

Begrenzte Ressourcen und qualitativ minderwertige Metadaten verschärfen diese Big-Data-Probleme zusätzlich. Dieses Defizit lässt sich unter der Überschrift „Big-Data-Probleme ohne Big Data“ zusammenfassen. Es ist deutlich geworden, dass es einer maßgeschneiderten Kombination aus neuen und vorhandenen Technologien bedarf, um ein robustes, flexibles und skalierbares System zu entwickeln.

Digitale Werkzeuge für die Zukunft

Die von uns entwickelte Plattform folgt den drei Grundprinzipien Bedienkomfort, Anpassbarkeit und Weiterverwendung. Die Nutzer*innen arbeiten mit einer vertrauten und intuitiven Suchmaschine, die es ihnen erlaubt, zwischen den Sammlungen hin- und her zu wechseln und in einem Transkript schnell die relevanten Passagen zu finden. Mit Weiterverwendung meinen wir die Möglichkeit, Daten außerhalb der von uns entwickelten Suchoberfläche zu nutzen. Das ganze System ist so gestaltet, dass Mikrosites, die unsere Daten nutzen, leicht an unser System anknüpfen können. Gemeinsam mit Kolleg*innen beim MPIWG haben wir schon früh auf die Stärken von Large Language Models gesetzt, um die Extraktion von Metadaten zu automatisieren und zu optimieren, und damit bereits einige vielversprechende Ergebnisse erzielt.



04 Filmfotos aus den Oral History Archives des ANCP am UCLA.

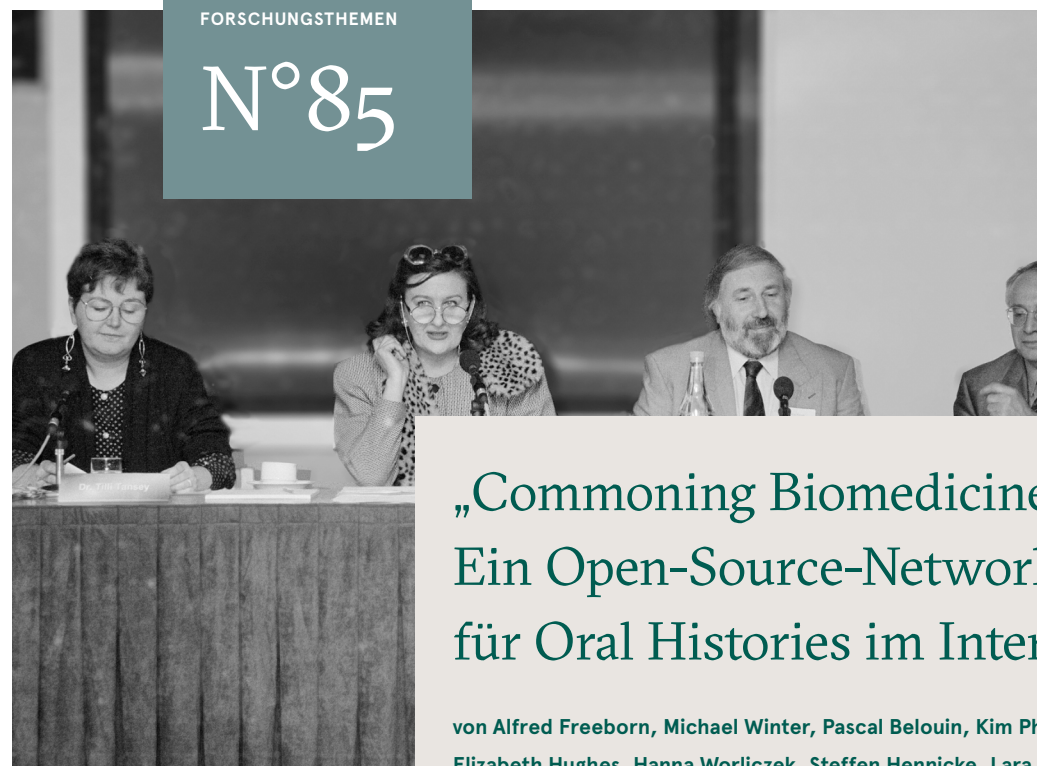
Nach dem Start einer Beta-Version der Plattform im Juli 2023 haben wir uns darauf konzentriert, immer mehr Sammlungen in das Netzwerk einzufügen und die Qualität der Metadaten zu verbessern. Grundsätzlich sind unsere Methoden generalisierbar, um Forschenden die Nutzung von und den Zugang zu digitalen Oral-History-Plattformen auch jenseits des Themas „Geschichte der Biomedizin“ zu erleichtern. Das ComBio-Projekt soll Forschende durch den Zugriff auf grundlegende Wissensressourcen dabei unterstützen, die Geschichte der Biomedizin besser zu schreiben. Doch auf einer anderen und womöglich grundsätzlicheren Ebene kann das Projekt als Form der wissenschaftlichen Zusammenarbeit im Bereich der Digital Humanities ein Umdenken darüber bewirken, wie wir Wissen im Sinne des Gemeinwohls erzeugen und bewahren wollen.

