

It's a Match

Digital Scholarship Services als zu gestaltender Freiraum zwischen Bibliotheken und Forschung

Michael Gasser, *ETH Zürich, ETH-Bibliothek, Sammlungen und Archive*

Christiane Sibille, *ETH Zürich, ETH-Bibliothek, Digital Scholarship Services*

Zusammenfassung

Das Potential digitaler Kollektionen ist enorm: Auf der einen Seite stellen Bibliotheken und andere Gedächtniseinrichtungen in hoher Qualität digitale und digitalisierte Inhalte in grossem Umfang und inhaltlicher Breite zur Verfügung. Auf der anderen Seite gibt es zahlreiche und rasch wachsende Forschungsfelder, die einen grossen Bedarf an solchen Datensätzen haben, um sie für ihre digitale Forschung zu nutzen. An dieser Schnittstelle zwischen Beständen und innovativen Formen der Nachnutzung entstehen Freiräume, in denen das Zusammenspiel von bibliothekarischen Inhalten und Services, Open Science und Co-Creation gestaltet werden kann und in denen die Bibliothek eine aktive Rolle als Kooperationspartner einnimmt. Um dieses Potential zu nutzen, richteten die Sammlungen und Archive der ETH-Bibliothek Mitte 2021 die Fachstelle Digital Scholarship Services ein. Im Zentrum des Beitrags stehen Beweggründe und Einblicke in die Praxis der Aufbauphase der Fachstelle sowie (Pilot-)Projekte aus dem Machine-Learning-Bereich. Welche Forschungsinteressen haben etwa Informatikprofessuren an Kollektionen digitalisierter Alt- oder Archivbestände? Wie können entsprechende Kooperationen initialisiert und verstetigt werden? Welche Anforderungen an Kompetenzen, Daten und Infrastruktur werden seitens der Forschenden gestellt und wo müssen auch innerhalb der Bibliothek neue Formen der Zusammenarbeit erprobt werden? Welche Mehrwerte können zu gegenseitigem Nutzen realisiert werden? Ein erstes Fazit zum Beitrag der Digital Scholarship Services zum Fokus, bestehende Inhalte KI-gestützt anzureichern und damit neue Arten der Nutzung zu ermöglichen, rundet den Beitrag ab.

Summary

The potential of digital collections is enormous: on the one hand, libraries and other memory institutions provide extensive, high-quality digital and digitised content in great volume and breadth of content. On the other hand, there are numerous and rapidly growing research fields that have a great need for such datasets in order to use them for their digital research. At this interface between collections and innovative forms of re-use, free spaces are created in which the interplay of library content and services, open science and co-creation can be shaped and in which the library takes on an active role as a cooperation partner. In order to make use of this potential, ETH Library's Collections and Archives set up the Digital Scholarship Services Unit in mid-2021. The article focuses on motivations and (pilot) projects from the machine learning area as well as insights into the practice of the initial phase of the specialist unit. What research interests, for example, do computer science professors have in collections of digitised old or archive holdings? How can corresponding collaborations be initialised and consolidated? What requirements do the researchers have in terms of competencies, data and infrastructure, and where do new forms of cooperation need to be tested within the library? What added value can be realised for mutual benefit? An initial conclusion on the contribution of digital scholarship services to the focus of enriching existing content with AI support and thus enabling new types of use rounds off the article.

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/5856>

Autorenidentifikation: Gasser, Michael: ORCID: <https://orcid.org/0000-0003-0390-1448>;
Sibille, Christiane: ORCID: <https://orcid.org/0000-0003-3689-2154>

Schlagwörter: Bibliothek ; Digital Scholarship ; Maschinelles Lernen ; Machine Learning ; Künstliche Intelligenz ; Artificial Intelligence (AI)

Dieses Werk steht unter der Lizenz [Creative Commons Namensnennung 4.0 International](https://creativecommons.org/licenses/by/4.0/).

1. Digital Scholarship als Zusammenspiel verschiedener Komponenten

Im vorliegenden Artikel, der die Aktivitäten im Bereich Digital Scholarship Services an der ETH-Bibliothek beschreibt, schlagen wir einen Ansatz vor, der das Zusammenspiel von wissenschaftlichen Fragestellungen, digitalen Methoden und Tools sowie die Nutzung digitaler Kollektionen ins Zentrum stellt. Die ausgewählten Fallbeispiele zeigen dabei auf, dass dies alles keine reine Support-Funktion hat, sondern dass die in diesem Kontext realisierten Kooperationen innovative Freiräume schaffen, in denen durch Co-Creation neue Datenebenen, explorative Services und vielfältige Möglichkeiten zum Wissenstransfer entwickelt werden.

In den vergangenen zehn Jahren haben sich, insbesondere in Bibliotheken im englischsprachigen Raum, unterschiedliche Dienstleistungen unter dem Sammelbegriff Digital Scholarship Services (DSS) etablieren können. Diese reichen von der Digitalisierung von Beständen über das Bereitstellen von Tools und digitalen Publikationsplattformen bis zur Beratung bei der konkreten Umsetzung digitaler Projekte.¹ Gemeinsames Ziel ist die Unterstützung von Digital Scholarship. Was der Begriff Digital Scholarship umfasst, wird in der Regel nur sehr lose definiert. So schreibt beispielsweise die University of Oxford auf ihrer Website:

«Scholarship’ is the collective attainments of experts working within a particular field of academic study, especially but not exclusively in the humanities. ‘Digital scholarship’ takes place when digital technology is used to enhance the materials or methods available to scholars.»²

