

Historische Schulbücher als Spielräume für Digital Humanities? Mapping von unterschiedlichen Metadatenformaten für Bibliotheken und linguistische Analysen

De Luca, Ernesto William

deluca@leibniz-gei.de

Georg-Eckert-Institut - Leibniz-Institut für internationale
Schulbuchforschung, Deutschland

Fallucchi, Francesca

fallucchi@leibniz-gei.de

Georg-Eckert-Institut - Leibniz-Institut für internationale
Schulbuchforschung, Deutschland

Hertling, Anke

hertling@leibniz-gei.de

Georg-Eckert-Institut - Leibniz-Institut für internationale
Schulbuchforschung, Deutschland

Klaes, Jan Sebastian

klaes@leibniz-gei.de

Georg-Eckert-Institut - Leibniz-Institut für internationale
Schulbuchforschung, Deutschland

Schmitz, Claudia

schmitz@leibniz-gei.de

Georg-Eckert-Institut - Leibniz-Institut für internationale
Schulbuchforschung, Deutschland

Towara, Nadine

towara@leibniz-gei.de

Georg-Eckert-Institut - Leibniz-Institut für internationale
Schulbuchforschung, Deutschland

Schulbücher transportieren gesellschaftliche und staatlich sanktionierte Werte und Normen. Als Quellengattung stellen sie einen vielversprechenden Gegenstand für zahlreiche wissenschaftliche Fragestellungen dar. Schulbücher werden in zahlreichen Bibliotheken gesammelt, aber in den seltensten Fällen werden sie systematisch erschlossen und für die Digitalisierung genießen sie in der Regel keine hohe Priorität.

Im Rahmen der digitalen Schulbuchbibliothek GEI-Digital¹ wurden in den letzten 10 Jahren historische deutsche Schulbücher der Fächer Geschichte, Geographie und Politik, sowie Realien- und (Erst-)Lesebücher von den Anfängen der Schulbuchproduktion im 17. Jahrhundert bis zum Ende des Ersten Weltkriegs digital zugänglich gemacht (Hertling/Klaes 2018). Digitalisiert und integriert wurden dabei sowohl Schulbücher aus den Beständen der Forschungsbibliothek des Georg-Eckert-Instituts – Leibniz-Institut für internationale Schulbuchforschung (GEI) als auch Schulbücher aus zahlreichen Partner-Bibliotheken im deutschsprachigen Raum. Die externen Schulbuchbestände wurde dem GEI zum Zwecke der Digitalisierung im Rahmen von Kooperationen leihweise überlassen oder als Fremddigitalisate virtuell in die GEI-Digital-Sammlung integriert.

Die bibliothekarische Erschließung des historischen Schulbuch-Korpus folgt den spezifischen Bedürfnissen der Schulbuchforschung. Typisches Kennzeichen von Schulbüchern sind dabei viele Ausgaben und die unterschiedlichen Bände eines Schulbuchs. Neben Angaben zum Verlag und Erscheinungsjahr werden zusätzlich auch Schulfächer und Schulstufen als deskriptive Metadaten erfasst. Sie stehen auf GEI-Digital als MARCXML, Metadata Object Description Schema (MODS) oder Dublin Core (DC) zur Nachnutzung zur Verfügung. Die Erschließung umfasst auch die intellektuelle Verknüpfung der Schulbuchautoren mit Normdateneinträgen in der Gemeinsamen Normdatei (GND), um das Korpus für biographische Forschungsansätze zu öffnen. Darüber hinaus werden die digitalisierten Schulbücher in Form einer Tiefererschließung in ihrer Struktur erschlossen und die Elemente, wie Titelblätter, Inhaltsverzeichnisse, Abbildungen etc. im Metadata Encoding & Transmission Standard (METS) ausgewiesen.

Die über 1,5 Millionen in GEI-Digital gescannten Seiten wurden zudem einer Optical Character Recognition (OCR) unterzogen und stehen als durchsuchbare Volltexte über eine OAI-PMH-Schnittstelle zur Verfügung. Die Resultate der Volltexterkennung werden im Zuge des Digitalisierungsworkflows im XML-Schema Analyzed Layout and Text Object (ALTO) ausgegeben.

Mit GEI-Digital ist für die Digital Humanities ein einzigartiges Korpus mit über 6.100 digitalisierten Schulbüchern entstanden, dass die gesamte Epoche der deutschen Schulbücher von deren Entstehung bis 1918 mit hoher Vollständigkeit virtuell zusammenführt. Die Digitalisate und Daten werden in zahlreichen Digital-Humanities-Projekten bereits nachgenutzt, z.B. im Projekt „Welt der Kinder“² (Heuwing/Weiß 2018 und Nieländer/Weiß 2018), in dem das Korpus mit Topic Modeling-Verfahren untersucht wurde. Das Portal GeoPortOst³ nutzt u.a. das in GEI-Digital vorhandene Kartenmaterial für Georeferenzierungen.

