

Archivalische Bestände in Bibliotheken und universitären Sammlungen: Das Digitalisierungsprojekt „Sammlung Zeitgeschichte – Online“

Emily Herkommer, Eva Hofmann

Zusammenfassung: Der vorliegende Projektbericht dokumentiert die Umsetzung des Digitalisierungsprojekts „Sammlung Zeitgeschichte – Online“, dessen Ziel es war, archivalische Bestände aus dem Institut für Zeitgeschichte und der Fachbereichsbibliothek Zeitgeschichte der Universität Wien digital zu erschließen und zugänglich zu machen. Im Zentrum standen die digitale Verfügbarkeit der Findbehelfe des Dokumentenarchivs sowie die Migration und Verbesserung des bereits online veröffentlichten „Gaupresse“-Archivs. Der technische Workflow umfasste die OCR-gestützte Texterkennung mittels *Transkribus*, die Standardisierung der Metadaten nach dem RNAB-Regelwerk sowie die Implementation der Open-Source-Software *Goobi*. Die Zusammenarbeit mit den PHAIDRA-Services der Universitätsbibliothek Wien ermöglichte eine nachhaltige und institutionsinterne Lösung für die digitale Präsentation und Langzeitarchivierung. Das Projekt verdeutlicht die Möglichkeiten, die sich aus der Nutzung von an der Universität Wien bereits vorhandenen Ressourcen ergeben und dient als Modell für ähnliche Digitalisierungsvorhaben in Bibliotheken, Archiven und universitären Sammlungen.

Schlagwörter: Archiv, Digitalisierung, Universitäre Sammlungen, Sammlungsdaten

Archival holdings in libraries and university collections: The digitization project “Sammlung Zeitgeschichte – Online”

Abstract: This project report outlines the implementation of the digitization initiative “Sammlung Zeitgeschichte – Online” at the Department for Contemporary History at the University Vienna. The project aimed to make archival holdings from the Institute of Contemporary History and the Special Library Contemporary History at the University of Vienna digitally accessible. Central objectives included the online availability of the archive’s finding aids and the migration and enhancement of the “Gaupresse”-Archive already published online. The technical workflow involved OCR-based text recognition using *Transkribus*, metadata

standardization according to RNAB guidelines, and the integration of the open-source platform *Goobi*. Collaboration with the PHAIDRA-Services of the Vienna University Library enabled a sustainable, institution-wide supported solution for online presentation and long-term archiving. Strong cooperation between various academic units was crucial to the project's success, which now serves as a reference model for similar digitization efforts in academic libraries, archives, and university collections.

Keywords: Archives, digitization, university collections, collection data

DOI: <https://doi.org/10.31263/voebm.v78i1.9662>

Dieses Werk ist – exkl. einzelner Logos und Abbildungen – lizenziert unter einer Creative-Commons-Lizenz Namensnennung 4.0 International-Lizenz

1. Ausgangssituation und Projektdesign

Die Digitalisierung und digitale Bereitstellung von Sammlungsbeständen an der Universität Wien wurde in den letzten Jahren stetig vorangetrieben. So auch am Institut für Zeitgeschichte und der Fachbereichsbibliothek Zeitgeschichte, wo im Rahmen eines Digitalisierungsprojekts die Plattform „Sammlungen Zeitgeschichte“ entstanden ist, um historische Bestände digital zugänglich zu machen. Das Digitalisierungsprojekt, das zwischen April 2023 und Dezember 2024 mit universitären Mitteln durchgeführt wurde und in dem verschiedene universitäre Abteilungen zusammenarbeiteten, zielte darauf ab, zwei zentrale Herausforderungen zu lösen, die in den Sammlungen des Instituts für Zeitgeschichte und der Fachbereichsbibliothek Zeitgeschichte aufgetreten sind: Zum einen sollten die Findbehelfe des Dokumentenarchivs am Institut für Zeitgeschichte nicht nur analog vor Ort, sondern auch digital auffindbar gemacht werden, um so die langfristige Nutzbarkeit der Bestände zu sichern und die Sichtbarkeit des Archivs zu erhöhen. Zum anderen sollte eine Lösung gefunden werden, um das sogenannte, bereits veröffentlichte „Gaupresse“-Archiv weiterhin digital zur Verfügung stellen zu können, da die *Drupal*-Version der bisherigen Datenbank nicht weiter serviciert wird und damit eine Migration der Datenbank in jedem Fall vorzunehmen gewesen wäre.

