

---

## e-rara.ch: Volltext 152-005

e-rara.ch: Nachnutzung für die Wissenschaft – Der Weg zum Volltext

Die Forschung im Feld der Digital Humanities, aber auch darüber hinaus, hat sich in den letzten Jahren rapide weiterentwickelt. Neue Technologien ermöglichen es, künftig auch den Volltext (e-rara.ch) bzw. Tools für das Erstellen des Volltexts und von Annotationen (e-manuscripta.ch) bereitzustellen. Dies ist der konsequente nächste Schritt in der Weiterentwicklung der Plattformen. Mit der Schaffung der Voraussetzungen für Volltext bzw. Annotation für die Plattformen e-rara.ch und e-manuscripta.ch wird aus Sicht der Wissenschaft und der Kooperationspartner von e-rara.ch und e-manuscripta.ch ein zentrales nationales und internationales Desiderat adressiert. Die Projektergebnisse werden die Forschung in der Schweiz und die Forschung über die Schweiz im In- und Ausland massgeblich unterstützen und verändern.

Der Antrag „e-rara.ch: Nachnutzung für die Wissenschaft – Der Weg zum Volltext“ verfolgt folgende Teilziele:

1. Einführung der Texterkennung (Optical Character Recognition OCR) für die in e-rara.ch vorliegenden qualitativ hochwertigen Image-Digitalisate von gemeinfreien Alten Drucken. Der inhaltliche Schwerpunkt liegt aus technischen Gründen auf Drucken mit Antiqua-Schrift aus dem Erscheinungszeitraum von ca. 1830 bis 1920.
2. Erweiterung der Webpräsentation (Frontend) für die Recherche nach und die Anzeige der Rechercheresultate von volltexterkannten Drucken.
3. Einbindung von im Rahmen von eBooks on Demand erstellten OCR-Dateien zu Drucken, die noch nicht in e-rara.ch vorhanden sind.
4. Prüfung und ggf. Entwicklung von technischen Workflows für rekursive OCR-Bearbeitungen – sprich: Die Wiederholung der OCR-Bearbeitung bzw. der Abgleich mit Ergebnissen von OCR-Bearbeitungen an anderen Orten mit dem Ziel der Verbesserung der Genauigkeitsquote.
5. Anpassung der bestehenden technischen Infrastruktur für e-rara.ch an der ETH-Bibliothek. Für die rechnerintensive Erstellung, Verarbeitung und Speicherung der Volltexte ist einerseits eine Erweiterung des virtuellen Servers und der Speicherkapazität notwendig. Andererseits muss der Visual Library Manager (mandantenfähiges Erschliessungssystem) zur Steuerung der (externen) OCR-Engine erweitert werden (Backend).
6. Definition der zusätzlichen organisatorischen Workflows. Mit der Einführung der Texterkennung erweitert sich die Palette der betrieblichen Prozesse, die analysiert, festgehalten und schlussendlich betrieben werden.

Der Antrag zur Weiterentwicklung von e-manuscripta.ch wird als Folgeantrag in der kommenden Förderperiode von SUK P2 eingereicht werden.