Hierbei werden zwei wichtige Aspekte deutlich. Erstens, dass der Begriff explizit eine breite fachliche Ausrichtung impliziert, aber häufig der Schwerpunkt in den Geisteswissenschaften liegt. Dies legt wiederum die Frage nahe, wodurch sich Digital Scholarship und Digital Humanities unterscheiden. Tatsächlich wurden beide Begriffe in der Vergangenheit häufig synonym verwendet, wobei die Digital

1 Wiggins et al. bieten eine aktuelle systematische Querschnittsstudie zur Rolle von Digital Scholarship an US-amerikanischen Universitäten: Wiggins, Benjamin; Hennesy, Cody; Vetruba, Brian et al.: Digital Scholarship Programs in Practice, in: *College & Research Libraries* 83 (4), 2022, 01.07.2022. Online: <<https://doi.org/10.5860/crl.83.4.568>>.

2 What is Digital Scholarship?, <<https://digitalscholarship.web.ox.ac.uk/what-digital-scholarship>>, Stand: 26.08.2022.

Humanities dabei als die Umsetzung von Digital Scholarship in den Geisteswissenschaften galten.³ Die oben zitierte Definition deutet jedoch darauf hin, dass in jüngerer Zeit eine Ausdifferenzierung stattfindet, die Digital Scholarship primär als die Anwendung digitaler Methoden, unabhängig von der fachlichen Zugehörigkeit, umschreibt.⁴ Ein zweiter wichtiger Aspekt ist die Inklusivität des vorgeschlagenen Ansatzes. Digital Scholarship umfasst nicht nur ein breites Feld an Fächern, sondern bezieht – in der obigen Definition auffallend allgemein formuliert – alle Expert*innen mit ein, die in dem jeweiligen akademischen Feld arbeiten. In der Praxis bedeutet dies, dass die Aktivitäten im Bereich Digital Scholarship häufig an den Bibliotheken und nicht in den Fächern selbst angesiedelt sind, wie es im oben erwähnten Beispiel aus Oxford und an der ETH Zürich der Fall ist. Wie sinnvoll die breite, viele Fachrichtungen einschliessende Definition von Digital Scholarship ist, zeigt sich dabei im Fall der ETH-Bibliothek, der grössten öffentlichen naturwissenschaftlichen und technischen Bibliothek der Schweiz, besonders deutlich. Neben möglichen Kooperationen mit dem Departement Geistes-, Sozial- und Staatswissenschaften der ETH Zürich waren es in der Vergangenheit auch andere Fachbereiche, wie Architektur und Informatik, mit denen entsprechende Projekte realisiert werden konnten.

Während der disziplinäre Zugang zu Digital Scholarship ebenso breit sein kann wie der institutionelle Hintergrund der beteiligten Expert*innen, gibt es aber ein wichtiges verbindendes Element: die Arbeit mit digitalen Sammlungen bzw. das Konzept von *collections as data*. Zu den Sammlungs- und Archivbeständen der ETH-Bibliothek gehören unter anderem das Hochschularchiv, die Sammlung Alter Drucke und Karten, das Bildarchiv, das Kunstinventar der ETH, die Sammlung wissenschaftlicher Instrumente und Lehrmittel sowie zwei Literaturarchive (Max-Frisch-Archiv und Thomas-Mann-Archiv). Zusätzlich zur Digitalisierung von Material aus diesen Beständen hat sich die ETH-Bibliothek in der Vergangenheit an weiteren umfangreichen Digitalisierungsprojekten beteiligt. Diese digitalen Sammlungen stehen im Zentrum der im Folgenden vorgestellten Projekte.

2. Digitale Kollektionen im Zentrum

Maarten Delbeke, Professor für Geschichte und Theorie der Architektur am Departement Architektur der ETH Zürich und Kooperationspartner der ETH-Bibliothek im Projekt «Graph – Text Reuse in Rare Books» beschreibt die Möglichkeiten der computer- und datengestützten Forschung wie folgt:

The exponentially increased availability of digitised sources in standardized format brings new opportunities for automated interaction with corpora that far exceed the capacity of manual exploration. Computational methods and the rise of artificial intelligence have opened new avenues in how researchers can process and analyse documents. The contents of digitised materials can be extracted more faithfully than ever. Text and images can be understood at an unprecedented level.⁵

3 2009 verwendete Julia Flanders auch die Bezeichnung «Digital Humanities Scholarship»: Flanders, Julia: The Productive Unease of 21st-century Digital Scholarship, in: Digital Humanities Quarterly 3 (3), 2009. Online: <<http://www.digitalhumanities.org/dhq/vol/3/3/000055/000055.html>>; Ausserdem: Thomas, William G., III: The Promise of the Digital Humanities and the Contested Nature of Digital Scholarship, in: A New Companion to Digital Humanities, Chichester 2015, S. 524–537. Online: <<https://doi.org/10.1002/9781118680605.ch36>>.

4 What is Digital Scholarship?, <<https://digitalscholarship.web.ox.ac.uk/what-digital-scholarship>>, Stand: 26.08.2022.

5 Delbeke, Maarten: About [the Graph] project, Graph – Text reuse in rare books, <<https://graph-rare-books.ethz.ch/#/about>>, Stand: 26.08.2022.

Dieses Zitat bringt in knapper Form zum Ausdruck, welche vier Komponenten im Bereich Digital Scholarship an der ETH-Bibliothek die Hauptrollen spielen: Ausgangspunkt sind *wissenschaftliche Fragestellungen*, die mithilfe *digitaler Methoden* und den entsprechenden *Tools* auf *digitale Kollektionen* angewendet werden.

