



Content Conversion Specialists

Fortschrittsbericht Kitodo & docWorks

Kitodo Mitgliederversammlung 01.06.17 - Frankfurt

Daniel Lanz

CCS GmbH

Wer ist CCS?

- CCS - Content Conversion Specialists – spezialisiert auf Bibliotheks-Digitalisierungs-Software und –Projektumsetzung.
- Gegründet 1976 , 40 Mitarbeiter
- Teilnahme an diversen Forschungsprojekten:
 - SLUB Dresden – Kitodo Anbindung (2016)
 - ENP – Europeana Newspapers Projekt (2012 – 2015)
 - Library of Congress (2004), National Digital Newspaper Program, technische Spezifikation
 - METAe – The Metadata Engine (2000 – 2003)
- Größte Digitalisierungsprojekte
 - Königliche Bibliothek der Niederlande (9 Million Zeitungsseiten, 2008 – 2012)
 - British Library (68.000 Bücher, 2007 – 2008)

Referenzen

Digitalisierungs-Software

15 Nationalbibliotheken

40 Akademische Bibliotheken

10 Digitalisierungs-Dienstleister

Digitalisierungsprojekte

F.A.Z.

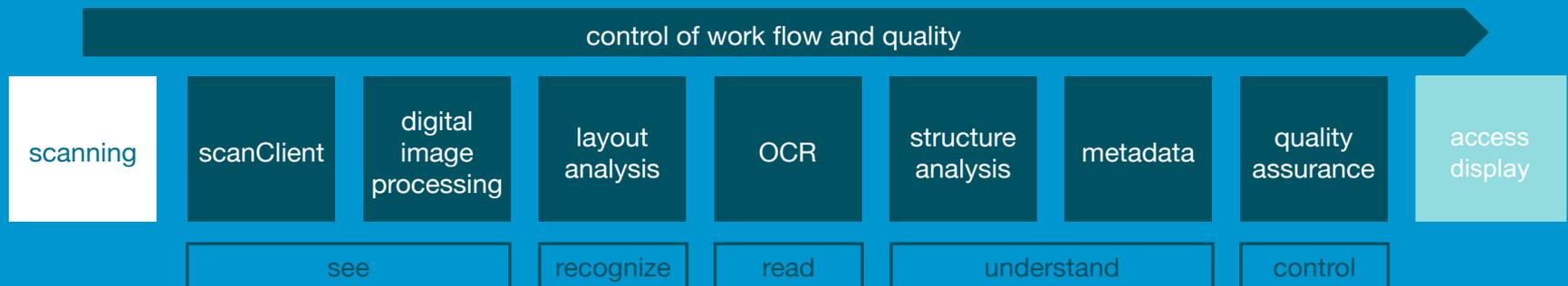
Axel Springer

10 Nationalbibliotheken

Cengage, FindMyPast, ProQuest, Microsoft

Was ist docWorks?

- Digitalisierungs-Software, die Bildbearbeitung, Layout-Analyse, Texterkennung, Struktur-Analyse und automatische Metadaten-Erfassung beinhaltet
- docWorks gibt es als Einzelplatzlösung und als Lösung mit bis zu 500 Arbeitsplätzen



- Je nach Anforderung werden manche Schritte übersprungen

Struktur-Erkennung & Metadaten-Erfassung

Monografie



Struktur-Erkennung & Metadaten-Erfassung

Zeitung

- Mit einem Regelsatz werden Wörter, Zeilen, Spalten, Linien und grafische Elemente erkannt und Elemente klassifiziert



- Die Struktur-Analyse fasst die verschiedenen Elemente eines Artikels zusammen und berücksichtigt dabei auch Fortsetzungen



Kitodo & docWorks – Überblick

- Entwicklungspartnerschaft SLUB Dresden mit CCS in 2016
- Ziel: Einbindung folgender docWorks Funktionalitäten in Kitodo:
 - Layout-Analyse
 - Texterkennung mit Frakturschrift-Erkennung (Tesseract 3.0.2, ab September 2017 Version 3.0.5)
 - Strukturanalyse (Monografien, Zeitschriften und Zeitungen)
 - Automatische Metadaten-Erfassung

