

Introduction to Digital History

Ina Serif

Table of contents

Welcome	1
1 Was ist digital history?	3
2 Forschung und Lehre	5
2.1 Digitale Tools zur Analyse	7
2.2 Digitale Tools zur Kommunikation	9
2.3 Digitale Elemente in der Hochschullehre	9
2.4 Projekte und Ressourcen	9
2.4.1 Alte Geschichte	9
2.4.2 Mittelalter und Frühe Neuzeit	9
2.4.3 Moderne und Zeitgeschichte	10
2.4.4 Jüdische Geschichte	10
2.4.5 Geschichte Afrikas	11
2.4.6 Osteuropäische Geschichte	11
2.4.7 Epochen-/Areaübergreifend:	11
3 Digital Literacy, Digital Criticism	13
3.1 Digital Criticism, Data Criticism	14
4 Datenerhebung und -analyse	19
5 FAIR und CARE	21
6 shell	23
Appendices	25
A Appendix	27
B Further Ressources	29
B.1 Digital Literacy, Digital Criticism	29
B.2 Tools für digital history (n.b.: free/open source)	30
B.2.1 Allgemein	30

B.2.2	Text-/Korpusanalyse	30
B.2.3	Visualisierung	30
B.3	Terminal/Command Line	30
B.4	Regular Expressions	30

Welcome

Der vorliegende Guide begleitet die Einführungskurse im Fach Geschichte an der Universität Basel und soll einen ersten Einblick in den Bereich digital history geben. Die epochen- und areaspezifischen Inhalte der verschiedenen Einführungskurse sollen dabei berücksichtigt werden, die Verweise auf verschiedene digital-history-Projekte aus unterschiedlichen Epochen mit der Zeit also anwachsen.

Der Guide wird für die Teilnehmer:innen der Einführungskurse von einer Präsenzsitzung begleitet, bietet aber hoffentlich auch unabhängig davon einen Mehrwert. Für Kommentare, Anregungen oder Beschwerden freue ich mich über eine Nachricht.

Der Guide ist in zwei Teile gegliedert: Die Kapitel 1–5 sollen eine erste Übersicht über “digital history” bieten und den Blick auf Neuerungen und Veränderungen richten, die sich in den Geschichtswissenschaften aus der Nutzung digitaler Methoden ergeben. Der anschließende Praxisteil zeigt an einem konkreten Beispiel die Anwendung verschiedener Techniken auf, die sich (nicht nur) für Historiker:innen bei der Arbeit mit Quellenmaterial anbieten. Der Praxisteil verfolgt dabei zwei Ziele: Zum einen sollen Hemmungen bei der Arbeit mit dem Computer, die über die Nutzung als elektronische Schreibmaschine hinausgeht, abgebaut werden. Zum zweiten soll ein grundlegendes Verständnis dafür hergestellt werden, welche Möglichkeiten computergestützte Analysen bieten und wie diese in der historischen Arbeit eingesetzt werden können.

Die Übersicht soll möglichst knapp gehalten werden – es gibt zahlreiche ausführliche Grundlagenwerke, weswegen viele Themen nur kurz angeschnitten, dafür aber mit weiterführenden Verweisen versehen werden. Dasselbe gilt für den Praxisteil: Weiterreichende Anleitungen, Tutorials oder Onlinekurse werden an entsprechender Stelle verlinkt. Vollständigkeit wird an keiner Stelle beansprucht; Hinweise auf weitere Online-Angebote nehme ich gerne auf.

Chapter 1

Was ist digital history?

Über die Antwort zur Frage, was digital history ist oder umfasst, kann man ausgiebig diskutieren; als Teilgebiet der digital humanities, der digitalen Geisteswissenschaften, kann die folgende aktuelle und pragmatische Definition von Blaney et al. hilfreich sein:

Digital humanities, in our view, is a question of approach: if you are actively and critically using digital tools to aid your work in researching, teaching or learning, you are probably doing digital humanities. We would encourage anyone to learn to program if they are interested in doing so, but we do not see it as a defining characteristic of work in digital humanities.¹