Die Vorstellung von Digital Scholarship als Zusammenspiel dieser vier Komponenten eröffnet der Bibliothek einen breiten Handlungsspielraum, der vielfältige Kooperationsmöglichkeiten mit unterschiedlichen Fachbereichen und grossem Potential für interdisziplinären Austausch ermöglicht. Aus der Perspektive der in der Sektion Sammlungen und Archive der ETH-Bibliothek angesiedelten Digital Scholarship Services sind dabei die eigenen digitalen Sammlungen Dreh- und Angelpunkt dieser Aktivitäten.⁶

Wie andere Gedächtnisinstitutionen auch, hat die ETH-Bibliothek in den vergangenen Jahrzehnten in grossem Umfang analoge Publikationen und Archivalien digitalisiert und bereitgestellt. Zu den wichtigsten Plattformen, die die ETH-Bibliothek – zum Teil im Verbund mit anderen Schweizer Bibliotheken – betreibt, gehören unter anderem e-rara.ch, e-manuscripta.ch, E-Periodica und E-Pics. Jedes dieser Angebote birgt ein grosses Potential für die Nutzung im Kontext von Digital Scholarship, nicht zuletzt dank der breiten verfügbaren Datenbasis. So enthält e-rara.ch mehr als 91 000 digitalisierte alte und seltene Drucke, darunter auch historische Karten. Das E-Pics-Bildarchiv stellt über 640 000 digitalisierte Bilder zur Verfügung, die in vielen Fällen durch entsprechende Rechteabklärungen mit Public Domain Mark bzw. Creative-Commons-Lizenzen zur Verfügung gestellt werden. Zudem erprobt das Bildarchiv unterschiedliche Formen neuer Erschliessungstechnologien, beispielsweise mit Hilfe von Crowdsourcing oder durch den Einsatz künstlicher Intelligenz.⁷ Die Plattform e-manuscripta.ch, die von der ETH-Bibliothek technisch gehostet wird, enthält annähernd 130 000 handschriftliche Quellen, wozu neben Texthandschriften beispielsweise auch eine grosse Sammlung sogenannter Sonnenflecken-Zeichnungen gehört, auf denen an Beobachtungstagen nach bestimmtem Schema die auf der Sonne sichtbaren Sonnenflecken festgehalten wurden. E-Periodica umfasst schliesslich einen Bestand von ca. 450 digitalisierten Zeitschriften, in denen ungefähr eine Million Artikel erschlossen sind, was einer Gesamtmenge von weit über 9 Millionen Seiten entspricht.

Die Vielfalt dieser digitalen Daten eröffnet ein nahezu unerschöpfliches Reservoir für mögliche Digital-Scholarship-Anwendungen, insbesondere im Bereich der Künstlichen Intelligenz bzw. des Machine Learnings. Man denke nur an verbesserte Verfahren zur Handschriftenerkennung und Dokumentenstrukturierung, an die Nutzung von Computer-Vision-Verfahren für die Analyse historischer Bilder und Ähnliches. Erste Projekte zur Umsetzung dieser Möglichkeiten sollen im Folgenden kurz skizziert werden.

6 Dieser Fokus ist unter anderem dadurch möglich, dass für andere bibliothekarische Dienstleistungen wie beispielsweise das Forschungsdatenmanagement und E-Publishing eigene spezialisierte Einheiten vorhanden sind, siehe: Organigramm der ETH-Bibliothek (1. Juni 2022), <https://ethz.ch/content/dam/ethz/associates/ethlibrary-dam/documents/UEber-uns/Organigramme/Organigramm_20220601_web_d.pdf>, Stand: 26.08.2022.

7 Graf, Nicole: «Wissen Sie mehr?» Erfolgsfaktoren beim Crowdsourcing: Ein Erfahrungsbericht des Bildarchivs der ETH-Bibliothek, Zürich, in: Rundbrief Fotografie 27 (1), 2020, S. 28–38. Online: <<https://doi.org/10.3929/ethz-b-000410818>>; Dies.: Alles unter Kontrolle? KI im Einsatz im Bildarchiv der ETH-Bibliothek, Folien der Präsentation am 8. Bibliothekskongress in Leipzig, 31. Mai 2022. Online: <<https://doi.org/10.3929/ethz-b-000551925>>, Stand: 26.08.2022.

3. Veranschaulichtes Potential: Graph – Text Reuse in Rare Books

Ein Beispiel für die Nutzung digitalisierter Bestände aus e-rara ist das Projekt «Graph – Text Reuse in Rare Books», das im Jahr 2020, also noch vor der Einrichtung einer eigenen Fachstelle für Digital Scholarship Services, realisiert wurde. Es handelt sich dabei um eine Kooperation zwischen der ETH-Proessur für Geschichte und Theorie der Architektur von Maarten Delbeke und der ETH-Bibliothek. Das Projekt und die daraus resultierende Plattform haben zum Ziel, die Frage zu beantworten, wie sich in Architekturtraktaten des 17. bis 19. Jahrhunderts Textpassagen verbreiteten. Es geht dabei nicht um historische Plagiatsforschung, sondern darum, mittels computergestützter Textanalysen und Visualisierungen ein Rechercheinstrument bereitzustellen, das die Forschung bei der Analyse unterstützt, wie innerhalb des abgedeckten Bereichs damals Wissen kumuliert, weitergegeben und erweitert wurde.

