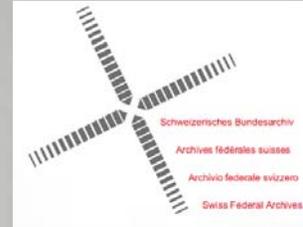


Hauptanforderungen

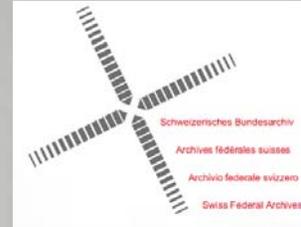


- Elektronische Übernahme der digitalisierten (gescannten) Seiten (TIF)
- OCR* und Erstellung von PDF (1Mio. Seiten)
- OCR* Güte 99.5% für Normalschrift, 85% für Frakturschrift
- Titel- und Textunit Erkennung (100%)
- Schnittstellen zu Umsystemen (Import)
- Skalierbar, ausbaubar
- Sehr gute Suchfunktionen, gute Performance
- Benutzerfreundliche Oberfläche
- Einhaltung Terminplan



* Scannen: Digitalisierung von Papier

* OCR: Zeichenerkennung auf gescannten Dokumenten für Inhaltsanalyse und Volltextsuche



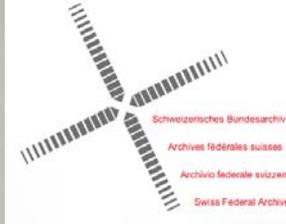
- OCR und Erstellung von PDF (1Mio. Seiten)
- OCR Güte 99.5% für Normalschrift, 85% für Frakturschrift
- Titel- und Textunit Erkennung (100%)

➔ **Projektfokus**

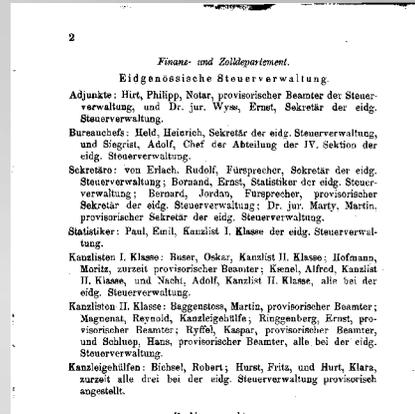
- Metadaten Extraktion
- Schnittstellen zu Umsystemen (Import)
- Skalierbar, ausbaubar
- Sehr gute Suchfunktionen, gute Performance
- Benutzerfreundliche Oberfläche

➔ **Standardprodukt
Convera
RetrievalWare**



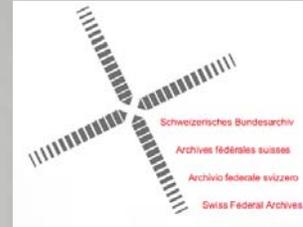


- Die geforderte OCR Güte und die automatische Titel-/Textunterkennung kann nur mit einem semiautomatischen Lösungsansatz erreicht werden.
- Die Qualität der Images ist äusserst unterschiedlich.
-> Iteratives OCR Verfahren, abhängig von der Image Qualität
- OCR Area Erkennung, Noise Remove und weitere Image Enhancements werden bei schlechter OCR Qualität durchgeführt

