

Digitale Editionen als Forschungsservice

Pilotprojekt der ZB Zürich zu neuen Formen der digitalen Bestandspräsentation

1. Einstieg in die digitale Editorik

Die Zentralbibliothek Zürich (ZB) verfügt seit über zwei Jahrzehnten über ein eigenes, leistungsstarkes Digitalisierungszentrum, in dem sie kontinuierlich die eigenen Bestände digitalisiert. Ebenso lange produziert sie für wissenschaftliche Editionsprojekte Digitalisate und erhebt dazugehörige und allenfalls weitere projektspezifische Metadaten. So hat sie z.B. innerhalb von sechs Jahren für die digitale Briefedition Johann Caspar Lavater über 23.000 Briefe von und an Lavater erschlossen und digitalisiert.¹ Für andere Editionsprojekte hat sie geringere Mengen beigegeben, etwa für die digitale Briefedition von Conrad Ferdinand Meyer, Heinrich Bullinger oder das Projekt zu den Frühneuzeitlichen Ärztebriefen des deutschsprachigen Raums (1500–1700).² Seit Anfang 2024 realisiert die ZB nun erstmals selbst zwei digitale Editionen und ist dabei, Kompetenzen, Workflows und technische Lösungen im Bereich der digitalen Editorik zu entwickeln und aufzubauen. Damit schlägt die ZB einen für sie neuen Weg ein und probiert ein neues Format aus, die eigenen Bestände für Forschung und Wissenschaft sowie die allgemeine breite Öffentlichkeit zu präsentieren.³

Die ZB testet damit zugleich einen Forschungsservice, mit dem sie ihr Serviceportfolio innovativ ausbauen möchte. In beiden Projekten sollen nachhaltige und interoperable Workflows entwickelt werden, die zugleich bibliothekarischen Standards und Anforderungen der *Digital Humanities* gerecht werden. Ziel ist es, projektspezifische und zugleich allgemeine Workflows und Datenmodelle zu etablieren, die zukünftig leicht nachnutzbar sind.

In beiden Vorhaben stehen Verfahren der *Automatic Text Recognition* (ATR) im Vordergrund: Im ersten Teilprojekt «Digitale Edition der Korrespondenzen von Hans Georg und Hermann Nägeli» werden alle vom Zürcher Musikverleger Hans Georg Nägeli (1773–1836) und seinem Sohn Hermann (1811–1872) verfassten und an sie gerichteten Briefe ediert, die in den Beständen der ZB überliefert sind. Es handelt sich um insgesamt ca. 3.800 Briefe, die im Zeitraum von 1790 bis 1872 entstanden sind. Mit Hilfe der Software Transkribus wird von allen Briefen jeweils ein Rohtext erzeugt, der daran anschließend korrigiert, gemäß TEI/XML-Standards codiert und mit einem Werk-, Orts- und Personenregister verknüpft wird. Mit Hilfe des TEI-Publisher werden die Briefe schließlich auf sources-online.org publiziert.⁴

1 Siehe Johann Caspar Lavater Online Briefedition, 15.07.2025, <https://jclavater-briefwechsel.ch/home>, Stand: 11.08.2025. Zum Lavater-Projekt vgl. auch Reiling, Jesko: Bibliotheken und ihr Beitrag zum Aufbau von Forschungsinfrastrukturen, in: Schweizerische Zeitschrift für die Erforschung des 18. Jahrhunderts xviii.ch 11, 2020, S. 52–62, <https://doi.org/10.24894/2673-4419.00005>, Stand: 11.08.2025.

2 Siehe Lukas, Wolfgang; Landshuter, Stephan: C. F. Meyer: Digitale Briefedition, 2024, <https://cfmeyer-briefwechsel.org/home>, Stand: 11.08.2025; Institut für Computerlinguistik und Institut für Schweizerische Reformationgeschichte der Universität Zürich: BullingerDigital, 2020–2025, <https://www.bullinger-digital.ch/index.html>, Stand: 11.08.2025; Stolberg, Michael: Frühneuzeitliche Ärztebriefe des deutschsprachigen Raums (1500–1700), <https://www.medizingeschichte.uni-wuerzburg.de/akademie/index.html>, Stand: 11.08.2025.