Das zugrundeliegende Textkorpus umfasst gut 1 300 Architekturtraktate aus dem Bestand der Alten und Seltenen Drucke der ETH-Bibliothek. Diese Traktate waren bereits zu Projektbeginn in digitalisierter Form auf der Plattform e-rara frei verfügbar. Die Metadaten konnten von dort über einen OAI-PMH-Endpoint, die Digitalisate über eine mit den Standards des International Image Interoperability Framework (IIIF) konforme API eingebunden werden.

Die aus dem Projekt resultierende und frei zugängliche Applikation «Graph – Text Reuse in Rare Books» funktioniert als Recherche- und Analysetool auf drei Ebenen:

- Auf der obersten Ebene wird – im Sinne eines Überblicks über den gesamten Korpus – als Graph sichtbar gemacht, in welchen Traktaten gleiche oder sehr ähnliche Textstellen zu finden sind.
- Durch die Auswahl einer Verbindung zwischen zwei Traktaten lassen sich auf unterster Ebene die innerhalb der beiden Werke identifizierten ähnlichen Textstellen anzeigen und direkt miteinander vergleichen.
- Als Erweiterung wurde 2022 zusätzlich eine mittlere Ebene implementiert. Sie visualisiert auf Stufe der Inhaltsverzeichnisse zweier Titel, wie sich wiederverwendete Textstellen verteilen. Ausmaß und Muster in der Weiterverwendung von Textpassagen werden dadurch augenfällig dargestellt.

Das Beispiel des Projekts und der daraus entstandenen Applikation «Graph – Text Reuse in Rare Books» illustriert, welche innovativen und interessanten Möglichkeiten das Feld der Digital Scholarship für konkrete und forschungsnahe Kooperationsprojekte zwischen Bibliotheken und Forschenden eröffnet. Übertragen auf das einfache Digital-Scholarship-Modell mit seinen vier Komponenten *wissenschaftliche Fragestellungen, digitale Methoden, Tools und digitale Kollektionen* gilt für dieses Projekt, dass sich die Rolle der ETH-Bibliothek in erster Linie auf den Bereich der digitalen Kollektionen und das damit verbundene Wissen über diese Bestände konzentrierte. Als Projektleiterin seitens der ETH-Bibliothek war Meda Hotea, die Leiterin der Gruppe Rara und Karten, ganz wesentlich an der Identifikation von Architekturtraktaten und der Zusammenstellung des Korpus beteiligt. Bibliotheksinterne

IT-Unterstützung erhielt das Projekt in der Datenbereitstellung und beim Betrieb der Applikation. Benoit Seguin sorgte aus fachlicher Perspektive unter Verwendung und Kombination existierender Methoden und Tools auch für die computergestützte Analyse des Textkorpus und die Realisierung der Applikation.⁸ Somit gab die Professur nicht nur den inhaltlichen Anstoss zum Projekt, sondern übernahm auch die eigentliche technische Umsetzung.

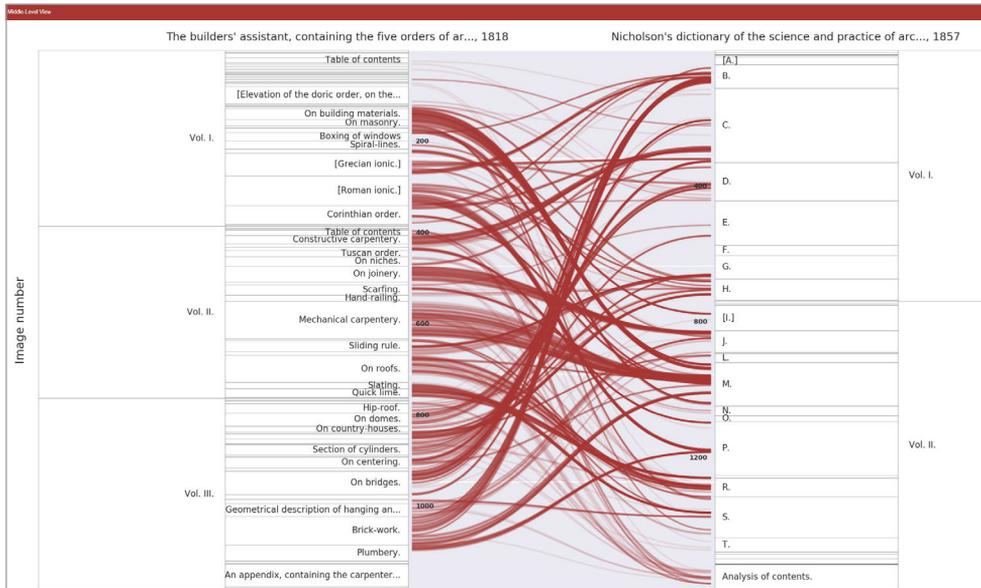


Abb. 1: Der neue Middle Layer von «Graph – Text Reuse in Rare Books» macht die Weiterverwendung von Textstellen auf Ebene der Inhaltsverzeichnisse zweier Titel sichtbar

4. Einrichtung der Fachstelle Digital Scholarship Services

Die Umsetzung des Projekts «Graph – Text Reuse in Rare Books» legte eine weiterführende Frage nahe: Wenn aus der damals bestehenden Organisationsstruktur heraus forschungsnahe Kooperationsprojekte im Bereich Digital Scholarship initialisiert und umgesetzt werden konnten, um wie viel effektiver könnte zusätzliches Potential genutzt und ausgeschöpft werden, wenn die Bibliothek selbst – insbesondere im Hinblick auf die Nutzung ihrer digitalisierten Bestände – über mehr eigene Kompetenzen in entsprechenden Methoden und Tools verfügen würde? Wie internationale Beispiele aus anderen wissenschaftlichen Bibliotheken zeigen, spielen gerade kleine, spezialisierte Einheiten, die häufig mit dem Sammelbegriff «GLAM Labs» bezeichnet werden, eine wichtige Rolle dabei, in Zusammenarbeit mit der Forschung die digitale Nutzung digitaler Kollektionen im Sinne von *collections as*

