**027.7 Zeitschrift für Bibliothekskultur / Journal for Library Culture •
Digital Humanities und wissenschaftliche Bibliotheken**

Ein großer Berg Daten? Zur bibliothekswissenschaftlichen Dimension des korpusliteraturwissenschaftlichen Digital Humanities- Projekts „High Mountains – Deutschschweizer Erzählliteratur 1880–1930“

J. Berenike Herrmann, Giulia Grisot, Susanne Gubser, Elias Kreyenbühl

Published on: Apr 20, 2021

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

ABSTRACT

Der Beitrag schildert die intensive Zusammenarbeit zwischen Fachwissenschaftler:innen und der Universitätsbibliothek Basel in einem korpusliteraturwissenschaftlichen DH-Projekt. Mit dem Auftrag zur Erstellung eines möglichst umfassenden Metadatensatzes Deutschschweizer Werke eines bestimmten Zeitraums mit definierten Parametern und der daran anschliessenden Retrodigitalisierung ausgewählter Objekte übernahm die Bibliothek die Aufgabe, eine valide Datengrundlage für das Forschungsprojekt zu schaffen. Erreichter Meilenstein im Forschungsprojekt war die erstmalige datenbasierte Schätzung des Umfangs fiktionaler Erzählliteratur 1880–1930 von Deutschschweizer Autor:innen (7'000–10'000 Titel). Im Beitrag werden nicht nur die Methoden zur Erhebung der benötigten Daten beschrieben, sondern in der Retrospektive das Vorgehen auch kritisch hinterfragt, um abschliessend Lessons Learned und Desiderata bei der Umsetzung solcher Projekte in Kooperation mit Bibliotheken zu formulieren.

This article describes the intensive cooperation between researchers and the University Library of Basel in a corpus literary studies DH project. With the task of creating a comprehensive metadata set of Swiss-German works of a certain period with defined parameters and the subsequent retro-digitisation of selected objects, the library took on the task of creating a valid meta-data foundation for the research project. The milestone achieved in the research project was the first data-based estimate of the volume of fictional narrative literature 1880–1930 by Swiss-German authors (7,000–10,000 titles). The article not only describes the methods used to collect the necessary data, but also critically examines the procedure in retrospect in order to formulate lessons learned and desiderata for the implementation of such projects in cooperation with libraries.

1. Einleitung

*Er treibt den trägen Schwarm von schwer=beleibten Kühen,
Mit freudigem Gebrüll, sich im bethauten Steg,
Sie irren langsam um, wo Klee und Muttern blühen,
Und mäh'n das zarte Gras mit scharfen Zungen weg:
Er aber setzt sich bey einem Wasser=Falle,
Und ruft mit seinem Horn dem lauten Widerhalle.*

[\(Haller, 1762, V. 185ff\)](#)

In seinem Lehrgedicht *Die Alpen* zeichnet der Frühaufklärer Albrecht von Haller 1729 ein neuerdings *positiv* getöntes Bild der Gebirgskette. Mit Haller beginnt hier eine literarische Tradition, die die ehemals *unberechenbaren* Berge nun als „schönen“ *Naturraum* enkodiert, der, oftmals von *urbanen* oder gar industriell geprägten Topographien abgegrenzt, unter anderem zur positiven Bezugsgröße für nationale Identität jenseits der innerhelvetischen Sprachgrenzen wird.

Aber auch jenseits der Alpen und vordergründiger Angebote für nationale Identitätsbildung gilt, dass Literatur stets fiktionale Räume gestaltet und diese affektiv enkodiert. Dieser Zusammenhang von Gefühl und dargestelltem Raum wird mit digitalen Methoden des „High Mountains“-Projekts untersucht¹, das mittels der *Big Data* der digitalen Kataloge, Open Linked Data und digitalen Textkorpora ganz neue Forschungsfragen und alte Forschungsfragen neu stellt.² Im Folgenden stellen wir als *Use Case* die Zusammenarbeit zwischen Digitaler Literaturwissenschaft und der Universitätsbibliothek Basel (UB Basel) als „Datenlieferantin“ vor.³

Die „Korpusliteraturwissenschaft“ braucht als datengetriebene und datenintensive Wissenschaft *hochskalierte*, aber auch *belastbare* Daten ([Herrmann & Lauer, 2018](#)). Dabei geht es um vielfältige Informationstypen. Neben den digitalisierten Volltexten selbst stehen die Metadaten – Informationen über Autor:in mit Gender, Nationalität und Lebensdaten, und über die Texte, wie Gattung, Sprache, Übersetzung, sowie Einträge zu Publikations- und die Editions-geschichte. In unserem Fall geht es um Romane, Novellen und Erzählungen, die ursprünglich von 1880-1930 auf Deutsch erschienen sind und Schweizer Autorinstanzen zugeordnet werden können.

Die datenintensive Literaturwissenschaft hat einen statistischen Blick. Sie muss dem Anspruch genügen, sich auf die Kenntnis der „Grundgesamtheit“ der publizierten literarischen Texte eines Zeitraums zu beziehen um die Reichweite ihrer Aussagen einordnen zu können. An dieser Stelle, wo Literaturwissenschaft eben auch *Sammelforschung* (ein alter Name der Statistik) ist, werden die wissenschaftlichen Bibliotheken in der Arbeit mit Datenbeständen zu unerlässlichen Partnern auf Augenhöhe ([Meister & Veit, 2014](#)). Als Gedächtnisinstitutionen erweitern sie ihren Auftrag vom Bewahren und Nutzbarmachen ihrer Sammlungen um das aktive Anbieten von Metadaten. Auf diese Weise wird digitale geisteswissenschaftliche Forschung insbesondere im Big-Data-Kontext überhaupt erst ermöglicht ([Döhl, 2019](#)).

Im Rahmen unserer Zusammenarbeit waren es vor allem zwei Anfragen, die vom Forschungsprojekt „High Mountains“ an die UB Basel gestellt wurden: (a.) Die

Erstellung eines möglichst umfassenden Metadatenatzes Deutschschweizer Werke des beschriebenen Zeitraums mit den genannten Parametern und
(b.) die daran anschließende Retrodigitalisierung ausgewählter, bislang nur physisch vorliegender Objekte.

Im Folgenden möchten wir vor allem unsere Erfahrung zum aktiven Anbieten von Informationen in Bezug auf (a.) beschreiben: Ausgehend von der Annahme, dass ein Metadatenabzug der Deutschschweizer Literatur aus einem Bibliothekskatalog möglich sei, setzte rasch Ernüchterung ein: Weder gibt es eine umfassende Bibliographie, auf die Bezug genommen werden könnte, noch hält der Bibliothekskatalog der UB Basel dieser Frage stand. Selbst eine Linked Data-Abfrage der Gemeinsamen Normdatei (GND)⁴ konnte keine direkt verwertbare Liste von Deutschschweizer Autor:innen liefern. Um die Informationen anzubieten, folgte ein langer Weg, der über automatische Abfragen sowie manuelle und automatische Datenanreicherung führte. Zwischenetappen waren die händische Erstellung einer *philologischen Liste* und das Problem, anhand der Metadaten aus den GND-Autoritätsdateien den biographischen Bezug von Autor:innen als „Deutschschweizer:innen“ festzustellen. Das Ziel der Gesamtsicht in Form einer umfassenden Datenbank, die philologische und informationswissenschaftliche Informationstypen vereint, sowie idealerweise die Volltexte selbst vorhält, ist derzeit noch nicht erreicht. Es liegen jedoch erstmals umfassende und bereinigte Datensätze (als .csv-Tabellen) vor.

2. „Gebt uns eine Liste“. Zwischen Metadatenabzug und intellektueller Erstellung einer Bibliographie

Für das Forschungsprojekt sollte zunächst der Umfang und die Zusammensetzung der Deutschschweizer Erzählliteratur (Romane, Novellen, Erzählungen) der Jahre 1880–1930 eingeschätzt werden. Daher wurde ursprünglich ein Metadatenabzug aus dem Bibliothekskatalog bestellt („Liste deutschsprachiger CH-Belletristik“, #1, siehe Tab. 1). Da dieser Auftrag mangels vorgehaltener Metadaten nicht möglich war, wurde die Fachreferentin für Germanistik beauftragt, unter den gegebenen Bedingungen eine möglichst vollständige Bibliographie zu erstellen. Sowohl kanonische wie auch nicht kanonische Werke sollten Eingang finden und als solche ausgewiesen werden (ebenso wie bestimmte Textlängen).

Arbeitsschritt	Ressourcen	Möglich?	Ansprechpartner:in	Umfang des bearbeiteten Datensatzes
#1 Liste CH-DE-Werke	Bibliothekskatalog der Basler UB	nein	Fachreferentin Susanne Gubser	---
#2 Liste CH-DE-AutorInnen	GND-Abzug bei DNB	ja, mit Abstrichen: false positives (bes. Publikationsdatum), false negatives	Koordinatorin GND-Redaktion Esther Ernst, Systembibliothekar IDS Basel/Bern Basil Marti, Fachreferentin	N = 1'857 AutorInnen
#3 "Philologische Liste" (CH-DE-Werke und CH-DE-AutorInnen)	Literaturgeschichten, Nachschlagewerke, Bibliographien usw. (physisch / elektronisch)	ja, aber: zeitintensiv, nur Auswahl möglich; unklare Fälle (Werkgattung)	Fachreferentin	N = 281 AutorInnen
#4.1 Erweiterte Liste CH-DE-AutorInnen (merge #2 + #3, Anreicherung LOD)	LOD Schnittstellen (VIAF, Wikidata, GND lobid)	ja, aber: Überprüfung und LOD für einige AutorInnen aus #3 nötig (n= 49)	Data Librarian Jonas Waerber	N = 1'900 AutorInnen
#4.2 Erweiterte Liste CH-DE-Werke (merge #2 + #3, Anreicherung LOD)	LOD Schnittstellen (VIAF, Wikidata, swissbib)	ja, aber: weiteres Clean-Up nötig	Data Librarian und Postdoc Giulia Grisot	N = 25'410 Werke nach erster Bereinigung n = 7'963 Werke für "1880"- "1930" n = 868 AutorInnen für "1880"- "1930" n = 2'184 NAs Feld "Datum"
#5 Liste erhältlicher CH-DE-Volltexte für #4	Gutenberg.de, DTA; textgrid; wikisource, etc.; ja KOLIMO		Data Librarian und Postdoc	in Bearbeitung

Tab. 1: Stationen des Projekts „Index CH-DE-Belletristik 1880–1930“

Hierzu gab es für die *European Literary Text Collection* ([Odebrecht, Burnard, Eder, & Schöch, 2019](#); [Herrmann, Odebrecht, Santos, & Francois, 2020](#)), aber auch spezifisch für das „High Mountains“-Projekt diverse Vorgaben. Da nicht viel Zeit zu Verfügung stand, lag der Fokus von Anfang an auf elektronischen Möglichkeiten der Bibliographie-Erstellung und pragmatischen Vorgehensweisen. Tab. 1 zeigt die aufeinander aufbauenden, bzw. sich auseinander ergebenden, Arbeitsschritte zur Erstellung eines Gesamt-Metadatenatzes. Aufgeführt sind die jeweiligen digitalen oder physischen Ressourcen, die Angabe, inwieweit der avisierte Arbeitsschritt erfolgreich durchgeführt werden konnte, der/die Ansprechpartner:in, sowie gegebenenfalls der Umfang des jeweiligen Datensatzes.