3 Dieser Beitrag basiert auf dem Vortrag der Autor*innen mit gleichem Titel am 26.06.2025 auf dem 9. Bibliothekskongress 2025 (zugleich 113. BiblioCon) in Bremen.

Im zweiten Teilprojekt «Jeanne Hersch: Digitale Neuauflage der Schriften» sollen die verstreut publizierten Texte der Schweizer Philosophin digital zugänglich gemacht werden. Auch hier wird mit Transkribus der Rohtext erstellt, der anschließend korrigiert, nach den Richtlinien des Deutschen Textarchivs (DTA) ausgezeichnet sowie mit einem Werk- und Personen- bzw. Organisationenregister verbunden wird.

2. Archiv-Editionen als work-in-progress

In beiden Projekten realisiert die ZB eine sogenannte Archiv-Edition. Der Fokus richtet sich vor allem auf den Inhalt der Dokumente. Demzufolge geht es darum, einen guten Lesetext inklusive Register anzubieten, der das Interesse der allgemeinen Leserschaft, die Perspektiven verschiedener wissenschaftlicher Analysen sowie die Nachnutzung und Weiterbearbeitung durch die spezialisierte Editionsphilologie zu befriedigen vermag. Die Edition bietet einen sauberen Text in basaler Auszeichnung; es werden längst nicht alle inhaltlichen und/oder textlichen Phänomene vermerkt und kommentiert: Sowohl die philologischen als auch die inhaltlichen Auszeichnungen sind bewusst generisch gehalten.

Beide an der ZB entstehenden Archiv-Editionen besitzen die Offenheit für verschiedenartige Fragestellungen, die man an sie stellen kann beziehungsweise mit ihnen verfolgen könnte. Sie sind nicht wie viele gedruckte Editionen als einmalige, abgeschlossene Editionen anzusehen, sondern als Work-in-progress-Editionen, welche die erste Stufe der editorischen Aufarbeitung darstellen: Sie bilden die Basis für weitere, zukünftige editorische Vorhaben.

Damit ist ihr aktueller Status bestimmt, der sich in Zukunft ggf. ändern mag. Auch wenn die Archiv-Edition in ihrer Anlage jenes transitorische Momentum beinhaltet, das heutzutage als Charakteristikum für digitale Archive angesehen wird,⁵ unterscheidet sie sich dadurch von digitalen Archiven, dass sie neben dem Editionstext auch eine grundlegende Texterschließung sowie eine historische Einbettung der Dokumente und flexible Zugriffe auf das edierte Material enthält.

Eine solche Edition verfolgt das Prinzip eines modularen Aufbaus und ist in ihrem Kern offen für Nachnutzung und Anreicherung. Ein solches Editionsverständnis ist genuin in der Digitalität der Edition begründet. Es führt auch dazu, dass die Dokumentation eine entscheidende Bedeutung erhält. Auszeichnungsrichtlinien, editorische Entscheide sowie sämtliche technischen Informationen müssen ebenso à jour gehalten und zur Verfügung gestellt werden wie das edierte Textmaterial.

Das Erarbeiten von digitalen Editionen ist nicht nur Grundlagenforschung oder eine Dienstleistung für Forschung und Wissenschaft, sondern auch ein Beitrag zum Aufbau von Forschungsinfrastrukturen. Insbesondere für Gedächtnisinstitutionen stellen Archiv-Editionen ein geeignetes (digitales) Format dar, um einen nachhaltigen und wichtigen Beitrag zum wissenschaftlichen Diskurs zu leisten.

4 Weitere Projektinformationen online unter: Reiling, Jesko; Kreyenbühl, Elias; Kohler, Daniela: Digitale Edition der Korrespondenzen von Hans Georg und Hermann Nägeli, 2024, naegeli-briefe.ch, Stand: 11.08.2025.

5 Vgl. etwa Sahle, Patrick: Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels. Teil 2: Befunde, Theorie und Methodik. Norderstedt 2013, S. 33–34 und 152–154.