⁸ Seguin, Benoit: Method, Graph – Text reuse in rare books, <<https://graph-rare-books.ethz.ch/#/about>>, Stand: 26.08.2022.

data zu fördern und zu entwickeln.⁹ Auffällig ist, dass auch mehrere Nationalbibliotheken – etwa die Labs der British Library¹⁰ oder das KB Lab der niederländischen Nationalbibliothek¹¹ – über entsprechende Einheiten die digitale Nutzung ihrer digitalen Kollektionen aktiv stärken und vorantreiben.

Um strategisch und gezielt in diesen Bereich zu investieren, entschied die ETH-Bibliothek, Mitte 2021 in der Sektion Sammlungen und Archive die neue Fachstelle Digital Scholarship Services einzurichten. Die Ausrichtung der neuen Fachstelle umfasst dabei in erster Linie folgende Ziele:

- Kompetenzen im Bereich Digital Scholarship innerhalb der Bibliothek ausbauen,
- aktiv neue Kooperationen mit Forschenden der ETH Zürich eingehen,
- Pilotprojekte initiieren und weiterentwickeln,
- neue Services der ETH-Bibliothek im Bereich Digital Scholarship aufbauen.

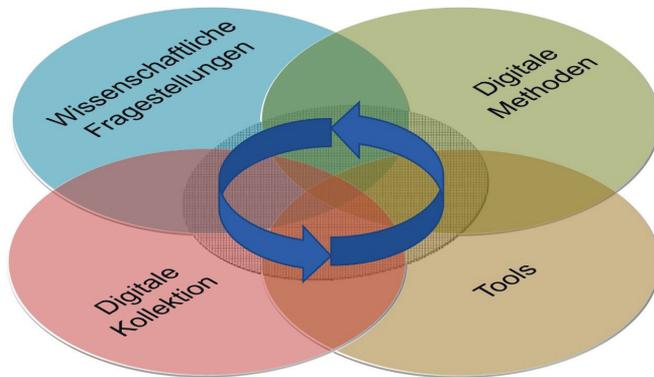


Abb. 2: Die Digital Scholarship Services der ETH-Bibliothek gestalten die Schnittstelle zwischen wissenschaftlichen Fragestellungen, digitalen Kollektionen, Methoden und Tools als Raum der Co-Creation zwischen Forschung und Bibliothek

Verglichen mit diesen weitreichenden Zielen ist die personelle Ausstattung der Fachstelle Digital Scholarship bisher verhältnismäßig bescheiden. Neben der festangestellten Verantwortlichen der Fachstelle (80 %-Stelle) umfasst sie ein zusätzliches 80 %-Pensum, das mit einer wechselnden Anzahl Masterstudierender unterschiedlicher Studienrichtungen (z.B. Informatik oder Computerlinguistik) besetzt wird, die projektspezifisch ausgewählt und eingesetzt werden. Ein Schwerpunkt der Fachstelle ist es, die digitale Perspektive der Sektion Sammlungen und Archive in Kooperationen einzubringen, Pilotprojekte zu realisieren und die so gewonnenen Erkenntnisse und Daten in unterschiedlicher Form nachnutzbar zu machen.

9 Vgl. Mahey, Mahendra et al.: Open a GLAM Lab. Digital Cultural Heritage Innovation Labs, Book Sprint, Doha, Qatar, 23.–27.09.2019. Online <<https://glamlabs.io/books/open-a-glam-lab/>>, Stand: 26.08.2022 oder Candela, Gustavo et al.: Reusing digital collections from GLAM institutions, in: Journal of Information Science 24, August 2020, S. 1–17. Online: <<https://doi.org/10.1177/0165551520950246>>.

10 British Library Labs, <<https://www.bl.uk/projects/british-library-labs>>, Stand 26.08.2022.

11 KBLAB, <<https://lab.kb.nl/>>, Stand 26.08.2022.

Nicht zuletzt dank einer konsequent auf Vernetzung und Kooperation innerhalb der ETH-Bibliothek und der ganzen ETH Zürich ausgerichteten Herangehensweise zeigt sich bereits nach einem Jahr, wie die Fachstelle auf den verschiedenen Gebieten dazu beiträgt, die Bibliothek auf Augenhöhe der Forschung als Partnerin in Digital-Scholarship-Projekten zu positionieren.

