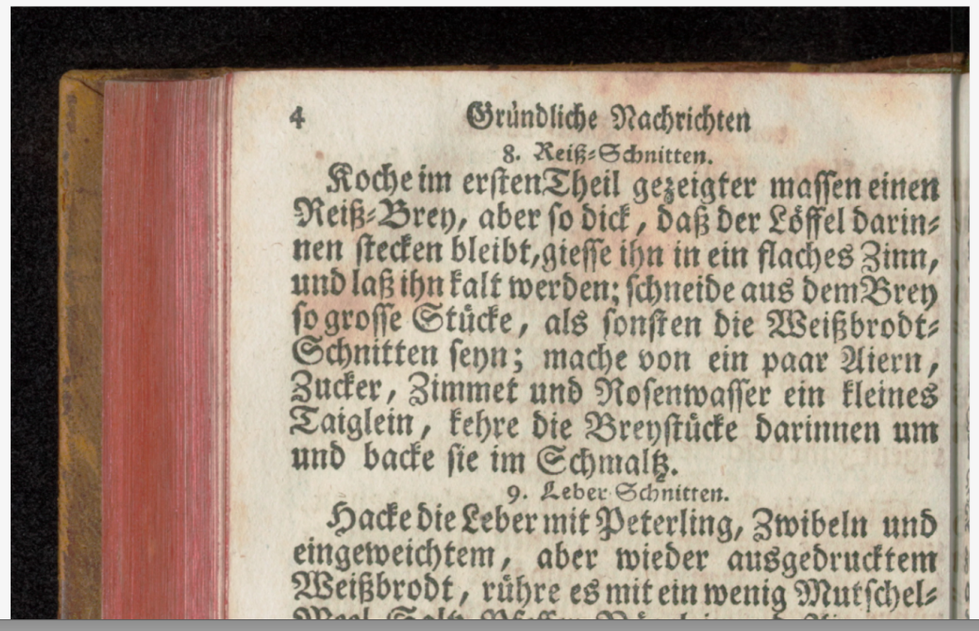
26 / 9.493

Zurück zur Trefferliste

Beschreibung

Zweyter Theil: Aus mehr als 750. Speisen bestehend, Nebst einer kleinen Hauß-Apotheck und andern Dem Frauenzimmer dienlichen raren Kunst-Stücken, Absonderlich Welchen an Säuberung Edelsteine, Perlen, und andern Sachen gelegen

14:4



4 Gründliche Nachrichten

8. Reiß-Schnitten.

Koche im ersten Theil gezeigter massen einen Reiß-Brey, aber so dick, daß der Löffel darinnen stecken bleibt, giesse ihn in ein flaches Zinn, und laß ihn kalt werden; schneide aus dem Brey so grosse Stücke, als sonsten die Weißbrodt-Schnitten seyn; mache von ein paar Aiern, Zucker, Zimmet und Rosenwasser ein kleines Taiglein, kehre die Breystücke darinnen um und backe sie im Schmaltz.