2.1 „Schöne“ Literatur vom Katalog?

Die Frage nach einem Datenabzug der Belletristik aus dem Bibliothekskatalog der UB Basel musste die Fachreferentin abschlägig beantworten. Belletristik wird erst seit 2017 durch eine entsprechende Codierung markiert.⁵ Eine Abfrage nach Belletristik hätte nur über die Namen bereits bekannter Autor:innen oder über eine Stichwortsuche nach Gattungsbegriffen wie „Roman“ oder „Novelle“ (wie sie öfter im Untertitel vorliegen) erfolgen können und nur sehr unvollständige Ergebnisse geliefert. Ähnlich stellt sich die Situation in anderen (Schweizer) Bibliothekskatalogen dar.⁶

Da kein Metadatenabzug möglich war, wurde zunächst geklärt, ob bereits eine Bibliographie der sogenannten „schönen Literatur“ (sprich: der literarischen Werke, Belletristik) der *deutschsprachigen* Schweiz existiert – dies ist jedoch nicht der Fall, weder elektronisch noch gedruckt. Eine Bibliographie der Schweiz gibt es zwar seit 1878⁷, die Jahrgänge bis 2006 liegen jedoch nur gedruckt vor und wurden erst ab 1916 nach Fachgebieten gegliedert (inklusive einer Abteilung „Schöne Literatur“, die jedoch nicht nur die deutschsprachige Belletristik enthält und auch nicht ausschliesslich die Belletristik von Schweizer Autor:innen⁸). Eine Auswertung dieser Hefte wurde daher als zu zeitintensiv verworfen. Kurz wurde die Digitalisierung der Bibliographie erwogen, aber auch dies hätte noch einiges an Bearbeitung nach sich gezogen.

Mit dem Projektteam und Kolleg:innen aus verschiedenen Bereichen der UB Basel wurden weitere Möglichkeiten eines unabhängigen elektronischen Datenabzugs für eine Deutschschweizer Autor:innen- bzw. Werkliste besprochen. Erwogen wurde ein Datenabzug via Linked-Open-Data-Schnittstellen wie Lobid und Linked Swissbib. Tests zeigten jedoch eine unzureichende Sensitivität und Spezifität für unsere Anfrage.⁹ In der Folge konzentrierten wir uns zunächst auf die GND.

2.2 Die GND-Liste: Autor:innen

Dank der Koordinatorin der Basler GND-Redaktion und dem Systembibliothekar IDS Basel/Bern konnte über den Metadatendienst der Deutschen Nationalbibliothek (DNB) ein Gesamtabzug der GND in MARC21-Format bezogen und in ein Tabellenformat (.csv) umgewandelt werden, um daraus eine erste, noch weit gefasste Autor:innenliste zu gewinnen (Arbeitsschritt #2). Kriterien waren der Ländercode („Schweiz“), der Beruf („Schriftsteller*“ und andere¹⁰) sowie Lebens- oder Wirkungsdaten (sehr breit gefasst: „18* OR 19*“).

Die Originalliste umfasste rund $N = 7'000$ Einträge. Sie wurde zunächst manuell von der Fachreferentin auf $n = 1'857$ Einträge reduziert, indem Personen mit nicht zum Wirkungszeitraum 1880–1930 passenden Lebensdaten entfernt wurden.¹¹

Anzumerken gilt, dass die Liste auch die französisch-, italienisch- und rätoromanischsprachigen Schweizer Autor:innen enthält, da die GND keine Codierung für die Schreibsprache von Autor:innen verwendet („false positives“). Daneben können zum Beispiel auch Verfasser:innen von ausschliesslich nicht-literarischen Texten oder reine Lyriker:innen enthalten sein, da die Berufsbezeichnungen nicht immer gleich konsistent in der gleichen Tiefe vergeben wurden.¹² So lieferten die ausgewählten

Berufsbezeichnungen¹³ tendenziell zu viele Treffer und werden – zusammen mit den doch breit gefassten Lebensdaten – noch einmal genauer überprüft. Umgekehrt sind auch keineswegs alle Schweizer Autor:innen auf der Liste enthalten („false negatives“). Wie spätere Stichproben ergaben, haben zwar sehr viele, aber bei weitem nicht alle Schweizer Autor:innen des Zeitraums bislang eine GND-Nummer erhalten (insbesondere im Bereich der nicht-kanonischen Autor:innen gibt es Lücken). Die Einträge weiterer Autor:innen wiederum sind nicht vollständig (das heisst Berufsbezeichnung, Ländercode oder Lebensdaten fehlen, bekommen im Datensatz ein „NA“)¹⁴ und konnten daher mit der durchgeführten Abfrage nicht gefunden werden.¹⁵

Mitextrahiert, wo vorhanden, wurde aus der GND das Feld „Geographischer Bezug“, das Geburts- und Sterbeort, sowie manchmal den Wirkungsort enthält. Von diesen Angaben wie auch den Ländercodes (oft liegen zu einer Person mehrere vor) lässt sich aber noch nicht zwingend auf den Sozialisations- beziehungsweise Wirkungsort schliessen. Gerade der Sozialisationsort der Autor:innen („Wo ist die Person aufgewachsen und zur Schule gegangen?“) wurde jedoch als massgebendes Kriterium für die Zuordnung von „Nationalität“ definiert. Auch über die Kanonizität der Autor:innen wird in dieser Liste nichts ausgesagt. Eine Nachbearbeitung der Metadaten, besonders von „Nationalität“ erwies sich daher für das Projekt als zwingend.

Zudem war die Autor:innenliste nur Mittel zum Zweck der Erstellung einer Werkliste. Daher galt es, ausgehend von den indizierten Autor:innen, die von selbigen verfassten und den Kriterien entsprechenden Werke aufzunehmen (Arbeitsschritt #4.2). Auf diese Datenanreicherung aus verschiedenen Quellen wird unter 2.4 eingegangen.

2.3 Die philologische Liste

Parallel zu diesen Arbeiten rund um elektronische Datenabzüge wurde von der Fachreferentin eine zweite Werkliste auf philologischer Basis begonnen (Arbeitsschritt #3). Ziel war die Erstellung einer möglichst detaillierten Bibliographie. Diese Grundlagenforschung war nötig, da bei den oben beschriebenen Datenabzügen Einschränkungen evident waren. Zudem sollten die Sensitivität und Spezifität der Datenabzüge eingeschätzt werden: Wie viele der relevanten Autor:innen und Werke werden gefunden, wie viele nicht relevante werden ausgeschlossen (bzw. fälschlich eingeschlossen)? Die aus unterschiedlichen Quellen erstellte philologische Liste umfasste schließlich N = 1'694 Werkeinträge von insgesamt N = 281 Autor:innen. Zunächst wurde hierfür eine Quellenbibliographie mit möglichen relevanten gedruckten wie elektronischen Quellen erstellt (Literaturgeschichten,

Nachschlagewerke, Bibliographien und Verzeichnisse, Anthologien, Schriftenreihen und Zeitschriften). Aus Zeitgründen konnten nur wenige dieser Quellen systematisch ausgewertet werden und so wurde eine die COST-ELTeC-Kriterien bedienende Priorisierung unter Berücksichtigung einschlägiger und ausgewählter spezieller Ressourcen vorgenommen. Die philologische Liste nähert sich also einer vollumfänglichen *ground truth* auch erst einmal nur an, ist aber ein wichtiger Schritt auf diese zu.

2.3.1 Gedruckte Quellen

Als erstes wurden drei Literaturgeschichten, zwei moderne ([Petzold, 2007](#); [Rusterholz & Solbach, 2007](#)) und eine zeitgenössische ([Nadler, 1932](#)), sowie ein zeitgenössisches Lexikon ([Aellen, 1918](#)) ausgewertet. Petzold und Rusterholz sollten dabei die auch heute noch bekannten Namen (Stichwort *Kanonizität*) liefern, Nadler und Aellen diejenigen der von den Zeitgenossen beachteten Autor:innen. Zusätzliche Werktitel zu den hier gefundenen Namen wurden von der Fachreferentin jeweils aus verschiedenen elektronischen Quellen ergänzt (Bibliothekskatalog, Wikipedia oder *Deutsches Literatur-Lexikon*).