5. Digital Scholarship Services in der Praxis

Wie bereits beschrieben, bilden aus der Perspektive der Bibliothek die digitalen Sammlungen das Drehkreuz für die Aktivitäten im Bereich der Digital Scholarship Services. Im Hinblick auf die weiteren Elemente (Fragestellungen, Methoden, Tools) lassen sich weitere Gemeinsamkeiten erkennen. Die wichtigste Rolle spielt hierbei die Erarbeitung zusätzlicher Daten-Ebenen, die Schicht für Schicht das ursprüngliche Digitalisat mit zusätzlichen Informationen anreichern. Ein Beispiel für die Umsetzung eines solchen Layers ist die Nutzung der Software Transkribus für die automatisierte Transkription handschriftlicher Texte mit Hilfe von Machine Learning.¹² Hierbei handelt es sich um ein Anwendungsbeispiel, das sich im engeren Kompetenzbereich der Sektion Sammlungen und Archive bewegt. Durch das Vorhandensein des entsprechenden Tools für die eigentliche Transformation ist zunächst keine weitere Kooperation mit zusätzlichen Partnern notwendig. Das in diesem Kontext gewonnene Wissen über die Nutzung des Tools, das Erstellen von Modellen und das einfachere Vermitteln handschriftlicher Texte kann jedoch in unterschiedliche Form weitergegeben werden. So bestehen bereits Partnerschaften innerhalb der Bibliothek, zum Beispiel mit der im Jahr 2022 gestarteten Edition der Notizhefte von Max Frisch sowie einem Projekt zur Transformation der handschriftlichen ETH-Schulratsprotokolle, die von der Gründung bis ins 20. Jahrhundert reichen. Außerhalb der Bibliothek, aber innerhalb der ETH Zürich wird aktuell ein Projekt für die mögliche Nutzung transformierter handschriftlicher Texte in Zusammenarbeit mit dem AI + Art Programm des ETH AI Centers geplant. Schließlich werden außerdem Kooperationen im Bereich der (geschichtswissenschaftlichen) Lehre angestrebt, in deren Rahmen die Studierenden anhand von Reiseberichten und Tagebüchern von ETH-Forschenden sowie anderen handschriftlichen Archivbeständen Transkriptions- und erste Editionserfahrungen sammeln können, die dann auch das Erstellen von Modellen für die Nutzung in Transkribus beinhalten.

Die durch die maschinelle Transkription und andere Verfahren gewonnenen Datenschichten eröffnen wiederum selbst interessante Nachnutzungsmöglichkeiten. Sei es dadurch, dass der Inhalt einfacher erfasst werden kann, oder dadurch, dass der transformierte Text selbst wieder Ausgangspunkt für weitere Layers wird, die beispielsweise durch Natural Language Processing (NLP) Verfahren erschlossen werden. Die Integration dieser Datenschichten in bestehende Systeme wird dabei angestrebt.

Als erstes größeres NLP-Projekt, das vom bibliotheksinternen DigiCenter geleitet wurde und an dem die Fachstelle Digital Scholarship Services beteiligt war, konnte im Juli 2022 das Projekt «E-Periodica Next Level Access» abgeschlossen und in einer ersten Beta-Version online gestellt werden.¹³ Auf

12 Transkribus | KI-gestützte Handschrifterkennung, <<https://readcoop.eu/de/transkribus/>>, Stand: 26.08.2022.

13 E-Periodica: Next Level Access, <<https://nla.e-periodica.ch>>, Stand: 26.08.2022.

der Basis eines Vorprojekts¹⁴ wurde im Rahmen von «E-Periodica Next Level Access» zunächst in Zusammenarbeit mit Computerlinguist*innen auf dem gesamten Bestand von E-Periodica – aktuell mehr als 9 Millionen Seiten verteilt auf ca. 450 unterschiedliche schweizerische Zeitschriften – ein Named Entity Recognition Modell trainiert, das ca. 40 Millionen Personennamen erkannt hat. Diese konnten auf einen Kernbestand von 20 Millionen Namen reduziert werden, die in einem nächsten Schritt mit Hilfe von regelbasierten Verfahren mit der Gemeinsamen Normdatei (GND) der Deutschen Nationalbibliothek abgeglichen und auf Übereinstimmungen bei Vornamen, Nachnamen und Berufen überprüft wurden. Auf diese Weise entstanden ca. 800 000 automatische Verknüpfungen zur GND. Abhängig von der Vollständigkeit der verwendeten Basisparameter kann diese Zuordnung derzeit noch unterschiedlich präzise sein. Diese Unschärfe wurde jedoch durch die Einführung eines Präzisionslevels aufgefangen.

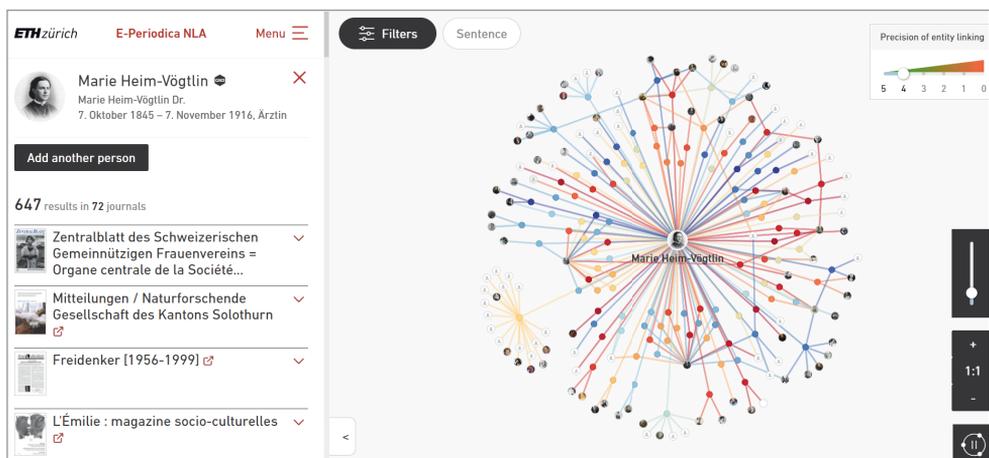


Abb. 3: Screenshot des Netzwerks von Marie Heim-Vögtlin in E-Periodica Next Level Access. Online: <https://nla.e-periodica.ch/gnd-118977962>