Für die Durchführung des Projekts wurden das Dokumentenarchiv am Institut für Zeitgeschichte und das „Gaupresse“-Archiv zum Teil getrennt bearbeitet, da die Ausgangslagen und die Anforderungen in den beiden Sammlungen unterschiedlich waren. Beim „Gaupresse“-Archiv handelt es sich um ca. 200.000 Zeitungsausschnitte, die von der NS-Gauleitung in Wien nach propagandistischen Auswahlkriterien gesammelt und geordnet worden waren. Die gesamte Sammlung war bereits seit 2018 online einsehbar und dementsprechend waren alle Zeitungsausschnitte zu Projektbeginn bereits gescannt, sortiert und teilweise (manuell) mit Metadaten versehen. Da die Highlighting-Funktion für Suchergebnisse auf der ursprünglichen Website nicht gegeben war und sich die Volltexterkennung in den letzten Jahren stark verbessert hat, wurde eine erneute OCR (Optical Character Recognition) durchgeführt. Außerdem sollte versucht werden, mit dem neuen Volltext eine automatisierte Anreicherung der Metadaten zu erreichen und so weitere Sucheinstiege zu ermöglichen.

Das Dokumentenarchiv ist Teil des Archivs am Institut für Zeitgeschichte, einer universitären Sammlung, die seit den 1960er Jahren laufend erweitert wird. Es handelt sich um sehr heterogene Bestände mit einem Schwerpunkt auf österreichischer Geschichte. Die im Rahmen dieses Projekts bearbeiteten Findbehelfe beinhalten die Metadateninformationen der ca. 1.500 Archivbestände. Für die Online-Präsentation sollten diese Metadaten normdatenbasiert nach dem RNAB-Regelwerk strukturiert und mit Schlagworten versehen werden, um die langfristige Erhaltung des Archivs und die Auffindbarkeit der Bestände zu ermöglichen. Die bereits vor Projektbeginn gescannten Findbehelfe wurden zudem OCR-erfasst, um eine Volltextsuche zu ermöglichen.

2. Durchführung

Die Volltexterkennung des „Gaupresse“-Archivs und der Findbehelfe des Dokumentenarchivs wurde mit der Software *Transkribus* durchgeführt. Da für Druckschriften bereits zahlreiche, mit großen Datensätzen trainierte Machine-Learning-Modelle zur Verfügung stehen und die Testergebnisse zufriedenstellend waren, wurde kein eigenes Modell erstellt. Die Transkription der ca. 200.000 Seiten nahm inklusive Upload der Scans mehrere Monate in Anspruch.

Ausgehend von diesen Volltexten wurde versucht, die Metadaten jener Zeitungsartikel des „Gaupresse“-Archivs, die noch nicht manuell vom Personal der Fachbereichsbibliothek Zeitgeschichte erfasst worden waren, automatisiert zu ergänzen. Dies war möglich, weil oberhalb der affichierten Zeitungsausschnitte Erscheinungsdatum, Ort und Zeitungsname gestempelt sind. Diese Informationen befinden sich in den meisten Fällen in der ersten Zeile des Volltexts und können automatisiert ausgelesen werden. Bereinigt wurden die Metadaten durch einen Abgleich mit einer Liste der im „Gaupresse“-Archiv vorhandenen Zeitungen und einem standardisierten Datumsformat. Dadurch konnten bei einem Großteil der Zeitungsausschnitte die Metadaten automatisiert ergänzt werden. In der Datenbank wird darauf hingewiesen, dass manche Metadaten automatisiert erfasst wurden, denn eine intellektuelle Kontrolle ist aufgrund der großen Datenmenge nicht möglich.