Einen Glücksfall stellt die Spezialbibliographie „Deutschsprachige Schriftstellerinnen der Schweiz 1700–1945“ ([Stump, Widmer, & Wyss, 1994](#)) dar. In den Literaturgeschichten wie auch bei Aellen kommen die Frauen eindeutig zu kurz. Viele werden gar nicht erwähnt, andere nur in knappen Exkursen zu „Unterhaltungs-“ beziehungsweise „Kinderliteratur“ abgehandelt. Für diese Bibliographie jedoch wurden die literarischen Werke von Deutschschweizer Frauen minutiös recherchiert.¹⁶

Weitere gedruckte Quellen waren weniger ergiebig oder ihre Bearbeitung musste als zu zeitintensiv verworfen werden, so zum Beispiel bei den drei Buchhandelsverzeichnissen [Schweizerischer Buchhändlerverein & Verein Schweizerischer Verlagsbuchhändler](#) (1914), [Schweizerischer Buchhändlerverein & Verein Schweizerischer Verlagsbuchhändler](#) (1931) und [Schweizerischer Buchhändlerverein & Verein Schweizerischer Verlagsbuchhändler](#) (1939). Einerseits enthalten diese Verzeichnisse auch in Schweizer Verlagen erschienene Werke ausländischer Autor:innen sowie Übersetzungen ins Deutsche, andererseits sind auch Neuauflagen nicht immer auf den ersten Blick als solche erkennbar. Hier wäre einiges an Rechercharbeit angefallen. Umgekehrt fehlen in den Schweizer Buchhandelsverzeichnissen die in ausländischen Verlagen publizierten Werke von Schweizer Autor:innen. Insgesamt wären diese drei Verzeichnisse aber schneller auszuwerten gewesen als die Schweizer Bibliographie (die allerdings im Gegenzug

auch die im Ausland erschienenen Werke Schweizer Autor:innen enthält).

Insbesondere für den Bereich der nicht-kanonischen Werke hätten sie sich vermutlich als ergiebig erwiesen.

2.3.2 Digitale Quellen

Während bei Bibliographien und Verzeichnissen meist Vollständigkeit angestrebt wird, nehmen Quellen wie Literaturgeschichten und Nachschlagewerke mit der getroffenen Auswahl und Gewichtung gleichzeitig eine Wertung vor. Dies gilt natürlich nicht nur für die gedruckten Nachschlagewerke, sondern auch für ihre elektronischen Pendants. Zudem sind bei den elektronischen Quellen meist weitere Schwierigkeiten zu berücksichtigen. Die frei zugänglichen Nachschlagewerke *Historisches Lexikon der Schweiz* und *Wikipedia*¹⁷ wurden verwendet, um im ersten philologischen Zugriff die Quellenlage für deutschsprachige Schweizer Autor:innen zu sichern. Diese Arbeit lieferte zwar auf Anhieb je eine hilfreiche Seite mit namentlichen Einträgen zur Deutschschweizer Literatur, Stichproben zeigten jedoch, dass viele Autor:innen aus den Nachschlagewerken nicht über den thematischen Sucheinstieg auffindbar sind. Zudem fehlen viele Autor:innen und die Detailliertheit der Werkverzeichnisse der enthaltenen Autor:innen variieren sehr.¹⁸

Als Enttäuschung erwiesen sich in diesem Zusammenhang insbesondere diverse lizenzierte elektronische Nachschlagewerke. Eine Abfrage nach deutschsprachigen Schweizer Autor:innen war entweder nicht wirklich ergiebig, da die Schweiz nur am Rande berücksichtigt wird (zum Beispiel *Kritisches Lexikon zur deutschsprachigen Gegenwartsliteratur*), oder nicht möglich, da die erforderlichen Suchkriterien entweder nicht erfasst oder nicht gezielt recherchierbar sind (zum Beispiel *Killy Literaturlexikon*, *Deutsches Literatur Lexikon Online*). So weist etwa das *Deutsche Literatur Lexikon Online (DLLO)*, welches die beiden umfangreichen gedruckten Nachschlagewerke *Deutsches Literatur Lexikon (DLL)* und *Deutsches Literatur Lexikon. Das 20. Jahrhundert* in einer Datenbank vereint, Schweizer Autor:innen sehr umfassend nach. Auch ist das Schriftenverzeichnis zu den Personeneinträgen jeweils ausführlich und enthält auch häufig Angaben zur Gattung eines Werkes. Leider wurden jedoch die Inhalte der gedruckten Bände für die Datenbank nicht standardisiert und aufbereitet¹⁹, was diese für eine komplexe Recherche leider unbrauchbar macht. Auch die elektronische Version des *DLL* konnte also faktisch nur als händisches Nachschlagewerk zur Ergänzung von Werktiteln genutzt werden.

2.3.3 Regionen

Da sich im Verlauf des Projekts ein Fokus auf die Erzählliteratur der Regionen Basel, Bern und Zürich herauskristallisierte, wurden abschliessend zwei Verzeichnisse regionaler Belletristik ausgewertet: das Basler Literarische Archiv (BLA) und die Literapedia Bern. Die Werktitel beider Verzeichnisse sind im Bibliothekskatalog codiert²⁰ und konnten als Liste exportiert werden. Die Einträge des Basler Literarischen Archivs wurden systematisch überprüft und die extrahierten Werklisten durch weitere Titel aus anderen Quellen ergänzt, da das Verzeichnis insbesondere Werke vor 1915 nur punktuell nachweist. Für den Raum Zürich gibt es leider kein Pendant, wie eine Nachfrage an der Zentralbibliothek ergeben hat. Es müssten dafür die lokalen Bibliographien ausgewertet werden, die bisher nur gedruckt existieren.²¹

Wo bekannt, wurden in der philologischen Liste Wirkungsort und Werkgattung verzeichnet. Insbesondere die *Gattung* eines Werkes ist im Projekt ein entscheidendes Metadatum. Gattung wird entweder im Paratext (Untertitel) des Werkes selbst erwähnt oder von Autoritäten wie den erwähnten Literaturgeschichten und Nachschlagewerken zugeschrieben. Dass diese Genreklassifizierung letztlich hoch variabel ist (vgl. z.B. [Underwood, 2019](#), Kapitel 2), ist eine bekannte literaturhistorische und -theoretische Problematik, die wir pragmatisch lösten, indem wir, wo verfügbar, die unterschiedlichen Informationstypen zu Gattung in den Datensatz aufnahmen. Häufig fehlen jedoch Gattungs-Angaben („NA“-Einträge), was für unseren Datensatz bedeutet, dass Gattung oftmals nur autoptisch²² festgestellt und händisch in den Datensatz eingefügt wird.

Daneben wurden für die Einschätzung der Kanonizität nützliche Hinweise zu Personen und Werken in die Liste aufgenommen (wie zum Beispiel Hinweise auf mehrere Auflagen oder Auszeichnungen).²³ Der Bereich der kanonischen Literatur ist durch unser Vorgehen, das sich auf einschlägige Kompendien beruft, insgesamt gut abgedeckt. Der Bereich der nicht-kanonischen Werke stellt sich schwieriger dar, nicht zuletzt, weil hier oftmals bereichsspezifisches Expertenwissen intellektuell erhoben werden muss. Nicht-kanonische Werke wurden von uns vor allem in Bezug auf die Literatur von Frauen, sowie die regionale Literatur aus Basel und Bern erhoben. Für eine weitere Anreicherung wäre die Auswertung der Schweizer Bibliographien und Buchhandelsverzeichnisse interessant.

2.4 Maschinelle Erweiterung der Datenbasis

Ausgehend von der oben beschriebenen #2 GND-Autor:innenliste und der #3 philologischen Liste wurde nun das Ziel einer möglichst umfänglichen Metadatenbank (Erhebung der Grundgesamtheit) mit maschinellen Methoden weiterverfolgt. Zunächst sollten die Autor:innen-Einträge mit möglichst vielen relevanten Merkmalen und Identifikatoren angereichert werden (2.4.1), um dann die Werktitel der Autor:innen aufzufinden und als Metadaten abzulegen (2.4.2).

2.4.1 Anreicherung der Autor:innen-Liste

Da die philologische Liste (Schritt #2) von Hand erstellt wurde (siehe 2.3), mussten die Autor:innen zunächst mit der GND-„Personen ID“ (kurz „GND-Nummer“) ergänzt werden. Diese dient als zentraler Identifikator gewissermaßen als Anker, um die Liste um weitere Identifikatoren und Metadaten anzureichern. Das Auffinden der GND-Nummer für unsere philologische Liste wird durch den Reconciliation-Service von OpenRefine²⁴ erfreulich vereinfacht. Auf Anhieb konnten zwei Drittel der $N = 281$ Autor:innen anhand der Namen zugeordnet und angereicht werden ($n = 123$ Autor:innen in beiden Datensätzen). In einem zweiten Schritt wurden die unsicheren Treffer intellektuell überprüft und sichere Treffer ausgewählt.

Nun konnten in Arbeitsschritt #4.1 die beiden Datensätze (#2 GND-Autor:innenliste; die Autor:innen der #3 philologischen Liste) miteinander abgeglichen werden. Diese neue, fusionierte Autor:innenliste umfasste nun ca. $N = 1'900$ Einträge und bildete die Basis für eine ganze Reihe weiterer Anreicherungsprozesse. In dieser Liste werden zum gegenwärtigen Zeitpunkt einige Autor:innen aus der ursprünglichen philologischen Liste nicht berücksichtigt ($n = 49$). Diese müssen in einem nächsten Schritt auf die Verfügbarkeit der GND-ID überprüft werden und – samt Werken – in die LOD-Anreicherung einbezogen werden.

Bei der Anreicherung wurde die Strategie verfolgt, möglichst viele Datenquellen nacheinander auszuschöpfen, um zu einem immer kompletteren Datenset zu gelangen. Um diese Abfragen zu erleichtern, wurde die Liste zunächst mit weiteren Autor:innen-IDs angereichert (siehe Tab. 2). Über die Datenschnittstelle von swissbib konnten die zugehörigen VIAF²⁵- und Wikidata²⁶-Identifikatoren für die Autor:innen gezogen werden.²⁷ Unsere lokale Quellenangabe zeigt, ob ein Item *sowohl* in der GND-Liste und der philologischen Liste („sgu#gnd“) oder bislang *nur* in der philologischen Liste („sgu“) vorhanden war.