Das Projekt beschränkte sich jedoch nicht nur auf die rein technische Umsetzung dieser Anreicherungsebene, sondern zielte explizit darauf ab, die so gewonnenen Informationen auch in einer eigenen Applikation online zugänglich zu machen. Auf diese Weise sollten zusätzliche Formen der Content-Exploration getestet werden. Hierfür wurde eine Visualisierung gewählt, die Personen, die im gleichen Satz, auf der gleichen Seite, bzw. im gleichen Inhaltselement vorkommen, in einem interaktiven Graph aufbereitet und darstellt. Zudem können Userinnen und User durch weitere Filtermöglichkeiten einen bestimmten Zeitraum auswählen oder die Suche auf einzelne Zeitschriften beschränken. Durch die interaktive Verknüpfung des Netzwerks mit den digitalisierten Beständen auf der einen Seite und die Hervorhebung der erkannten Entitäten innerhalb des bestehenden E-Periodica-Portals auf der anderen Seite besteht die Möglichkeit, zwischen beiden Zugangsmöglichkeiten hin und her zu navigieren.

14 Gasser, Michael; Wanger, Regina; Prada, Ismail: Wenn Algorithmen Zeitschriften lesen. Vom Mehrwert automatisierter Textanreicherung, in: o-bib. Das offene Bibliotheksjournal 5 (4), 2018, S. 181-192. Online: <<https://doi.org/10.5282/o-bib/2018H4S181-192>>.

Ein wichtiger Aspekt dieser stark aus der Bibliothek selbst betriebenen Initiativen ist die Nachhaltigkeit. Das bedeutet unter anderem, dass Lösungen angestrebt werden, die nicht nur auf einer einmaligen Datenprozessierung basieren, sondern so aufgebaut sind, dass auch künftige neue Daten möglichst automatisiert in den jeweiligen Workflow integriert werden können. Zudem wird kontinuierlich geprüft, ob Lösungen, die im Hinblick auf einen spezifischen Bestand entwickelt wurden, auch in anderen Kontexten nachgenutzt werden können.

Hierfür können die Digital Scholarship Services zunächst auf die Angebote der bibliothekseigenen IT-Services zurückgreifen, die neben der Applikationsbetreuung auch Dienstleistungen im Bereich Data Science und Research Support anbieten und ein eigenes API-Portal betreiben, und kooperieren zusätzlich mit den zentralen Scientific IT-Services der ETH Zürich, aktuell beispielsweise zur Weiterentwicklung des oben beschriebenen Named Entity Linkings hin zu einem auf Machine Learning basierenden Verfahren. Zudem kann innerhalb der Bibliothek auch auf weitere Kompetenzen, beispielsweise im Bereich der Forschungsdienstleistungen, zurückgegriffen werden.

An einer technischen Hochschule wie der ETH Zürich wird die Bedeutung eines breit angelegten, möglichst alle Fachrichtungen einschließenden Verständnisses von Digital Scholarship Services besonders deutlich. Da die Forschenden in der Regel über hervorragende digitale Kompetenzen und Infrastrukturen verfügen, bzw. durch ihre Arbeiten oft auch im hohen Masse zur Weiterentwicklung digitaler Technologien beitragen, ist Unterstützung durch die Bibliothek in diesem Bereich kaum ein Thema. Gleichzeitig gewinnt jedoch ein anderer Aspekt an Bedeutung, nämlich die Bereitstellung und Nutzung von *collections as data*, wozu hier auch das spezifische, in der Bibliothek vorhandene, Domänenwissen über Bestände, Digitalisierungsverfahren und Metadaten gezählt werden kann. Ein Beispiel für eine solche Kooperation zwischen den Digital Scholarship Services und dem Department Informatik der ETH Zürich ist die Zusammenarbeit mit dem DS3Lab.¹⁵ Für die Arbeiten von Ce Zhang und seinem Team konnten verschiedene Datenbestände für die Nutzung in Forschung und Lehre (z.B. BA- und MA-Arbeiten) im Umfeld des DocParser-Projekts zur Verfügung gestellt werden.¹⁶

In den Fällen, in denen sich die Bibliothek in dieser Form als Anbieterin von *collections as data* beteiligt, entsteht durch den zunehmend etablierten FAIR-Data Gedanken ein eigener Datenkreislauf, in dem am Ende des Projekts neue, angereicherte Datenebenen oder Modelle auch der Bibliothek zur Verfügung stehen. Diese können wiederum in bestehende Services integriert werden und verbessern so das bestehende Angebot. Einer der wichtigsten Faktoren ist jedoch der Gewinn an Wissen über neue Datentechnologien und ihre Anwendungen sowie die Schaffung eines Kooperationsnetzwerks. Auf diese Weise überführen die Digital Scholarship Services eine wichtige Tradition analoger Archive und Sammlungen ins digitale Zeitalter, nämlich die gemeinsame interdisziplinäre Auseinandersetzung mit den jeweiligen Beständen und Objekten.

15 DS3Lab: Data Sciences, Data Systems, & Data Services, <<https://ds3lab.inf.ethz.ch/>>, Stand: 26.08.2022.

16 Rausch, Johannes; Martinez, Octavio; Bissig, Fabian u.a.: DocParser: Hierarchical Document Structure Parsing from Renderings, in: Proceedings of the AAAI Conference on Artificial Intelligence 35 (5), 2021, S. 4328–4338. Online: <<https://ojs.aaai.org/index.php/AAAI/article/view/16558>>. Stand: 26.08.2022.