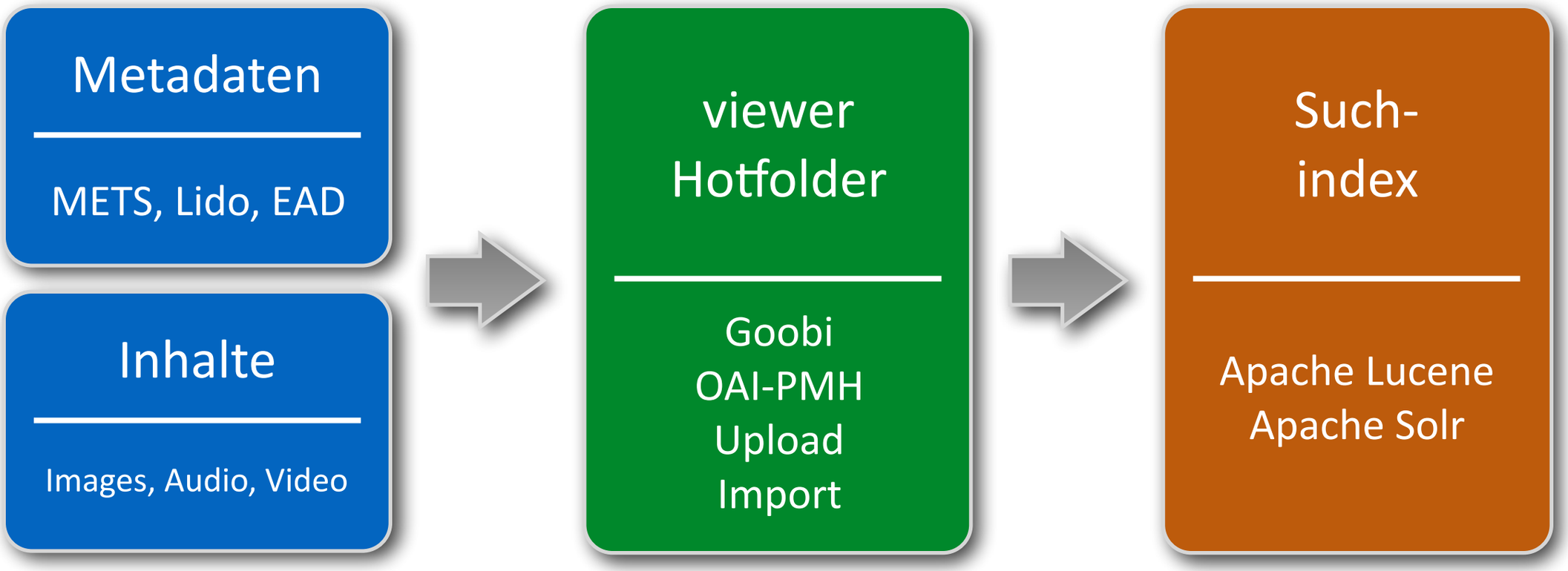
9. Leber-Schnitten.

Hacke die Leber mit Peterling, Zwibeln und eingeweichtem, aber wieder ausgedrucktem Weißbrodt, rühre es mit ein wenig Mutschel-

Verglichen dazu: Die Injests der anderen?

- Bisherige andere Injests in andere Systeme:
  - Goobi viewer
  - Preservica
  - Rosetta
  - Fedora

Ingest in Goobi viewer



Arbeitsweise bisher über das Dateisystem

Ingest in Preservica

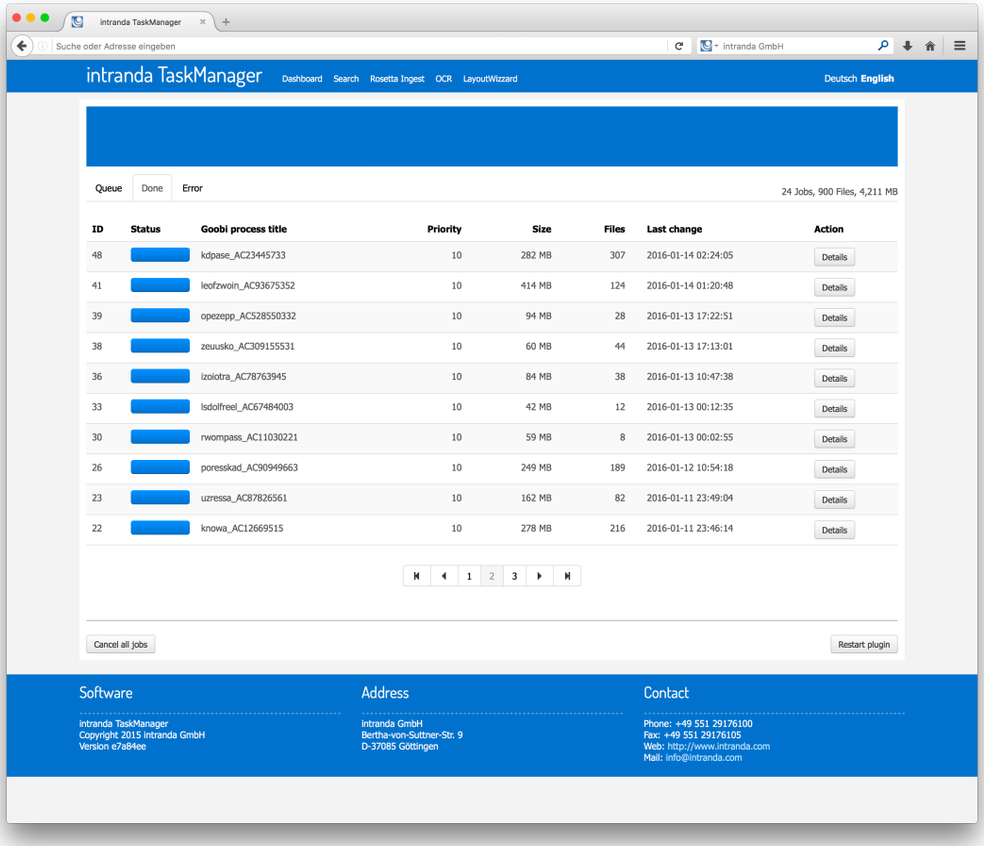
Goobi

- Bibliographische Aufnahme
- Scannen
- […]
- Struktur- und Metadaten
- Derivate (Bilder, OCR, Epub …)
- SOAP Generator
- Preservica Ingest
- Export zu Präsentationssystem