116009128	https://data.swissbib.ch/person/5e533a1-aaaa-30b0-84b0-3e017a20c3b5	98456177	Q1146857	squl#grnd	Aellen, Hermann	1887-05-24	Oberbalm	4535846-2	1929-09-26	Münster	453748-5	M	Schriftsteller#Journalist
116294272	https://data.swissbib.ch/person/5b32c50-6ad3-3a6b-93b8-975415049b65	27818259	sgu		Altherr, Paul	1887			1959			M	
136594619	https://data.swissbib.ch/person/5b26231e-7ea8-3a6d-9d8a-082a94e4c3d5	281096300	Q57338497	squl#grnd	Altherr, Alfred	1843-03-14	Grub (Appenzel, Aussenroden)	4543206-9	1918-01-18	Basel	4004817-5	M	Planer#Schriftsteller#Theologie stud #Theologie
139437819	https://data.swissbib.ch/person/5a22a24-69b8-3a71-a04e-e6423d8e4e02	262323012	Q63486299	squl#grnd	Amberger, Olga	1882-08-08	Zürich	4068038-1	1970-08-21	Zürich	4068038-1	W	Schriftstellerin#Schriftstellerin#Journalist#Mal
106697992	https://data.swissbib.ch/person/7b729a07-6074-3a6d-9e46-413d9d8d5b37	245440723	Q73602809	squl#grnd	Annatz, Ulrich	1879-01-09	Bräuslingen	4749374-4	1934-07-18	Bern	4005762-8	M	Schriftsteller
117763047	https://data.swissbib.ch/person/51566e07-23ac-3a47-b0ac-3a2c3d468173	22927224		squl#grnd	Annheim, Theodor	1891-05-29						M	Schriftsteller
102526671	https://data.swissbib.ch/person/3c58c99e-51ca-30b2-9534-08be779e5d57	254914642	Q1423263	squl#grnd	Anner, Hedwig	1888-02-05	Thun	4059991-7	1969-05-08	Givis	4453844-3	W	Dr. phil akad #Folklorist#Schriftstellerin
10692349	https://data.swissbib.ch/person/06e6d2e-617b-3a6d-9d8a-082a94e4c3d5	15147491	Q1294389	squl#grnd	Annet, Edwin	1901-05-11	Zürich	4068038-1	1962-11-27	Zürich	4068038-1	M	Journalist#Redakteur#Schriftsteller
116373817	https://data.swissbib.ch/person/5b26231e-7ea8-3a6d-9d8a-082a94e4c3d5	762201	Q360786	squl#grnd	Attenhofer, Adolf	1879-05-14	Chur	4010189-7	1950-12-24	Zürich	4068038-1	M	Schriftsteller#Endologe
119290804	https://data.swissbib.ch/person/5b26231e-7ea8-3a6d-9d8a-082a94e4c3d5	30226278	Q146037	squl#grnd	Auer, Gerthe	1871-06-25	Wien	4066009-6	1940-07-16	Berlin	4005728-8	W	Schriftstellerin
117764574	https://data.swissbib.ch/person/5b26231e-7ea8-3a6d-9d8a-082a94e4c3d5	67251112	Q1283562	squl#grnd	Bach, Marie Luise	1836-06-26	Bern	4005762-8	1904-08-25			W	Schriftstellerin#Lehrerin
102500485	https://data.swissbib.ch/person/5b26231e-7ea8-3a6d-9d8a-082a94e4c3d5	247469081	squl#grnd		Bachthor-Bucktorf, Anna	1839-06-10	Bern	4005762-8	1909-02-02	Männedorf	4372043-2	W	Autoren
125964065	https://data.swissbib.ch/person/5b26231e-7ea8-3a6d-9d8a-082a94e4c3d5	28049628	Q15436429	squl#grnd	Baerwart, Theobald	1872			1942			M	Schriftsteller
104672765	https://data.swissbib.ch/person/5b26231e-7ea8-3a6d-9d8a-082a94e4c3d5	74293331	Q1357452	squl#grnd	Batzli, Ernst	1902-04-10	Bolligen	4007613-1	1959-01-03	Bolligen	4007613-1	M	Schriftsteller#Mundartschriftsteller#Lehrer
104673117	https://data.swissbib.ch/person/5b26231e-7ea8-3a6d-9d8a-082a94e4c3d5	305927707	sgu		Batzli, Ida	1856			1927			W	
1047349752	https://data.swissbib.ch/person/5b26231e-7ea8-3a6d-9d8a-082a94e4c3d5	306290794	sgu		Baudenbacher, Emil	1874	Murten	4075106-3	1921			M	
102754486	https://data.swissbib.ch/person/5b26231e-7ea8-3a6d-9d8a-082a94e4c3d5	298459922	squl#grnd		Beerli, Hans	1892	Sankt Gallen	4051594-1	1957			M	Jurist#Schriftsteller#Journalist#Doktor akad
116109040	https://data.swissbib.ch/person/5b26231e-7ea8-3a6d-9d8a-082a94e4c3d5	8130824	Q19126011	squl#grnd	Bentschen, Alfred	1864			1924			M	Schriftsteller#Redakteur#Organist
13631015X	https://data.swissbib.ch/person/5b26231e-7ea8-3a6d-9d8a-082a94e4c3d5	39754967	Q5739582	squl#grnd	Bethens, Eduard	1884-04-20	Biel	4006509-1	1944-11-14	Locarno	4036106-8	M	Schriftsteller#Dramatiker#Journalist
116136808	https://data.swissbib.ch/person/7943b207-794e-3a6d-9d8a-082a94e4c3d5	64750310	Q1895904	sgu	Berlebach, Gosiwina von	1845-09-25	Erfurt	4015240-6	1916-04-10	Wien	4066009-6	W	
116828649	https://data.swissbib.ch/person/02e6f8a-3631-3118-8157-ca786d15d0	39425134	Q124698	squl#grnd	Bernoulli, Carl Albrecht	1905-01-10	Basel	4004817-5	1927-02-13	Arlesheim	4002929-3	M	Theologe#Kirchenhistoriker#Schriftsteller
116183168	https://data.swissbib.ch/person/5b26231e-7ea8-3a6d-9d8a-082a94e4c3d5	50015695	Q120196	squl#grnd	Baumann, Emil	1882-03-14	Basel	4004817-5	1951-02-05	Basel	4004817-5	M	Künstler#Schriftsteller#Malerei#Lithograf
104610956	https://data.swissbib.ch/person/5b26231e-7ea8-3a6d-9d8a-082a94e4c3d5	30542583	Q15453801	squl#grnd	Biedermann, Carl	1824			1894			M	Schriftsteller#Kaufmann
104825995	https://data.swissbib.ch/person/5b26231e-7ea8-3a6d-9d8a-082a94e4c3d5	15215754	Q1656493	squl#grnd	Biedschieder, Ida	1854-07-06	Zürich	4068038-1	1919-06-28	Zürich	4068038-1	W	Pädagogin#Schriftstellerin
104849221	https://data.swissbib.ch/person/5b26231e-7ea8-3a6d-9d8a-082a94e4c3d5	306825637	Q18269770	sgu	Birmel, Johann Georg	1858-03-28	Wetzell	4128525-6	1927-10-31	Romanshorn	4012593-3	M	
116203943	https://data.swissbib.ch/person/5b26231e-7ea8-3a6d-9d8a-082a94e4c3d5	61648027	Q17626760	squl#grnd	Bloesch, Hans	1879-12-26	Bern	4005762-8	1945-04-28	Bern	4005762-8	M	Schriftsteller#Bibliothekar#Germanistik, Philosophie
112044530	https://data.swissbib.ch/person/5b26231e-7ea8-3a6d-9d8a-082a94e4c3d5	59227654	Q11756884	sgu	Bögli, Lina	1858			1941			W	
137202709	https://data.swissbib.ch/person/5b26231e-7ea8-3a6d-9d8a-082a94e4c3d5	81426726	Q1990452	squl#grnd	Bott, Nikolaus	1864-06-16	Lichtensteig	4392453-0	1947-02-26	Riehen	4049979-0	M	Schriftsteller#Jugendbuchautor
118893017	https://data.swissbib.ch/person/5b26231e-7ea8-3a6d-9d8a-082a94e4c3d5	64113465	Q117227	squl#grnd	Bosshart, Jakob	1862-08-07	Oberembrach	4533096-7	1924-02-18	Clavadel	4560208-6	M	Pädagog#Schullehrer#Schriftsteller#Erzähler

Tab. 2: Ausschnitt aus der unbereinigten erweiterten Liste CH-DE-AutorInnen

Aus VIAF wurden die Identifikationsnummer und die Werte des Attributs *dbo:notableWorks* für die Autor:innen bezogen.²⁸ Aus Wikidata wurde die Identifikationsnummer wdt:P800²⁹ und das Geschlecht wdt:P21 angereichert. Via lobid³⁰ wurde schliesslich die GND nochmals abgefragt, um weitere Merkmale wie Geschlecht, Geburts- und Sterbedatum sowie -ort zu ergänzen. Externe Daten wurden nur berücksichtigt, wenn diese nicht bereits vorhanden waren. Die lokalen Daten hatten also immer Vorrang. Ebenfalls wurde die Identifikationsnummer des swissbib Linked-Data-Datensatzes ergänzt.³¹ Diese Identifikationsnummer ist wichtig, denn über sie werden die Werktitel aus dem swissbib Katalog abgefragt.

2.4.2 Werke finden

Zuerst wurden die Werke, welche in den *Notable Works*-Attributen aus VIAF und Wikidata referenziert wurden, geholt und mit weiteren relevanten Metadaten ergänzt. Dieser Prozess war nicht sehr ergiebig. Aus VIAF konnten so $n = 1'910$ Werke extrahiert werden und aus Wikidata nur $n = 130$ Werke. Die VIAF-Werke enthalten neben dem Metadatum „Titel“ nur das Metadatum „Sprache des Texts“. Die Metadaten-Bandbreite ist hier also sehr schmal, dafür sind diese aber fast komplett. Bei Wikidata konnte neben „Titel“ und „Sprache“ auch der „Typ des Werkes“³² (entspricht in etwa „Gattung“) extrahiert werden.