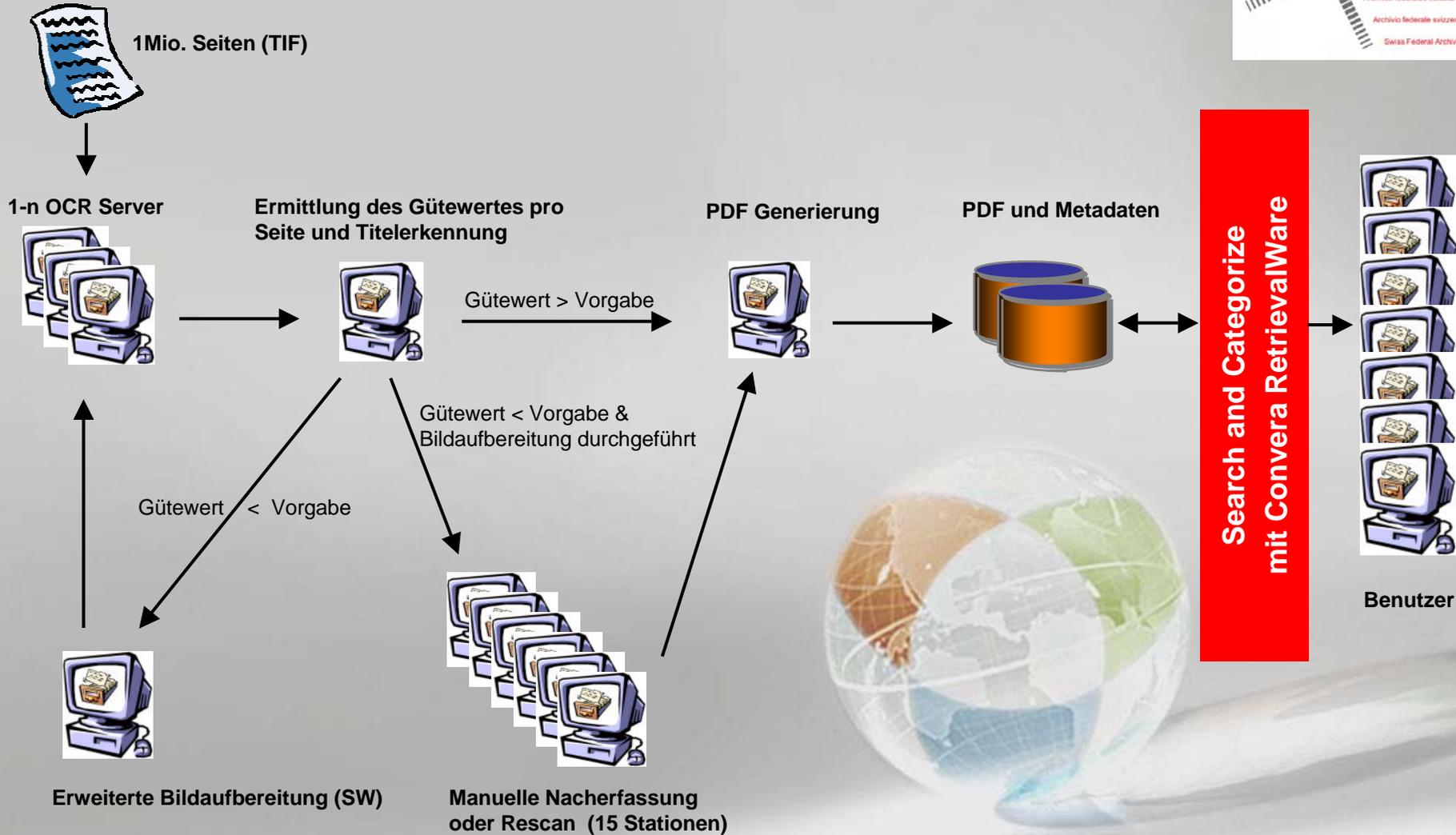


Herausforderungen

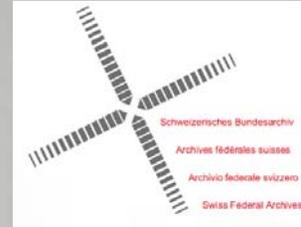
- Manueller Korrekturaufwand gering halten (Kosten)
- Automatische Titelerkennung
- OCR von Frakturschriften



Workflow (vereinfachte Darstellung)



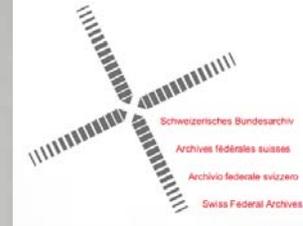
Folgerung



- Das Konzept der semiautomatischen Prozesse passte
- Dank einem Team, bestehend aus 15-20 motivierten Mitarbeitern (temporär über 2 Monate) konnten die Termine eingehalten werden.
- Aufgrund der schlechten Scannqualität wurden ca. 400'000 Seiten rescannt > Scann Mehraufwand, dafür geringerer Korrekturaufwand

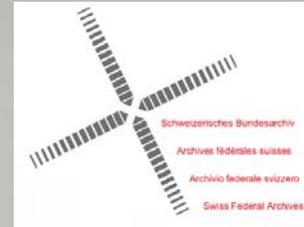


Ausblick aus Sicht Lieferanten



- Crosslinguale Suche
- Einbindung von semantischen Netzwerken
- Automatische Kategorisierung und Klassifizierung





Vielen Dank für Ihre Aufmerksamkeit!

Bruno Eugster

Convera AG

eugster@convera.ch