Das Projekt „Commoning Biomedicine“ ist eine Kooperation im Bereich Digital Humanities zwischen der Forschungsgruppe „Praktiken der Validierung in der Biomedizinischen Forschung“ und der Wissenschaftlichen IT-Gruppe des Max-Planck-Instituts für Wissenschaftsgeschichte, die zum Ziel hat, Oral Histories aus der Biomedizin als gemeinsame Wissensressource bereitzustellen.

Weitere Forschungsthemen sind auf der
Institutswebsite zugänglich:
www.mpiwg-berlin.mpg.de/de/forschungsthemen

MPI für Wissenschaftsgeschichte

Boltzmannstraße 22 · 14195 Berlin



„Commoning Biomedicine“: Ein Open-Source-Network für Oral Histories im Internet

von Alfred Freeborn, Michael Winter, Pascal Belouin, Kim Pham,
Elizabeth Hughes, Hanna Worliczek, Steffen Hennicke, Lara Keuck
DEZEMBER 2023

Mit Stand 2023 gibt es online etwa 4.000 englischsprachige Oral-History-Interviews mit Biomediziner*innen, Ärzt*innen, Pflegefachkräften, Patient*innen und Krankenhausmanager*innen. Ungefähr dieselbe Anzahl steht lediglich offline zur Verfügung, weil die Interviews entweder vertrauliche Informationen enthalten oder schlicht noch nicht digitalisiert worden sind. Die Durchführung, Bearbeitung und Sammlung der Interviews wird von unterschiedlichen Institutionen vorangetrieben. In den USA und Großbritannien zählen dazu einige der weltweit finanzstärksten biomedizinischen Stiftungen, aber auch öffentliche Bibliotheken, medizinische Fakultäten und Forschungseinrichtungen der Wissenschafts- und Medizingeschichte. Dieser schon jetzt beachtliche Korpus wächst stetig weiter mit der Aufzeichnung neuer und der Digitalisierung bestehender Sammlungen, die einzigartige Einblicke in die Geschichte der Medizin im 20. und 21. Jahrhundert bieten.

Mündliche Zeugnisse dieser Art bieten ausführliche persönliche Erfahrungsberichte zur Entwicklung von medizinischer Forschung und von Behandlungsmethoden sowie Informationen über Veranstaltungen, Quellen und Literatur, die andernorts nicht zu finden sind. Außerdem beleuchten sie die Perspektiven derer, die in der Medizingeschichte viel zu oft vernachlässigt werden. Als Wissensressourcen können sie einen Einfluss darauf nehmen, wie künftige Medizinhistoriker*innen die jüngere Vergangenheit bewerten.

Allerdings stoßen Forschende, die diese Quellen nutzen wollen, aktuell auf große Schwierigkeiten. Dabei ist der Zugang zu diesem Korpus das größte Problem. Zunächst einmal ist es nicht möglich, alle Sammlungen gleichzeitig zu durchsuchen. Die Interviews sind als unterschiedliche Medienformate bei verschiedenen Einrichtungen hinterlegt, die individuelle Metadaten- und Datenschutzbestimmungen zugrunde legen. Um zu ermitteln, welche Oral Histories überhaupt verfügbar sind, müssen die Forschenden in einem unstrukturierten und zeitaufwendigen Prozess etliche Websites durchforsten – weshalb die Wissenschafts-Community diese wertvollen Quellen häufig vernachlässigt. Darüber hinaus erschweren auch grundlegendere Einschränkungen den Zugang, wie zum Beispiel gesetzliche Bestimmungen darüber, wer diese sehr persönlichen Dokumente verwahren, einsehen und zitieren darf. Und schließlich besteht häufig die Gefahr, dass diese digitalen Quellen mit dem Abschalten von Projekt-Websites nicht mehr zur Verfügung stehen. Doch in erster Linie stehen Forschende heute vor dem praktischen Problem, dass sie einfach nicht wissen, was es in diesem Bereich alles gibt. Um diesen Korpus für Forschende aus dem weiten Feld der Medical Humanities als gemeinsame Ressource zu erschließen, entwickeln wir eine Open-Source-Website-Infrastruktur, um Oral-History-Bestände miteinander zu vernetzen. Das Projekt „Commoning Biomedicine“ ist eine Kooperation im Bereich Digital Humanities zwischen der Forschungsgruppe „Praktiken der Validierung in der Biomedizinischen Forschung“ unter Leitung von Lara Keuck und der Wissenschaftlichen IT-Gruppe des Max-Planck-Instituts für Wissenschaftsgeschichte, die zum Ziel hat, Oral Histories aus der Biomedizin als gemeinsame Wissensressource bereitzustellen.