Um die Anzahl der Werke zu erweitern, wurden auch solche aus dem Bibliothekskatalog *swissbib* ausgewertet. Leider ist auch diese Liste nicht vollständig, da nur Datensätze gefunden wurden, die im Katalog mit der GND (oder einer anderen Autoritätsidentifikationsnummer) verknüpft sind. Insgesamt konnten aber über $N =$

56'000 Katalogeinträge identifiziert werden. Allen gemeinsam ist, dass eine Person aus dem Autor:innen-Datensatz (#4.1) in irgendeiner Form, sei es beispielsweise als Herausgeber, zum Werk beigetragen hat.

Dieser grosse Datensatz musste jedoch noch nachbearbeitet werden, um Dubletten und nichtrelevante Einträge zu entfernen. Die meisten Werke sind in mehreren Bibliotheken vorhanden und erhalten so jeweils einen eigenen Eintrag in unserem Datensatz. Um den Datensatz zu verkleinern, konnte ein Wert genutzt werden, welcher solche Dubletten zumindest teilweise zusammenführt. Zusätzlich kommen in dieser Liste auch Übersetzungen vor (die nicht relevant sind, da das Projekt nur auf Deutsch verfasste Werke berücksichtigt). Mit diesen beiden Kriterien konnte der Datensatz auf $N = 25'410$ Einträge reduziert werden.³³

Ein Filtern nach den Datenwerten „1880“–„1930“ ergab eine Liste von $n = 7'963$ Werken von $n = 868$ Autor:innen. Jedoch sind im Feld „Datum“ insgesamt $n = 2'184$ NAs oder andere nicht-verwendbare Codes enthalten, die zumindest potenziell weitere relevante Werke (und Autor:innen) enthalten.

Unser *work in progress* berichtet also konservativ rund 11'000 literarische Werke von circa 850 Autor:innen Deutschschweizer literarischer Prosa für 1880–1930. Da hier noch einige irrelevante Items, zum Beispiel nichtliterarische Gattungen und nichtdeutschsprachige Texte, enthalten sind, handelt es sich um eine sicherlich noch „aufgeblähte“ Zahl. Unsere definitive Annäherung an die Grundgesamtheit erfordert einige weiteren Schritte, u.a. mit Filtern, an die sich eine intellektuelle Überprüfung anschließt. Dubletten können über Suchanfragen nach Metadaten (wie Titel, Sprache, Publikationsjahr, Publikationsort, und in manchen Fällen auch über das Thema) relativ rasch identifizieren werden. Wie erwähnt müssen insbesondere in den Feldern „Publikationsdatum“ und „Gattung“ fehlende Werte auch händisch ergänzt werden, um den Datensatz anschliessend aufräumen zu können.

Die Vorstellung, mit maschinellen Abfragen auch nur ein annähernd vollständiges Datenset *direkt* erstellen zu können, hat sich für unser Projekt zur Deutschschweizer Literatur noch als Utopie erwiesen. Hingegen wurde für unsere Fragestellung klar, dass die manuelle philologische Erhebung weiter eine unumgängliche „Grundlagenforschung“ darstellt, und zwar nicht nur bezüglich der nicht-kanonischen Werke, sondern auch weit in den kanonischen Bereich hinein.

Die Vorteile des maschinellen Ansatzes, dessen Potenzial auch andere Projekte im Bereich der Literaturwissenschaften erkannt haben³⁴, liegen aber auf der Hand. Was

er auch in unserem Projekt bereits leisten konnte, ist eine effiziente Anreicherung der philologischen Liste mit verschiedenen LOD-Identifikatoren und das schrittweise Anreichern des Datensatzes um Namen, Werke und weitere Metadaten, die sich in der finalen Bereinigungsarbeit als nützlich erweisen werden. Zudem sind forschungsrelevante Metadaten zu Autor:innen (wie Gender, biographische Daten) und Werken (Publikationsjahr, Gattung, Seiten-Umfang und Kanonizität) zumindest in Teilen systematisch vorhanden. Lücken und Inkonsistenzen sollen hier in einigen Folgeschritten sukzessive ausgeglichen werden.^{[35](#)}

Die gemeinsame Arbeit hat uns dem Ziel, die Schweizer Erzählliteratur für den Zeitraum um 1900 präzise zu beschreiben, um ein sehr großes Stück angenähert. Nicht nur für unser Projekt, sondern auch angesichts einer wachsenden Nachfrage aus literatur-, kultur- und datenwissenschaftlichen Disziplinen sind vollständigere, korrektere und einheitlichere Metadaten ein Desiderat.

Jedoch sind gänzlich saubere, einheitliche und vollständige Metadaten wohl noch länger nicht erreichbar, vereint doch swissbib insgesamt 900 Katalogsysteme aus der Schweiz. Es sind nicht nur die unterschiedlichen Erfassungsstandards, die sich in der Datenqualität niederschlagen, sondern auch die historische Gewachsenheit der Datenbestände, die zu je unterschiedlichen Zeitpunkten mit unterschiedlichen Regelwerken erfasst wurden. Doch für die Verbesserung der Datenlage und die Erprobung neuer Vorgehensweise bedarf es eben Projekten wie dem unseren.

Für unser Projekt kann konstatiert werden, dass wir nach Abschluss der Bereinigung mit einer für den Bereich der digitalen Literaturwissenschaft vergleichsweise hohen^{[36](#)} Abbildgenauigkeit der Daten rechnen können. Die maschinell verlinkte Datensuche eröffnet uns auch dank der philologischen Liste, sowie *Gender* und *Region* als Balancefaktoren für nichtkanonische Werke, einen Überblick der Erzählliteratur, der zwar sicherlich nicht die Grundgesamtheit selbst darstellt, aber unseres Wissens nach erstmalig einen *quantitativen Zugriff* auf das ideelle Gesamtkorpus erlaubt. Für unsere weitere Forschung stellt es also eine wertvolle Grundlage für das wohlinformierte Sampling repräsentativer Volltexte dar, um in der Folge Emotionsrepräsentationen in erzählten Räumen der Literatur quantitativ zu untersuchen.

3. Die Retrodigitalisierung

Für das affektive und räumliche Textmining der Deutschschweizer Literatur werden derzeit möglichst viele der Werke unseres Datensatzes als digitale Volltexte gesammelt und in Auswahl auch retrodigitalisiert.^{[37](#)}

Die Komplexität des Vorhabens der quantitativen Konzeptualisierung wurde nochmals deutlich, als es konkret darum ging, diejenigen Werke, die noch nicht digital vorlagen, für die Digitalisierung zu beschaffen. Obwohl der Begriff „Schweizer Literatur“ eine gewisse Selbstverständlichkeit besitzt, zeigte sich, dass die Deutschschweizer Literatur weder listenförmig noch in physischen Ausgaben leicht verfügbar ist. Rund 40% der Werke waren nicht oder nicht in der gewünschten Ausgabe in der UB Basel vorhanden und mussten aus anderen Bibliotheken beschafft werden – teils sogar aus Bibliotheken, zu denen die UB Basel keinen Kurierdienst besitzt. Dies war eine neue Erfahrung für die UB Basel, die mit ihrer Digitalisierungsstrategie auf die Verfügbarmachung ihrer eigenen Bestände fokussiert.

Auch beim Digitalisierungsworkflow musste sich die Bibliothek an die konkreten Anforderungen des Projekts anpassen. In erster Linie betraf dies das Ausgabeformat der Textdaten und die Organisation der fortlaufenden Datenübergabe an das Projekt. Auch weicht das Projekt in der Arbeitsweise vom klassischen Digitalisierungsauftrag ab, indem es nicht mit einer fertigen Titelliste mit Signaturen begann, sondern eine Kollaboration im Sinne eines *Work in Progress* war, bei der unter anderem auch die Abklärung des Urheberrechts virulent wurde, etwa, wo ein Roman von 1912 noch nicht in die Public Domain fällt, da die Autorin zum Publikationszeitpunkt sehr jung war und erst 1969 verstarb.

Eine zentrale Sorge betraf die Qualität der OCR-Lesung bei den Frakturschriften. Da das Projekt an den Erstausgaben der Werke interessiert ist, liegt ein Großteil noch in Frakturschrift vor. Glücklicherweise wiesen die meisten Werke einen guten physischen Erhaltungszustand auf und verfügten über eine gute Druckqualität. Dank der hohen Scanqualität konnte eine insgesamt sehr gute OCR-Qualität erzielt werden, eine systematische Überprüfung der Character- und Word-Error Rate war daher nicht nötig. Allerdings gab es dabei auch Ausnahmen: Bei Erzählungen, die umfangreiche Passagen in Mundart enthalten, schlug die OCR-Lesung systematisch fehl. Solche Texte mussten in der *post correction* manuell berichtigt oder gar ausgesondert werden. An dieser wie auch an vielen anderen Stellen zeigte sich, wie wichtig die Kooperation zwischen Forschungsprojekt und Bibliothek ist. So prüft das Bibliothekspersonal beispielsweise in der Digitalisierungsvorbereitung die Werke, um wirklich sicherzustellen, dass keine Lyrikbände oder reine Mundartwerke digitalisiert werden.

Da diese projektspezifische Digitalisierung in manchen Punkten von der bibliotheksinternen Routine abwich, war eine fortlaufende Abstimmung zwischen

Digitalisierung und Projekt notwendig, was eine hohe Flexibilität beider Seiten erforderte. Durch das gezielte Mitdenken der Voraussetzungen und Ziele des anderen konnten Synergieeffekte ausgeschöpft werden. So schlug die Bibliothek den Projektverantwortlichen die zusätzliche Publikation der gemeinfreien Werke auf der digitalen Bibliothek *e-rara*³⁸ vor. Dies war nicht nur ein Gewinn für die Bibliothek, die so die Sichtbarkeit ihrer Bestände erhöhen kann, sondern kommt auch dem Projekt zugute, indem dieser Teil der Forschungsdaten bereits kostenlos dauerhaft gesichert und verfügbar ist.