6. Fazit

Durch die beschriebene gemeinsame Arbeit mit und an digitalen Beständen gestalten die Digital Scholarship Services an der ETH-Bibliothek in Zusammenarbeit mit unterschiedlichen Partnern die Schnittstellen zwischen diesen Beständen, digitalen Tools und Methoden sowie wissenschaftlichen Fragestellungen. Das Schaffen dieser dynamischen und explorativen Freiräume erfordert in der Praxis ein hohes Mass an Flexibilität und Kommunikation. Neue Kooperationen müssen initiiert, Daten auf ihre Verwendbarkeit und Zugänglichkeit überprüft, Infrastrukturen aufgebaut und Technologien ausprobiert werden. All das aus unterschiedlichen Gründen gelegentlich auch in Zeiträumen, die kürzer sind als die gewöhnlichen bibliothekarischen Vorlauf- und Planungsphasen. Im Gegenzug bieten die so entstandenen Freiräume Gelegenheiten für das Austesten innovativer Lösungen, die in vielfältiger Form Mehrwerte schaffen, beispielsweise durch die Nachnutzung neuer Datenlayers und Tools oder auf sehr grundlegender Ebene durch den Wissenstransfer zwischen den beteiligten Partnerinnen und Partnern. Durch diese Kooperationen leisten die Digital Scholarship Services einen wichtigen Beitrag zur Umsetzung von Co-Creation in der bibliothekarischen Praxis.

Literaturverzeichnis

- British Library Labs, <<https://www.bl.uk/projects/british-library-labs>>, Stand 26.08.2022.
- Candela, Gustavo et al.: Reusing digital collections from GLAM institutions, in: *Journal of Information Science* 24, August 2020, S. 1–17. Online: <<https://doi.org/10.1177/0165551520950246>>.
- Delbeke, Maarten: About [the Graph] project, Graph – Text reuse in rare books, <<https://graph-rare-books.ethz.ch/#/about>>, Stand: 26.08.2022.
- DS3Lab: Data Sciences, Data Systems, & Data Services, <<https://ds3lab.inf.ethz.ch/>>, Stand: 26.08.2022.
- E-Periodica: Next Level Access, <<https://nla.e-periodica.ch>>, Stand: 26.08.2022.
- Flanders, Julia: The Productive Unease of 21st-century Digital Scholarship, in: *Digital Humanities Quarterly* 3 (3), 2009. Online: <<http://www.digitalhumanities.org/dhq/vol/3/3/000055/000055.html>>.
- Gasser, Michael; Wanger, Regina; Prada, Ismail: Wenn Algorithmen Zeitschriften lesen: Vom Mehrwert automatisierter Textanreicherung, in: *o-bib. Das offene Bibliotheksjournal* 5 (4), 2018, S. 181–192. Online: <<https://doi.org/10.5282/o-bib/2018H4S181-192>>.
- Graf, Nicole: «Wissen Sie mehr?» Erfolgsfaktoren beim Crowdsourcing: Ein Erfahrungsbericht des Bildarchivs der ETH-Bibliothek, Zürich, in: *Rundbrief Fotografie* 27 (1), 2020, S. 28–38. Online: <<https://doi.org/10.3929/ethz-b-000410818>>.
- Graf, Nicole: Alles unter Kontrolle? KI im Einsatz im Bildarchiv der ETH-Bibliothek, Folien der Präsentation am 8. Bibliothekskongress in Leipzig, 31. Mai 2022. Online: <<https://doi.org/10.3929/ethz-b-000551925>>, Stand: 26.08.2022.
- KBLAB, <<https://lab.kb.nl/>>, Stand 26.08.2022.
- Mahey, Mahendra et al.: Open a GLAM Lab. Digital Cultural Heritage Innovation Labs, Book Sprint, Doha, Qatar, 23.–27.09.2019. Online: <https://glamlabs.io/books/open-a-glam-lab/>, Stand: 26.08.2022

- Rausch, Johannes; Martinez, Octavio; Bissig, Fabian et al.: DocParser. Hierarchical Document Structure Parsing from Renderings, in: Proceedings of the AAAI Conference on Artificial Intelligence 35 (5), 2021, S. 4328–4338. Online: <<https://ojs.aaai.org/index.php/AAAI/article/view/16558>>. Stand: 26.08.2022.
- Thomas, William G., III: The Promise of the Digital Humanities and the Contested Nature of Digital Scholarship, in: A New Companion to Digital Humanities, Chichester 2015, S. 524–537. Online: <<https://doi.org/10.1002/9781118680605.ch36>>.
- Transkribus | KI-gestützte Handschrifterkennung, <<https://readcoop.eu/de/transkribus/>>, Stand: 26.08.2022.
- Organigramm der ETH-Bibliothek (1. Juni 2022), <https://ethz.ch/content/dam/ethz/associates/ethlibrary-dam/documents/Ueber-uns/Organigramme/Organigramm_20220601_web_d.pdf>, Stand: 26.08.2022.
- Seguin, Benoit: Method, Graph – Text reuse in rare books, <<https://graph-rare-books.ethz.ch/#/about>>, Stand: 26.08.2022.
- What is Digital Scholarship?, <<https://digitalscholarship.web.ox.ac.uk/what-digital-scholarship>>, Stand: 26.08.2022.
- Wiggins, Benjamin; Hennesy, Cody; Vetruba, Brian et al.: Digital Scholarship Programs in Practice, in: College & Research Libraries 83 (4), 2022, 01.07.2022. Online: <<https://doi.org/10.5860/crl.83.4.568>>.