Die Findbehelfe des Dokumentenarchivs mussten aufgrund ihrer heterogenen Layouts und unterschiedlichen Erschließungstiefe manuell erfasst und strukturiert werden. Dieser aufwendige Prozess, in dem bei fehlenden Informationen die physischen Archivalien begutachtet, ein eigener Thesaurus erstellt und Schlagworte vergeben wurden, ermöglichte eine genaue Bestandsaufnahme des Archivs und erleichtert in Zukunft sowohl die Arbeit der Mitarbeiter:innen als auch die Recherchen der Benutzer:innen. Im Rahmen dieses Projekts wurde ein Teil der Bestände neu erfasst und ein Workflow für die kontinuierliche Be- schlagwortung erarbeitet. Obwohl es im Rahmen dieses Projekts nicht möglich war, die Metadaten einzelner Archivalien in der Datenbank zu erfassen, können diese Informationen mit der Volltextsuche gefunden werden.

Für die Präsentation der Digitalisate und Volltexte sowie für die Darstellung der Bestände in ihrer hierarchischen Ordnung (Archivplan) wurde die Software *Goobi* an die spezifischen Anforderungen angepasst und als Datenbank „Sammlung Zeitgeschichte“ implementiert. Da das „Gaupresse“-Archiv nur für Forschung und Lehre genutzt werden soll, wurde hier ein Anmeldetool eingesetzt, das eine Registrierung der Nutzer:innen ermöglicht. Die Langzeitarchivierung der Bestände erfolgt in PHAIDRA, dem digitalen Repositorium der Universität Wien.

3. Ergebnisse und Ausblick

Die neue Datenbank „Sammlung Zeitgeschichte“ markiert einen wesentlichen Fortschritt in der digitalen Erschließung der universitären Sammlungen am Institut für Zeitgeschichte. Für das Dokumentenarchiv wurde die Durchsuchbarkeit der Findbehelfe durch Volltexterkennung, Schlagworte und standardisierte Metadaten deutlich verbessert. Die automatisierte Anreicherung der Metadaten von Zeitungsausschnitten im „Gaupresse“-Archiv konnte auf Grundlage klar identifizierbarer Textmuster erfolgreich umgesetzt und das OCR erneuert und um die Highlighting-Funktion der Suchergebnisse ergänzt werden. Die Integration in die Goobi-Infrastruktur erlaubt die Darstellung komplexer Archivstrukturen einschließlich hierarchischer Ordnungen und verknüpfter Digitalisate.

Ausschlaggebend für die erfolgreiche Umsetzung des Projekts war die enge Zusammenarbeit verschiedener universitärer Abteilungen, eingebettet in ein prozessorientiertes Projektdesign. Anstatt von Beginn an alle Methoden festzulegen, wurden diese im Laufe des Projekts schrittweise entwickelt und laufend an die praktischen Anforderungen angepasst. In der Planungsphase erfolgte eine umfassende Erhebung der historischen Bestände und der zugehörigen Daten. Diese Bestandsaufnahme war entscheidend, um den Aufwand für vorbereitende Arbeiten – wie das Ordnen der Materialien, das Scannen sowie die Erfassung von Metadaten – realistisch einschätzen und Projektziele klar formulieren zu können. Ein zentraler Aspekt des Projekts ist die institutionelle Einbettung der Lösung in die bestehende IT-, Bibliotheks- und Archivinfrastruktur der Universität Wien. Durch die PHAIDRA-Services kann sichergestellt werden, dass die digitalen Bestände nicht nur nachhaltig archiviert, sondern die Infrastrukturen auch langfristig gewartet werden, zugänglich bleiben und für andere archivalische Bestände genutzt werden können.

Die enge interdisziplinäre Zusammenarbeit zwischen dem Institut für Zeitgeschichte und der Fachbereichsbibliothek Zeitgeschichte, die die historischen Bestände bewahren und zur Verfügung stellen, mit den PHAIDRA-Services, die an der Universitätsbibliothek Wien und dem Zentralen Informatikdienst (IT-Support for Research, Data Management) angesiedelt sind, sowie mit Beteiligten aus verschiedenen Ab-

teilungen der Universitätsbibliothek Wien, dem Universitätsarchiv und der Historisch-Kulturwissenschaftlichen Fakultät hat sich als äußerst produktiv erwiesen und stellt eine zentrale Lehre für künftige Vorhaben dar.

Der gesamte Projektbericht ist im PHAIDRA-Repositorium der Universität Wien öffentlich abrufbar: <https://phaidra.univie.ac.at/o:2156962>.

Emily Herkommer BA
emily.herkommer@univie.ac.at

Eva Hofmann BA
e.hofmann@univie.ac.at