4. Fazit

Die beschriebene Kooperation hat uns dem Ziel, den Zusammenhang von Gefühl und dargestelltem Raum während fünf Dekaden der Deutschschweizer Literatur zu erforschen, ein grosses Stück nähergebracht. Die gemeinsame Arbeit war eine heutzutage noch recht seltene, aber lohnende Investition, weil es für die datenintensive und datengetriebene Forschung hochskalierte, aber eben auch belastbare Daten braucht. Im Unterschied zu vielen anderen Projekten gingen wir also nicht von einer rein pragmatischen Sammlung bereits digitalisierter Volltexte aus, sondern starteten mit dem Meilenstein eines möglichst kompletten Index der relevanten Werke: Der Index wird bereits vor Akquise der Volltexte erstellt, um einschätzen zu können, an welchen Stellen gezielte Retrodigitalisierung erfolgen soll, um eine möglichst gute Repräsentativität der „wahren“ Population der Texte zu erzielen.

Ausgehend von unserer Datenbasis liegt die tatsächliche Anzahl Deutschschweizer literarischer Prosawerke 1880–1930 zwischen 7'000 und 10'000 Titeln. Dass diese Schätzung in einem engen Sinne datenbasiert ist, ist nicht trivial, sondern ein Novum. Dem Kriterium der *Hochskalierung* ist die Arbeit also bereits gerecht geworden: bei dieser literaturwissenschaftlichen Form der „big data“ geht nicht mehr um einige wenige exemplarische Titel, sondern um die tatsächliche Grundgesamtheit der gedruckten Werke.³⁹ Wie beschrieben, müssen unsere Daten aber noch weiter bereinigt werden, um auch *belastbar* zu sein. Dazu muss die *Spezifität* (Precision)⁴⁰ des Datensatzes erhöht werden, vor allem durch die Entfernung von irrtümlich eingeschlossenen Werken (false positives).

Es ist klar, dass auch die gesäuberte Liste nur eine Annäherung an den wahren Umfang und die Variabilität des fokussierten Publikationszeitraums und -genres sein kann. Zur weiteren Erhöhung der *Sensitivität* (*Recall*), besonders zur Entfernung eines

Kanonizitätsbias, ist also die Ergänzung der Liste um bislang nicht berücksichtigte Titel (false negatives) erforderlich. Diese Arbeit ist aufgrund des erhöhten intellektuellen Aufwands und des benötigten Expertenwissens nicht trivial. Sie sollte daher Sache eines kollaborativen Sammel- und Dokumentationsauftrags sein, der sich als Aufruf an spezialisierte Forschungsprojekte ganz unterschiedlicher Couleur wendet. Zur Kommunikation dieses Auftrags sind verschiedene Szenarien vorstellbar, so etwa der Einbezug der Akademien (SAGW, ÖAW, Akademienunion) sowie der zentralen Forschungsförderungsinstitutionen (SNF, DFG und FWF) und weiterer Institutionen über die Ländergrenzen hinweg.

Ein anderes Desiderat ist die Interoperabilität bestehender *linked data* Ressourcen in Echtzeit, unter automatischer Aktualisierung der Datensätze aus mehreren unabhängigen Repositorien zugleich. Alassi (2020, Kapitel 7) legt einen generischen Workflow für über RDF Triplette flexibel durchsuchbare digitale Editionen vor, der auch für andere Typen von linked data nutzbar wäre.

Doch zurück zu unserem Projekt: Hätte es im Nachhinein andere, gegebenenfalls einfachere Wege gegeben? Wir meinen: wohl kaum, oder aber nur in Nuancen. Unsere Hauptentscheidungen betrafen

- *Grundgesamtheit vor Textkorpus*: Wir verfolgen einen Ansatz, der vom *Usus* vieler Digital Humanities-Projekte etwas abweicht. Anstatt lediglich die bereits digital verfügbaren Werke als Textkorpus zu sammeln oder uns pragmatisch auf eine Untermenge zu verständigen, erheben wir zuerst die quantitative Grundgesamtheit der publizierten Werke. Wenn das Ziel eine quantitative Erhebung ist, sind eben (ausgehend von der in unserem Fall unerlässlichen manuellen Arbeit der „philologischen Liste“) quantitative Vorgehensweisen und digitale Ressourcen nötig.
- *Entwicklung des Workflows*: Wir haben in unserer Zusammenarbeit einen Workflow zwischen den unterschiedlichen Quellen (Katalog, digitale Ressourcen, LOD, philologische Quellen) erst entwickelt und dabei die stetige Anreicherung des Datensatzes in seinen verschiedenen Versionen vollzogen.

Bezüglich des Workflows ist aus Sicht des *data librarianships* sicherlich erneut die Frage zu stellen, ob die besten, funktionellen und aktuellen, LOD-Ressourcen⁴¹ genutzt und in der effizientesten Weise verknüpft wurden. So stellt sich insbesondere die Frage, ob die unterschiedlichen Arbeitsschritte #2–#4 besser hätten verzahnt werden können:

- *Wäre ein früherer Zugriff auf Wikidata oder DBPedia sinnvoll gewesen?* Wir meinen nicht, brauchten wir doch den Weg über die Autor:innen-GND, den philologischen „Goldstandard“ und das schrittweise Anreichern weiterer LOD-Indices. DBPedia ist momentan weniger aktuell als Wikidata.
- *Wäre eine frühere Quellenkritik zwischen den Schritten #3 und #4, also der weiteren Anreicherung, sinnvoll gewesen, um die Datenmenge kleiner zu halten?* Quellenkritik der Listen ist in jedem Fall notwendig, zum Beispiel, weil schon unser zentraler und erster LOD-Identifikator, die GND, für unsere Zwecke sehr ungenau ist. Aus den angereicherten Daten müssen die Einträge mit irrelevanten Lebensdaten, Berufsbezeichnungen (Regisseur:in und Übersetzer:in), und Sprache (Welschschweizer Autor:innen) entfernt werden. Sicher ist, dass die angereicherten Daten ausgedünnt werden müssen, unklar ist noch der optimale Zeitpunkt dieses Beschneidens: ob iterativ zwischen den Anreicherungen oder am Schluss?

Das Ausloten der Bibliothekskataloge als Forschungsinfrastruktur ist bei Weitem noch nicht abgeschlossen und weitere spezifische Vorhaben müssen hier ebenfalls Desiderata formulieren. Denn unser Projekt macht vor allem eines deutlich: datengetriebene Forschungsvorhaben wollen künftig immer häufiger mit solch präzisen Fragestellungen auf Katalogsysteme zugreifen und Katalogdaten als Forschungsdaten (oder Ausgangspunkt für Forschungsfragen) verwenden.

Aus Sicht der Bibliothek ist eine gute Metadaten-Qualität für (zukünftige) Forschungsprojekte daher ausschlaggebend. Maschinell kann aus dem Bibliothekskatalog wie auch aus Normdateien und Datenbanken aber nur abgefragt werden, was in den Daten enthalten ist. So kommt der Datenpflege und Qualitätssicherung auch unter dem Vorzeichen des Digitalen grosse Bedeutung zu. Diese klassische Bibliotheksaufgabe darf nicht unterschätzt werden. Neben einheitlichen Erfassungsstandards und Bereinigungsarbeiten (zum Beispiel nach Zusammenführung von Datenbeständen aus unterschiedlichen Katalogen) ist auch zu prüfen, ob und welche weiteren für Forschungsprojekte datenwissenschaftlicher Disziplinen relevanten Parameter neu erfasst werden könnten.

Die Bibliothek bekommt als Forschungsinfrastruktur zudem einen neuen Akzent. Ein aktives Anbieten von Informationen und eine gute Sichtbarkeit für die Wissenschaftler:innen ist nötig, die wissen müssen, dass ihre Fragen hier beantwortet werden und sie mit der Bibliothek projektorientiert zusammenarbeiten können. Ein Forschungsprojekt mit seiner konkreten Fragestellung ist ein echter Test der Bibliotheksinfrastruktur und wirft viele Fragen auf – nicht nur hinsichtlich der Qualität

der Erschliessung, dem Stellenwert von Katalog, Datenbanken, Referenzwerken, Bibliographien, sondern auch hinsichtlich forschungsbezogener Bibliotheksservices. Dabei heisst „digital“ noch lange nicht nutzbar. Problempunkte sind Interoperabilität und Datenqualität: Der Zugriff auf Linked Data ist derzeit noch kein standardmäßig und leicht verfügbarer Dienst.

Wir schliessen mit einigen konkreten „learned lessons“ und Desiderata:

Korpusliteraturwissenschaftliche Forschungsprojekte sollten Arbeitspakete für die Erhebung nicht nur von Volltexten, sondern auch von *Metadaten* einplanen. Deren Modellierung ist, gerade wenn die Datensätze noch gar nicht oder nur in Teilen vorhanden sind, ein wichtiger Teil des Forschungsprozesses. Diese neue *metadatenbasierte Literaturwissenschaft* ist zwischen Fachdisziplin und Data Librarianship angesiedelt.

Daher schlagen wir vor: Forschungsbibliotheken sollten idealerweise schon vor Projektbeginn Ansprechpartner:innen für (Literatur-)Wissenschaftler:innen sein, um Angaben über Datenlage und Datenqualität zu machen und gemeinsam den zeitlichen und personellen Aufwand abzuschätzen. Zudem sollte es auf Seiten der Bibliothek spezialisiertes Personal geben, das entweder im Rahmen des Bibliotheksdeputats mitarbeitet oder freigestellt würde, um aus Projektmitteln finanziert zu werden. Bei all diesem wäre eine zentrale Koordinationsstelle hilfreich, die Arbeitsabläufe und Kommunikationswege bereits im Vorfeld klärt und Kompetenzen bereitstellt.