Separater Aufruf des SOAP-Generators und erst im Anschluß Bereitstellung der Dateien in Hotfolder

Aufruf der Goobi-Web-API nach Abschluss des erfolgreichen Ingests, Rückgabe einer xml-Datei mit Hash-Werten und URLs zur Ergänzung der METS-Datei

TaskManager



Bereitstellung der Inhalte ohne Metadaten; Ungewisse Wartezeit bei teilweise ungewissem Ergebnis

Preservica

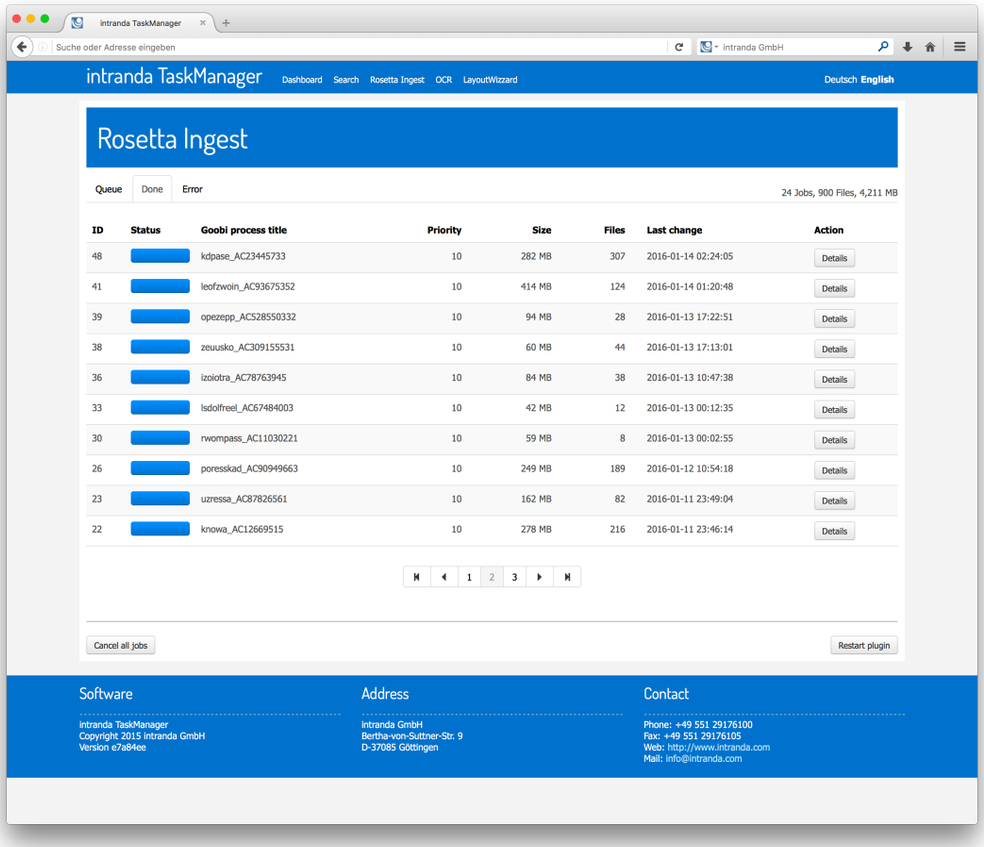
Ingest in Rosetta

Goobi

- Bibliographische Aufnahme
- Scannen
- […]
- Struktur- und Metadaten
- Derivate (Bilder, OCR, Epub …)
- […]
- **Rosetta Ingest**
- Export zu Präsentationssystem

Aufruf des TaskManagers via TaskClient als Commandline Aufruf innerhalb des regulären Workflows

TaskManager



Aufruf der Goobi-Web-API nach Abschluss des erfolgreichen Ingests

Bereitstellung einer aufbereiteten METS-Datei und der zugehörigen Streams für den Ingest in Rosetta