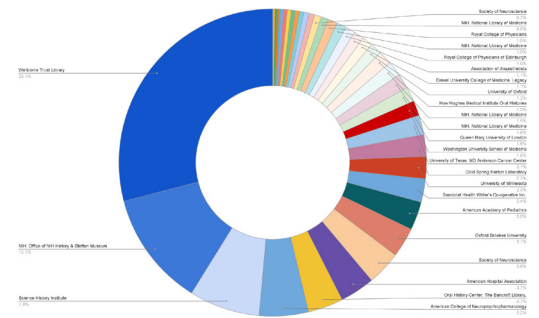


01 Sam McQueen, Unsplash.com

Commoning wissenschaftlicher Ressourcen: Collaboration, Conversations und Code

Was bedeutet „Commoning“? Der Begriff geht auf historische Auseinandersetzungen um Landeinhengungen im England der frühen Neuzeit zurück. In jüngerer Zeit wird er jedoch im Zusammenhang mit nachhaltigen Formen des Managements ökologischer und digitaler Ressourcen verwendet. Wir nutzen den Begriff, um das Engagement für den Einsatz digitaler Open-Source-Tools beim Aufbau zukunftsfähiger gemeinsamer Wissensressourcen zu beschreiben. In diesem Zusammenhang unterteilt sich das Commoning in die so genannten 3 Cs: **Collaboration**, **Conversations** und **Code**. Um Oral-History-Sammlungen dauerhaft besser zugänglich zu machen, reicht es nicht aus, Suchwerkzeuge zu entwickeln. Wir müssen auch die Zusammenarbeit (**Collaboration**) zwischen „Oral Historians“ und Archivar*innen verbessern. Beispielsweise arbeiten wir mit dem Science History Institute in Philadelphia zusammen, das eines der weltweit größten Online-

Archive für Oral-History-Sammlungen aus Wissenschaft und Medizin verwaltet. Wir veranstalten regelmäßig **Conversations** genannte Workshops, um Kontakte zwischen Forschenden aufzubauen, die mit digitalen mündlichen Zeugnissen arbeiten, und um die Wissenschafts-Community zu stärken. Im Mittelpunkt unserer Arbeit steht jedoch die Entwicklung von Open-Source-**Code**, um Online-Sammlungen von Wissensressourcen miteinander zu vernetzen.

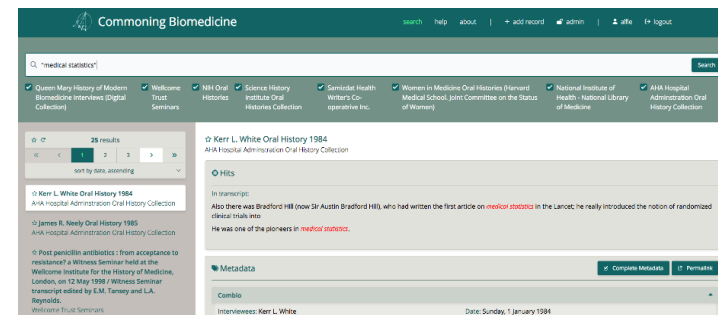


02 Diagramm zur Verteilung der Oral-History-Sammlungen im Bereich Biomedizin in den verschiedenen Institutionen, nach einer Übersicht über die Sammlungen aus dem Jahr 2022. Quelle: Alfred Freöborn.

Big-Data-Probleme ohne Big Data

Die Vernetzung dezentraler Oral-History-Sammlungen ist mit zahlreichen Herausforderungen verbunden, die häufig auch bei Big-Data-Projekten auftreten: die Föderation und Integration von Datenbanken, die Formalisierung von Schemata und Ontologien sowie die Bereinigung und Normalisierung, Skalierung und Sicherheit von Daten. Weniger technisch ausgedrückt geht es darum, Informationen aus vielen verschiedenen Datenbanken zu erfassen und miteinander zu integrieren, ohne die Originaldaten in einer einzigen zentralen Datenbank zusammenzufassen, und die Daten anschließend zu standardisieren, damit Nutzer*innen eine übergreifende Suche in den verschiedenen Datenbanken durchführen können.

Zu Projektbeginn haben wir eine umfassende Liste von Oral-History-Sammlungen im Bereich Biomedizin (sowohl online als auch offline) erstellt und verschiedene derzeit gängige Oral-History-Webframeworks getestet. Bei dieser Untersuchung trat ein Problem zutage, das bei zahlreichen Digital-Humanities-Projekten zu beobachten ist: das Datenmaterial war nicht umfangreich genug. Der aktuelle Oral-History-Bestand im Bereich Biomedizin fällt gemessen an der bei Big-Data-Projekten üblichen Datenmenge eher gering aus. Zudem enthalten Oral-History-Sammlungen häufig nur wenige und manchmal keine Metadaten – wie beispielsweise Angaben zum Zeitpunkt und Ort des Interviews und zu den Namen der Beteiligten.



03 Screenshot der Beta-Version der ComBio-Plattform.