Reiseführer in der Anfangszeit des Alpinismus stellten erste Handleitungen für die „rauen Wege der Alpen, die oft jähren und gefährlichen Pfade auf denselben“ ([Wytttenbach, 1777](#)) bereit. Ganz ähnlich haben wir aufgezeigt, wie das teils unwegsame und immer hochskalierte Gelände der „(Meta-)Datenberge“ unseres korpusliteraturwissenschaftlichen Projekts begehbar werden kann. Wo Wytttenbach mindestens einen langen Bergstock und beschlagene Schuhe empfiehlt, nebst Fusseisen und Überstrümpfen gegen die Insekten, haben wir – so steht zu hoffen – konkret gezeigt, wie „Reisepartien“ aus Bibliothek und Fachwissenschaft im Datengelände erfolgreich zum Ziel kommen können. Wir enden mit der schönen Aussicht, dass in absehbarer Zeit wohl auch die *Datenberge der Literaturwissenschaft* nicht nur rau aufgetürmt, sondern kartographiert verheißungsvoll zum Auskundschaften bereitstehen.

Footnotes

1. SNF-Projekt 189832 „High Mountains Low Arousal? Distant Reading Topographies of Sentiment in German Swiss Novels in the early 20th Century“, Kollaboration zwischen DHLab Basel und UB Basel. <http://p3.snf.ch/Project-189832>. Das Projekt ist mit der EU COST Action „Distant Reading“ www.distant-reading.net assoziiert. Im Sinne der Zielsetzung der COST Action „Distant Reading“ hat das Projekt in der Kooperation mit der UB interdisziplinär einen genuin neuen Ansatz der digitalen Literaturgeschichtsschreibung konkretisiert. Die zu diesem Ergebnis führende Forschung geht neue Wege und bedarf der Verzahnung unterschiedlicher Expertisen. [↪](#)
2. Angesichts einer wichtigen Phase der Literaturgeschichte – der Zeit des modernen Umbruchs um 1900, die mit narrativen Traditionen von Wahrnehmungs- und Affektdarstellung bricht, sie aber auch weiterführt – legen wir in einer komparatistischen Studie Verfahren des „Distant Reading“ an, um womöglich spezifisch schweizerische Verfahren zu beschreiben: die affektive Ausgestaltung fiktional repräsentierter Räume in Deutschschweizer Literatur im quantitativen Vergleich im Spannungsfeld zwischen realistisch und fantastisch ausgestalteten Topographien von Berg und Ebene, Stadt und Land, Natur und Industrie. [↪](#)
3. Eine zentrale Rolle in der dargestellten Forschung spielte als *Data Librarian* Jonas Waeber, dem wir für seinen Einsatz vielmals danken. Einen ersten Überblick des Workflows zeigten wir im Vortrag Herrmann, J. B., Grisot, G., & Waeber, J. (2020). *Distant Reading Topographies of Sentiment in German Swiss Novels in the early 20th Century. A methodological progress report*. [Invited talk] Research Meeting Research Area 5 “Building Digital Communities”. EXC 2020 Temporal Communities, Doing Literature in a Global Perspective, Freie Universität Berlin. [↪](#)
4. <https://www.dnb.de/DE/Professionell/Standardisierung/GND/gnd.html>. [↪](#)
5. MARC21-Feld 084 mit DNB-Nebensachgruppe „B“. [↪](#)
6. So verwenden etwa die Zentralbibliothek Zürich und die Deutsche Nationalbibliothek schon früher Codierungen (und Schlagwörter) als die UB Basel. Für den relevanten Erscheinungszeitraum sind diese jedoch nur vereinzelt im Katalog vorzufinden (eventuell spätere Eingänge?). Die Schweizerische Nationalbibliothek wiederum hat bei der Rekatalogisierung der Zettelkarten die Informationen des Sachkatalogs (Dezimalklassifikation, weist die Belletristik von Beginn an separat nach) nicht mit übertragen. Diese sind daher heute leider nur

über die digitalisierten Zettelkarten einsehbar

(http://siibns.ch/german/cat1_frame.htm), jedoch nicht via Katalog abfragbar
<https://www.helvetica.ch/>). ☹

7. Seit 1901 von der 1895 gegründeten Schweizerischen Landesbibliothek (heute Nationalbibliothek) herausgegeben, davor von zwei Antiquariaten. Schweizerisches Antiquariat (Ed.). (1871-1877). *Bibliographie der Schweiz: bibliographisch-kritische Revue der neuen Erscheinungen der Schweiz* = Bibliographie de la Suisse: revue bibliographique et critique des publications nouvelles de la Suisse.; Georg, H. (Ed.). (1878-1901). *Bibliographie und literarische Chronik der Schweiz* = Bibliographie et chronique littéraire de la Suisse.; Benteli, A. (Ed.). (1901-1942). *Bibliographisches Bulletin der Schweizerischen Landesbibliothek* = Bulletin bibliographique de la Bibliothèque nationale suisse.; Verlag des Schweizerischen Buchhändlervereins (Ed.). (1943-2007). *Das Schweizer Buch: bibliographisches Bulletin der Schweizerischen Landesbibliothek, Bern* = Le livre suisse: bulletin bibliographique de la Bibliothèque nationale suisse = Il libro svizzero: bollettino bibliografico della Biblioteca nazionale svizzera.; Schweizerische Nationalbibliothek (Ed.). (2007-). *Das Schweizer Buch: schweizerische Nationalbibliografie* = Le livre suisse: bibliographie nationale suisse = Il libro svizzero: bibliografia nazionale svizzera = Il cudesch svizzer: bibliografia naziunala svizra = The Swiss book: Swiss national bibliography. ☹

8. So werden auch literarische Werke Nichtschweizer Autor:innen verzeichnet, sofern diese bei Schweizer Verlagen erschienen sind oder über einen Schweizbezug verfügen. ☹

9. So fehlte in Lobid etwa der GND-Ländercode für die Schweiz und Linked Swissbib lieferte zu viele nicht relevante Namen. Eine Anreicherung mit DBpedia, VIAF und Wikidata wurde erst vorgenommen, als die GND-ID als Identifikator vorlag.

☹

10. Einbezogen wurden: „Schriftsteller* OR Autor* OR Romancier OR Romanschriftstellerin OR Reiseschriftsteller* OR Kriminalschriftsteller* OR Prosaist* OR Exilschriftsteller* OR Heimatschriftsteller* OR Kinderbuchautor* OR Jugendbuchautor*“. Nicht berücksichtigt wurden Bezeichnungen wie „Lyriker* OR Dramatiker* OR Mundartschriftsteller*“. ☹

11. Alle Zwischenschritte wurden als Forschungsdaten unter FAIR Principles dokumentiert und stehen auf Zenodo zur Verfügung:

<https://zenodo.org/record/4592279#.YEe237CSlPY>. ☹

12. „Schriftsteller:in“ statt „Lyriker:in“ oder „Autor:in“ statt „Schriftsteller:in“. Generell am häufigsten kommt die Berufsbezeichnung „Schriftsteller:in“ vor, zum Teil wurden aber auch engere Bezeichnungen wie „Lyriker:in“ oder „Romancier“ vergeben. Die Bezeichnung „Autor:in“ wird gemäss GND als Oberbegriff für Verfasser:innen sowohl literarischer wie auch wissenschaftlicher Werke verwendet – oder auch in Fällen, wo die Personenangaben nicht ausführlich recherchiert wurden (siehe auch Fussnote 15). [↵](#)
13. Vor allem „Autor*“: Hier wurden offenbar auch Berufsbezeichnungen wie „Kochbuchautor:in“ oder „Drehbuchautor:in“ aufgefunden, wie sich später zeigte. [↵](#)
14. NA für engl. „Not available“: fehlender Wert. [↵](#)
15. Da auch die Schweizer Bibliotheken an der GND mitarbeiten und die Deutsche Nationalbibliothek die deutschsprachige Literatur breit sammelt, ist die Abdeckung zwar sehr gut. Aber aus Aufwandsgründen ist es praktisch nicht möglich, für jede Person, die zu Lebzeiten eine Publikation veröffentlicht, einen Normdatensatz mit individualisierten Angaben zu erstellen. Zudem wurden in den letzten Jahren auch maschinell Altdatensätze verschiedener Bibliotheken und Bibliotheksverbünde in die GND eingespielt, die bisher nicht vollständig bereinigt sind. [↵](#)
16. Gerne hätte die Fachreferentin diese grosse Datenmenge elektronisch genutzt. Eine Anfrage bei den Autorinnen und beim Verlag stiess auf Wohlwollen, die Disketten waren aber nicht mehr auffindbar. Die Auswertung musste daher ab Band erfolgen und berücksichtigte nur Namen von Autorinnen, von denen mehr als ein Werk im relevanten Zeitraum erschienen ist. [↵](#)
17. <https://hls-dhs-dss.ch/> und <https://de.wikipedia.org/>. [↵](#)
18. Historisches Lexikon der Schweiz: „Personen -> Künste und Literaturen -> Literaturen -> Deutschsprachige Literatur“; Wikipedia: „Schweizer Literatur -> Literatur der deutschsprachigen Schweiz“ und „Liste Schweizer Schriftsteller -> Autoren der deutschen Schweiz“. [↵](#)
19. Berufsbezeichnungen oder Länder etwa sind gar nicht erfasst (nur Geburts-, Sterbe- und zum Teil Wirkungsorte), Gattungsbezeichnungen und diverse andere Angaben werden abgekürzt, eine Normierung der Daten findet nicht statt. Dies wirft durchaus Fragen bezüglich Kosten-Nutzen solcher Anschaffungen an der Bibliothek auf. [↵](#)