Rosetta

Lessons learned: Der Status quo

- Große Datenmengen
- Lange Transferzeit
- Überwachung des Ingestfortschritts
- Abschließende Statusüberprüfung

- Verläßliche Auskunft, ob die Daten vollständig und intakt angekommen sind

Goobi - 2.2

goobi01.fritz.box/goobi/uii/process_all.xhtml

intranda GmbH

Dashboard | Meine Aufgaben | Workflow | Administration | Controlling | Steffen Hankiewicz

Vorgangsdetails

4 Aktive Benutzer

17:58:58 Sonntag, 11. September 2016

Dashboard › Vorgänge › Vorgangsdetails

Vorgang

Vorgangstitel:	demo_mycore_ocr
Projekt:	Tests
Erstellungsdatum:	16.08.2016
Regelsatz:	Standard
Laufzettel:	Standard
In Auswahlliste anzeigen:	☐
Ist eine Vorlage:	☐
ID:	38666
Batch:	

Vorgangslog

Nachricht hinzufügen

Abfolge der Aufgaben

Nr.	Titel	Status	Aktionen
0	› Vorgang anlegen		
1	› Einspielen der Images		
2	› Tiff-Header erzeugen		
3	› Erstellung der komprimierten Derivate		
4	› Qualitätskontrolle		
5	› Struktur-u. Metadaten		
6	› OCR		
7	› MyCore preparation		
8	› MyCore ingest		

+ Aufgabe hinzufügen

Physische Vorlagen

Autoren:	Knowles, Lees; Winkel, G. G.
Erscheinungsjahr:	1914

Goobi - 2.2

goobi01.fritz.box/goobi/uii/process_edit_step.xhtml

intranda GmbH

Dashboard | Meine Aufgaben | Workflow | Administration | Controlling | Steffen Hankiewicz

Aufgabendetails

4 Aktive Benutzer

18:00:20 Sonntag, 11. September 2016

Dashboard › Vorgänge › Vorgangsdetails › Aufgabendetails

Details - demo_mycore_ocr

Titel *	MyCore preparation
Reihenfolge *	7
Priorität *	0
Metadaten	☐
Images lesen	☐
Images schreiben	☐
Beim Abschließen verifizieren	☐
Export DMS	☑
Aufgabe überspringen	☐
Automatische Aufgabe	☑
Skript Schritt	☐
Status *	Offen
Batch Schritt	☐
Plugin für Arbeitsschritt	plugin_intranda_mycore_export
Validierungsplugin	Nicht ausgewählt
Plugin für Zeitverzögerung	☐
Metadatenindex beim Abschließen aktualisieren	☐

Löschen | Abbrechen | Speichern

Berechtigte Personen

Administration

Stuart Lewis

Suche oder Adresse eingeben

intranda GmbH

mir mycore mods institutional repository

administrator

Suche Blättern Dokumente einreichen 0

Suche

Start / / Stuart Lewis

Ein anderer Tag diesmal ohne Corps-Studen... 6 von 9 IIIF-Image Test

Strahlsundische Zeitung

Vorschau

[1] - -

Strukturübersicht

- Strahlsundische Zeitung
  - Ausgabe
  - Ausgabe
  - Ausgabe
  - Ausgabe

31% - 0°

Dateien Aktionen

00000001.tif	31.08.2016	2.41 MB	
00000002.tif	31.08.2016	2.66 MB	
00000003.tif	31.08.2016	2.78 MB	
00000004.tif	31.08.2016	2.44 MB	

+ auf die Merkliste Aktionen

Zitieren

tweet teilen +1

Zitierform:
Stuart Lewis.

Zitier-Link kopieren

Export

BibTeX, MODS, RIS, ISI, PICA, DC

Systeminformation

Publikationsstatus:	eingereicht
Erstellt am:	19.05.2016 - 16:38:00
Erstellt von:	administrator
Letzte Änderung:	31.08.2016 - 11:39:01
Zuletzt geändert von:	administrator
MyCoRe ID:	mir_mods_00000018
Version:	2 (Gesamte Versionsgeschichte)

Lessons learned: Der Status quo

- Implementiert:
  - Erzeugung einer (einzigen) großen zip-Datei
  - Generierung von Dublin-Core Daten
  - Meldung der Dublin-Core Daten an MyCoRe
  - Hochladen der zip-Datei

- Derzeit noch problematisch für Werke größer als 10 GB:
  - (lange) warten
  - selbst manuell prüfen

Offene Fragen zur Vorgehensweise?

Fragen