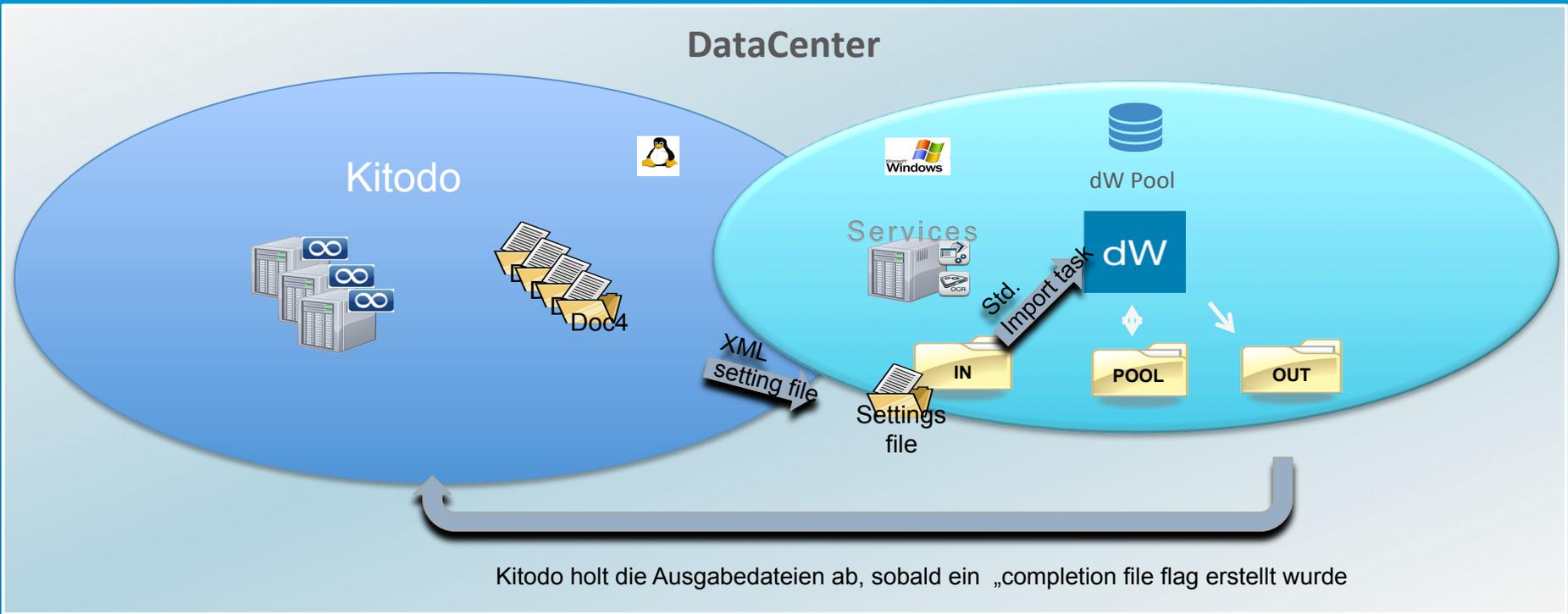
Kitodo & docWorks – Vorteile

- Welche Vorteile bietet der Einsatz der docWorks Features?
 - Bessere Ergebnisse für Zeitungs-Digitalisierung durch genaue Spalten- und Zonenerkennung (besonders wichtig bei Fortsetzungen)
 - Unbegrenzte Frakturschrift-Erkennung ohne Mehrkosten
 - Auswahl weiterer OCR Engines (ABBYY, Nuance)
 - Automatische Erfassung von Seitenzahlen
 - Automatische Struktur-Erfassung (Kapitel, Artikel)
 - Automatische Erfassung von Metadaten (z.B. Bildbeschriftungen)

Kitodo & docWorks – Workflow

- Kitodo übergibt Bilddateien, Kitodo XML und Kommando-XML Datei
- docWorks verarbeitet diese automatisch in einem Hintergrundprozess
- docWorks liefert Layout-, Text- und Strukturinformationen sowie Metadaten im XML Format, das für Kitodo lesbar ist
- Kitodo übernimmt die Ergebnisse automatisch

Kitodo & docWorks – Systemumgebung



Kitodo & docWorks – Nächste Schritte

- Nutzung durch weitere Organisationen ab Mitte 2017 möglich (jährliche Nutzungsgebühr)
- Anpassung durch CCS in die bestehende Kitodo Umgebung (kostenpflichtig)
- Weitere Verbesserung der Implementierung angestrebt
- Kontinuierliche Weiterentwicklung, um neue docWorks Funktionen in Kitodo nutzbar zu machen

Danke für Ihre Zeit!

Daniel Lanz

Geschäftsführer

CCS Content Conversion Specialists GmbH

Weidestr. 134

22083 Hamburg

Germany

T +49 40 227130-16

F +49 40 227130-11

M +49 176 12713016

d.lanz@content-conversion.com

www.content-conversion.com