In einem nächsten Schritt ist geplant, das Korpus in dem von CLARIN betriebenen Virtual Language Observatory

(VLO)⁴nachzuweisen, um es für weiterführende und v.a. linguistische Analysen zugänglich zu machen. Voraussetzung für einen Nachweis ist die Repräsentation der digitalisierten Schulbücher in der Component MetaData Infrastructure (CMDI). CMDI stellt ein Framework zur Verfügung, um Profile für Metadaten für die Beschreibung und Benutzung bereitzustellen.

Ausgehend von Metadatenformaten, die sich v.a. an bibliothekarischen Standards orientieren, werden auf dem Poster Anforderungen und Strategien für Mapping-Prozesse als Grundlage für Digital-Humanities-Projekte präsentiert. Im Mittelpunkt stehen dabei die in GEI-Digital gemachten Mapping-Erfahrungen mit den Formaten METS/MODS, TEI, CMDI und Dublin Core (DC) und die Herausforderung ihrer jeweiligen Interoperabilität.

In einem ersten Schritt In einer Machbarkeitsstudie stellte sich im Projekt GEI-Digital ein Mapping von METS zu CMDI als undurchführbar heraus (Fallucchi/De Luca 2019). Ein Mapping von Dublin Core (DC) zu CMDI als CLARIN-Empfehlung ist mit Blick auf die Besonderheiten der Erschließung von Schulbüchern insbesondere mit Blick auf die für die Forschung wichtigen Ausgabebezeichnungen und Bandangaben stark verlustbehaftet. Vor dem Hintergrund werden derzeit alternative Optionen diskutiert, die auf dem Poster aufgezeigt und erörtert werden sollen. Eine Möglichkeit stellt die Anreicherung von Dublin Core-Metadaten und ihre Konvertierung in CMDI dar. Eine weitere Option besteht in der Umwandlung von textbasierten ALTO-Dateien in CMDI.

Schulbuchforschung - Forschungsinfrastrukturen und Projekte. Göttingen: V&R unipress 21–44

Hertling, Anke / Klaes, Sebastian (2018): »GEI-Digital« als Grundlage für Digital-Humanities-Projekte: Erschließung und Datenaufbereitung, in: Maret Nieländer / Ernesto William De Luca (eds): Digital Humanities in der internationalen Schulbuchforschung - Forschungsinfrastrukturen und Projekte. Göttingen: V&R unipress 45–68

Heuwing, Ben / Weiß, Andreas (2018): Suche und Analyse in großen Textsammlungen: Neue Werkzeuge für die Schulbuchforschung in: Maret Nieländer / Ernesto William De Luca (eds): Digital Humanities in der internationalen Schulbuchforschung - Forschungsinfrastrukturen und Projekte. Göttingen: V&R unipress 145–170

Nieländer, Maret / Weiß, Andreas (2018): »Schönere Daten« – Nachnutzung und Aufbereitung für die Verwendung in Digital-Humanities-Projekten, in: Maret Nieländer / Ernesto William De Luca (eds): Digital Humanities in der internationalen Schulbuchforschung - Forschungsinfrastrukturen und Projekte. Göttingen: V&R unipress 91–116

Fußnoten

1. Projektwebseite GEI-Digital – die digitale Schulbuch-Bibliothek:: <http://gei-digital.gei.de/viewer/>
2. Projektwebseite Welt der Kinder: <http://welt-der-kinder.gei.de/>
3. Projektwebseite *GeoPortOst*: Portal für thematische und versteckte Karten zu Ost- und Südosteuropa: <http://geoportost.ios-regensburg.de/>
4. CLARIN Virtual Language Observatory: <https://vlo.clarin.eu/?jsessionid=6E94D6CA11D6A8A1889F2C3E6C6A79FE?0>

Bibliographie

Fallucchi, Francesca / De Luca, Ernesto William (2019): "Connecting and Mapping LOD and CMDI Through Knowledge Organization" Springer, Cham, pp. 291–301.

Hertling, Anke / Klaes, Sebastian (2018): Historische Schulbücher als digitales Korpus für die Forschung: Auswahl und Aufbau einer digitalen Schulbuchbibliothek, in: Maret Nieländer / Ernesto William De Luca (eds): Digital Humanities in der internationalen