20. Lokale Felder 909A_f\$ blaaut resp. 909B_\$b bb. [↵](#)
21. Im Gegensatz dazu liegen die *Basler Bibliographie* wie auch die *Bibliographie zur Geschichte des Kantons Bern* bereits vollständig digitalisiert vor. [↵](#)
22. *Autopsie* bedeutet im Bibliothekswesen die Katalogisierung anhand des tatsächlich vorliegenden Buches. [↵](#)
23. Hinweise auf die Kanonizität sollte auch die Quellenangabe selbst liefern (insbesondere die Erwähnung in Literaturgeschichten). [↵](#)
24. <https://openrefine.org/>. [↵](#)
25. Virtual International Authority File, siehe <https://viaf.org/>. [↵](#)
26. <https://www.wikidata.org/>. [↵](#)
27. Auf eine Anreicherung mit DBPedia wurde verzichtet, da der Datenbestand von VIAF und Wikidata auf neuem Stand ist. [↵](#)
28. Dieses Attribut enthält Referenzen zu VIAF-URIs von Werken von der indizierten Person. Wie Stichproben ergaben, ist diese Liste nicht vollständig. Sie stellt aber eine wertvolle Ressource dar. [↵](#)
29. Notable Works. Wie bei VIAF enthält dieses Attribut Referenzen zu Wikidata-URIs von Werken der indizierten Person. Auch diese Liste ist bei weitem nicht erschöpfend. [↵](#)
30. <https://lobid.org/>. [↵](#)
31. Dieser Datensatz wurde vom swissbib-Projekt basierend auf den Katalogdaten erstellt und ist frei verfügbar über eine Schnittstelle. Jedoch wurde in diesem Projekt ein direkter Zugriff auf die Datenbank verwendet, was deutlich performanter war. [↵](#)
32. Dazu wurde die Property wdt:P31 (instance of) ausgewertet. [↵](#)
33. Der Entstehungsprozess dieser (noch preliminären) Ergebnisse ist im Sinne der FAIR Principles nachvollziehbar und wiederholbar. Ein anderes Thema ist das Desiderat einer „absoluten“ Interoperabilität in Echtzeit: während zukünftige Mutationen in den Quelldaten bei uns zu numerisch abweichenden Ergebnissen führen, sollte über Möglichkeiten nachgedacht werden, unifizierte Datensätze aus unterschiedlichen Quellen in Echtzeit zu generieren. Vgl. Alassi, S. (2020). From the

Mechanics of Jacob Bernoulli to Digital History of Science. Infrastructure, Tools, and Methods (Dissertation, Universität Basel). <https://doi.org/10.5451/unibas-ep81737>. ↵

34. <https://www.zb.uzh.ch/de/events/informationen-digital-verpackt>. ↵

35. Als konkrete *todos* sind zu nennen (1) die Dedublierung der Werklisten (besonders um false positives mit Publikationsdatum ausserhalb 1880-1930, nichtlit. Genres (z.B. wiss. Werke), Serieneinträge, „Nichtschweizer“ Nationalität und Titelvarianten) und die (2) Ergänzung fehlender Werte (NAs) bzw. Spezifizierung bestimmter Werte (v.a. Publikationsdatum, -ort). Zudem wird derzeit (3) die Dokumentation der Forschungsdaten abgeschlossen und durch Publikation auf Zenodo geteilt. Zur (4) Erstellung des eigentlichen Textkorpus werden unsere Metadaten auf verfügbare Volltext-Objekte gematched, diese zum Textmining gesammelt und vorbereitet. Fortlaufend werden ausgewählte Titel retrodigitalisiert, die schliesslich als kleines „CH-DE Korpus“ (n = 100 Titel) im Rahmen des ELTeC publiziert werden (siehe <https://www.distant-reading.net/eltec/>). ↵

36. Man vergleiche die Aussagen über die Repräsentativität und Datenlage der grössten digitalen Volltextsammlung englischsprachiger Texte, dem HathiTrust, zum Beispiel bei Piper, A. (2018). *Enumerations: Data and literary study*. Chicago: University of Chicago Press. 102, der feststellt, dass die Metadatenlage des Hathi Archivs mit rund 21'000 Dokumenten nicht in aller Genauigkeit bekannt ist, oder Underwood, T. (2019). *Distant horizons: Digital evidence and literary change*. Chicago: University of Chicago Press. 117, der konstatiert, „this library is by no means a complete record of fiction published in English“ (siehe auch Underwood, 2019, Kapitel 4 und Appendix A). ↵

37. Workflow mit den folgenden Schritten: Auswahl der Titel aus der Bibliographie, Recherche nach digitalen Objekten in ausreichender Qualität, Identifikation von physischen Objekten im Bibliothekskatalog, Digitalisierung im engeren Sinne (mit OCR) und schließlich die XML-Encodierung im ELTeC-Format. Digitalisierte Werke werden über Abfrageskripte vor allem aus den Repositorien DTA (<https://www.deutschestextarchiv.de/>), KOLIMO, Gutenberg, <https://textgrid.de/> und Wikidata gesammelt. <https://www.projekt-gutenberg.org/> <https://fortext.net/ressourcen/textsammlungen/kolimo-korpus-der-literarischen-moderne>. ↵

38. <https://www.e-rara.ch>. ↵

39. Ohne „Goldstandard“ (den wir selbst in diesem Projekt erst erstellen) kann der *Recall* (die Wahrscheinlichkeit, mit der ein positives Objekt korrekt als positiv klassifiziert wird) nicht formal berechnet werden, wir schätzen ihn aber als recht hoch ein. [↵](#)

40. Die Wahrscheinlichkeit, mit der ein negatives Objekt korrekt als negativ klassifiziert wird. [↵](#)

41. <https://metagrid.ch/> zur schweizspezifischen Abfrage u.a. des Historischen Lexikons der Schweiz wurde bislang noch nicht angesprochen, wäre aber zukünftig eine gute Quelle. [↵](#)

Citations

1. Haller, A. von. (1762). Die Alpen. In *Versuch Schweizerischer Gedichte* (9th ed.). Göttingen. Retrieved from https://www.hs-augsburg.de/~harsch/germanica/Chronologie/18Jh/Haller/hal_v04.html#z6 [↵](#)
2. Herrmann, J. B., & Lauer, G. (2018). Korpusliteraturwissenschaft. Zur Konzeption und Praxis am Beispiel eines Korpus zur literarischen Moderne. *Osnabrücker Beiträge Zur Sprachtheorie*, 92, 127–156. Retrieved from http://gerhardlauer.io/files/2515/4755/9277/Herrmann_Lauer_Korpus.pdf [↵](#)
3. Meister, J. C., & Veit, J. (2014). Digital Humanities – Neue Netzwerke für die Geisteswissenschaften. *ZfBB*, 61(4–5), 263–266. <https://doi.org/10.3196/18642950146145174> [↵](#)
4. Döhl, F. (2019). Digital Humanities und Bibliotheken. Über technisch-organisatorische Infrastruktur hinausgedacht. *ZfBB*, 66(1), 4–18. <https://doi.org/10.3196/186429501966114> [↵](#)
5. Odebrecht, C., Burnard, L., Eder, M., & Schöch, C. (2019). *The European Literary Text Collection (ELTeC)*. Retrieved from <https://dev.clariah.nl/files/dh2019/boa/0715.html> [↵](#)
6. Herrmann, J. B., Odebrecht, C., Santos, D., & Francois, P. (2020). Towards Modeling the European Novel. Introducing ELTeC for Multilingual and Pluricultural Distant Reading. In *Book of Abstracts. International Conference DH2020, Ottawa, Canada*. Retrieved from https://dh2020.adho.org/wp-content/uploads/2020/07/212_TowardsModelingtheEuropeanNovelIntroducingELTeCforMultilingualandPluriculturalDistantReading.html [↵](#)

7. Petzold, K. (Ed.). (2007). *Schweizer Literaturgeschichte. Die deutschsprachige Literatur im 20. Jahrhundert*. Leipzig: Miltitzke. [↵](#)
8. Rusterholz, P., & Solbach, A. (Eds.). (2007). *Schweizer Literaturgeschichte*. Stuttgart: J. B. Metzler. [↵](#)
9. Nadler, J. (1932). *Literaturgeschichte der deutschen Schweiz*. Leipzig: Grethlein. [↵](#)
10. Aellen, H. (Ed.). (1918). *Schweizerisches Schriftsteller-Lexikon = Lexique des écrivains suisses = Lessico dei scrittori svizzeri*. Weinfelden: Schweizer Heimatkunst-Verlag. [↵](#)
11. Stump, D., Widmer, M., & Wyss, R. (1994). *Deutschsprachige Schriftstellerinnen in der Schweiz: 1700–1945: eine Bibliographie*. Zürich: Limmat Verlag. [↵](#)
12. Schweizerischer Buchhändlerverein, & Verein Schweizerischer Verlagsbuchhändler (Eds.). (1914). *Das Schweizerische Buch: Schweizerische Landesausstellung, 1896–1914*. Bern: A. Francke. [↵](#)
13. Schweizerischer Buchhändlerverein, & Verein Schweizerischer Verlagsbuchhändler (Eds.). (1931). *Das Schweizerische Buch: Katalog der noch lieferbaren Schweizer Bücher und Schriften aus den Jahren 1914–1930*. Zürich: Orell Füssli. [↵](#)
14. Schweizerischer Buchhändlerverein, & Verein Schweizerischer Verlagsbuchhändler (Eds.). (1939). *Livres suisses = Das Buch der Schweiz = Libri svizzeri: 1931–1938, mit Ergänzungen aus früheren Jahren*. Frauenfeld: Huber. [↵](#)
15. Underwood, T. (2019). *Distant horizons: Digital evidence and literary change*. Chicago: University of Chicago Press. [↵](#)
16. Alassi, S. (2020). *From the Mechanics of Jacob Bernoulli to Digital History of Science. Infrastructure, Tools, and Methods*. Dissertation, Universität Basel. <https://doi.org/10.5451/unibas-ep81737> [↵](#)
17. Wytttenbach, J. S. (1777). *Kurze Anleitung für diejenigen, welche eine Reise durch einen Theil der merkwürdigsten Alpgegenden des Lauterbrunnenthals, Grindelwald, und über Meyringen auf Bern zurück, machen wollen*. Bern: Bey Wagner, Hoch-Obrigkeitlichen Buchdrucker. <https://doi.org/10.3931/e-rara-32756> [↵](#)