3. Von der Kundenanfrage zum Forschungsservice

Beide Editionsprojekte wurden durch Kundenanfragen angestoßen. Das Nägeli-Projekt entwickelte sich durch ein geplantes Forschungsvorhaben am Musikwissenschaftlichen Institut der Universität Zürich, das Hersch-Projekt wurde durch die Jeanne Hersch-Gesellschaft angeregt, die aufgrund der besonderen Relevanz von Hersch als wichtiger Zeitstimme des 20. Jahrhunderts das Œuvre der Schweizer Philosophin digital verfügbar machen wollte.⁶ Auch am Fortgang des Projektes ist die Gesellschaft sehr interessiert und bemüht sich mit großem Engagement erfolgreich, Drittmittel für das Projekt einzuwerben.⁷

In der digitalen Neuauflage der Schriften von Jeanne Hersch werden erstmals sämtliche aktuell bekannten Texte von Hersch gemeinsam versammelt. Es handelt sich hierbei um verschiedene Gattungen: um Monografien, Sammelbandbeiträge, Zeitungs- und Zeitschriftenartikel oder gar auch Interviews und protokollierte Gesprächsrunden mit der Philosophin. Die rund 300 Texte werden in ihren Originalfassungen – vor allem französisch oder deutsch – präsentiert, auf allfällige spätere Übersetzungen wird systematisch hingewiesen.⁸

4. Bibliothekarische und editorische Anforderungen an den Bibliothekskatalog

Der Workflow im Hersch-Projekt startet im Bibliothekskatalog *Swisscovery*.⁹ Da viele Titel von Hersch bereits im Katalog erfasst sind, wird die Titelliste – das zentrale Arbeitsdokument mit allen zu edierenden Werken – direkt aus den bibliografischen Datensätzen des Bibliothekskatalogs generiert. Viele Titel sind allerdings nur unzureichend nachgewiesen und die Metadaten häufig unvollständig oder fehlerhaft; hier müssen Titelsätze bereinigt und ergänzt, gelegentlich auch einmal neu angelegt werden. Ziel ist es, für jeden Titel einen vollständigen und qualitativ hochwertigen bibliografischen Datensatz sicherzustellen. In der Regel erfolgt dies durch die Erstellung analytischer Aufnahmen; bestehende Monografiedatensätze werden bei Bedarf bereinigt. Dass die Kontrolle der Metadaten sowie die (Re-)Katalogisierung einer großen Anzahl einzelner Titel insgesamt so viel Zeit beansprucht, war freilich vor Projektstart nicht abzusehen.

Hier zeigt sich ein besonderer Vorteil von Editionsprojekten, die innerhalb von Bibliotheken realisiert werden: Die ZB kann auf ihre bibliothekarische Expertise zurückgreifen, für das Editionsprojekt selbst hochwertige Metadaten im Katalog erzeugen und führt damit bibliothekarische und editorische Arbeiten zugleich durch. Die erarbeiteten Metadaten bilden die Grundlage für die weiteren Arbeitsschritte

6 Die Anfrage an die ZB erfolgte, weil diese auch den Hersch-Nachlass aufbewahrt; vgl. ZBcollections, 2025, <https://zbcollections.ch/home/#/content/a08abfb9a2a94a27a224697165047344>, Stand: 11.08.2025.

7 Vorstand der Jeanne Hersch-Gesellschaft: Jeanne Hersch. Die herausragende Genfer Philosophin, <http://www.jeanne-hersch-gesellschaft.ch>, Stand: 11.08.2025.

8 Weitere Projektinformationen online unter: Reiling, Jesko; Kreyenbühl, Elias; Rom, Sharon: *Jeanne Hersch. Digitale Neuauflage der Schriften*, 2025, edition-jeanne-hersch.org, Stand: 11.08.2025.

9 Swisscovery, 2025, <https://swisscovery.slsp.ch>, Stand: 11.08.2025. Vgl. dazu auch Bibliotheks-informatik Zentralbibliothek Zürich: ZB Data Map, Juli 2025, <https://data.zb.uzh.ch/map/books/data-map-der-zentralbibliothek-zurich/page/alma-und-pri-move>, Stand: 11.08.2025.

und können dank eines lokal vergebenen Projektcodes leicht und direkt aus Alma übernommen werden.

Swisscovery erleichtert die institutionenübergreifende Zusammenarbeit bei der Erfassung von Titelsätzen. Als Bibliothekskatalog unterliegt er bibliothekarischen Anforderungen, so etwa auch beim Anlegen von thematischen Sammlungen, in denen Titel zu bestimmten Themen zusammengestellt werden können. Im Rahmen des Editionsvorhabens sollte eine Hersch-Kollektion eingerichtet werden. Dabei zeigte sich jedoch eine Einschränkung: In solche Sammlungen können ausschließlich Bestände aus der eigenen bibliothekarischen *Institutional Zone* (IZ) aufgenommen werden – also aus jenem Bereich, der die lokalen Datensätze und Bestände der eigenen Bibliothek enthält. Titel aus der *Network Zone* (NZ), dem gemeinsamen Bereich für bibliografische Datensätze, den alle teilnehmenden Institutionen nutzen, können hingegen nicht integriert werden. Da sich viele Hersch-Titel, die in der Edition ediert werden sollen, nicht im Bestand der ZB befinden, könnten diese in einer solchen Sammlung nicht aufgeführt werden.

Ein weiterer Sachverhalt kam im Projektverlauf ans Licht: Während von den *Frontends* der Editionsprojekte aus auf die entsprechenden Katalogeinträge verlinkt werden kann, ist es bislang nicht möglich, umgekehrt auf einfache – geschweige denn auf automatisierte – Weise vom Bibliothekskatalog auf die zugehörigen Editionsprojekte zu verweisen. Wünschenswert wäre eine technische Lösung, mit der bibliografische Datensätze direkt mit den zugehörigen Volltexten verlinkt und darüber hinaus auch in thematische Sammlungen integriert werden können.

5. Textauszeichnungen mit KI

Die digitale Neuauflage der Schriften von Hersch orientiert sich bei der Textauszeichnung grundsätzlich an den vom Deutschen Textarchiv (DTA) entwickelten «Richtlinien zur Transkription», die eine auf dem internationalen TEI-Standard basierende Spezifikation zur strukturierten Auszeichnung historischer deutschsprachiger Texte darstellen.¹⁰ Ergänzend dazu wurden projektspezifische Richtlinien formuliert.

Für die automatische Texterkennung und Erstellung der Rohtexte wird Transkribus mit dem Modell «Print M1» eingesetzt. Dieses Modell erzielt im Projekt bislang gute Ergebnisse und weist eine *Character Error Rate* (CER) von lediglich 2.2% auf. Herausfordernd sind jedoch insbesondere mehrspaltige Layouts, wie sie häufig in Zeitungsartikeln vorkommen. Transkribus stellt mit seinen Segmentierungsfunktionen und vordefinierten Tags zwar eine solide Basis für die TEI-Auszeichnung des Layouts bereit – etwa durch die Markierung von Absätzen mit <p>, Zeilenumbrüchen mit <lb> oder Überschriften mit entsprechenden Tags. Diese Tags reichen aber nicht aus, um alle Anforderungen des DTA-konformen TEI-Standards abzudecken, weshalb nachträgliche Anpassungen und Ergänzungen erforderlich sind.

¹⁰ Vgl. hierzu Richtlinien zur Transkription, 2025, <https://www.deutschestextarchiv.de/doku/basisformat/transkription.html>, Stand: 11.08.2025.

Im Zuge dieser Anpassungen kam die Idee auf, KI-gestützte Verfahren anzuwenden, um die formalen Strukturen direkt im gewünschten DTA-Format auszuzeichnen – ohne den Umweg über den Transkribus-Output. Da die Auszeichnung nach festen Regeln erfolgt und keine kreative, sondern eine regelgeleitete Tätigkeit darstellt, bietet sich für eine effizientere Verarbeitung grösserer Textmengen ein solches Verfahren durchaus an. Vor diesem Hintergrund testet das Projektteam derzeit den Einsatz generativer KI zur strukturierten Layouterkennung und automatischen Auszeichnung formaler Elemente wie Absätzen oder direkter Rede. Ziel ist die Generierung eines vollständigen, validen XML-Dokuments im TEI-Format, wofür aktuell ChatGPT eingesetzt wird. Sobald dieses Verfahren etabliert ist, sind weitere KI-gestützte Anwendungen geplant – etwa *Named Entity Recognition* und *-Linking*, um Personen- und Körperschaftsnamen automatisch zu identifizieren und mit der Gemeinsamen Normdatei (GND) zu verknüpfen.

Jesko Reiling, Zentralbibliothek Zürich, <https://orcid.org/0000-0001-8132-3562>

Sharon Rom, Zentralbibliothek Zürich, <https://orcid.org/0000-0002-1662-8885>

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/6184>

Dieses Werk steht unter der Lizenz [Creative Commons Namensnennung 4.0 International](#).