Dabei umfassen “digital tools” eine große Bandbreite – und es wird sich kaum eine:r finden, der:die Studium, Forschung oder Lehre völlig ohne die Nutzung digitaler Techniken betreibt. Wir sind alle Historiker:innen im digitalen Zeitalter, und als solche müssen wir ohnehin neue Kompetenzen entwickeln. Wir können uns aber zudem dafür entscheiden, für ein Forschungsprojekt Methoden und Techniken einzusetzen, die über die traditionellen Werkzeuge der Geschichtswissenschaften hinausgehen – Analyse und Interpretation von Quellen durch deren genaue Lektüre, sogenanntes *close reading* –, und uns durch den Computer unterstützen lassen. Ob wir hierbei auf vorhandene Software zurückgreifen oder selbst Programme schreiben, um uns nicht nur als Historiker:innen im digitalen Zeitalter, sondern auch als digitale Historiker:innen zu verstehen, mögen manche als Glaubensfrage auffassen; eine inkludierende Haltung zu dieser Frage scheint mir dabei nur Vorteile zu haben.²

¹Blaney, Jonathan; Winters, Jane; Milligan, Sarah u. a.: Doing digital history: a beginner’s guide to working with text as data, Manchester 2021 (IHR research guides), S. 6.

²Entgegen einer häufig zitierten Aussage von Emmanuel Le Roy Ladurie (*1929), der Historiker von morgen werde Programmierer sein, oder er werde nicht sein: “L’historien de demain sera programmeur ou il ne sera pas.” Le Roy Ladurie, Emmanuel: La fin des érudits, in: Le Nouvel Observateur, 08.1968.

Für eine erste Idee dafür, wie man historische Fragestellungen mithilfe digitaler Methoden beantworten kann und wie unterschiedlich digital unterstützte Forschungsprojekte aussehen können, bietet sich unter anderem der Übersichtsartikel “State of the Field: Digital History” von Romein et al. (2020) an.³ Eine anwachsende Liste an Beispielprojekten aus unterschiedlichen Epochen bzw. Themenbereichen findet sich unter Projekte und Ressourcen in Kapitel 2.

Um eine Annäherung an die aktive, kritische und reflektierte Nutzung digitaler Methoden in Forschung und Lehre mit einem Fokus auf deren Anwendung in den Geschichtswissenschaften geht es im vorliegenden Guide. Weiterführende Texte zur Frage, was digital history ist bzw. umfasst, finden sich unter Further Ressources

³Romein, C. Annemieke; Kemman, Max; Birkholz, Julie M. u. a.: State of the Field: Digital History, in: History 105 (365), 04.2020, S. 291–312. Online: <<https://doi.org/10.1111/1468-229X.12969>>, Stand: 15.09.2022.

Chapter 2

Forschung und Lehre

Die fortschreitende Digitalisierung in ganz unterschiedlichen Lebensbereichen zieht Veränderungen und Entwicklungen auch für die historische Arbeit mit sich, und dies auf mehreren Ebenen: in Bezug auf die Arbeit bzw. den Umgang mit Quellen, hinsichtlich des Einsatzes digitaler Methoden nicht nur zur Analyse von Forschungsergebnissen, sondern auch für deren Kommunikation, und schließlich für die Hochschullehre.

Als Historiker:innen steht die Arbeit mit Quellen im Mittelpunkt unserer Analysen. Das bedeutet gleichzeitig, dass der Zugang bzw. die Verfügbarkeit von Dokumenten einen Einfluss darauf hat, welche Fragen wir beantworten oder welche Analysen wir vornehmen können. Zugangsbeschränkungen, die die Größe und Zusammensetzung unseres Untersuchungskorpus beeinflussen, können dabei von Gedächtnisinstitutionen ausgehen, beispielsweise wenn bei zeitgenössischen Akten eine Schutzfrist festgesetzt wird oder wenn ein Objekt zu fragil für die Benutzung ist. Auch kann es aus finanziellen und/oder organisatorischen Gründen schwierig sein, bestimmte Archive an weiter entfernten Orten aufzusuchen, um weitere Dokumente für die Untersuchung zu berücksichtigen. Groß angelegte Digitalisierungsprojekte in Bibliotheken und Archiven bergen damit die Möglichkeit, zusätzliche Quellen nicht nur über einen Eintrag im Bibliothekskatalog zu finden, sondern die entsprechenden Dokumente in digitaler Form auf den eigenen Rechner zu laden. Gerade auch für wertvolle historische Bestände – antike Papyri, Handschriften aus dem Frühmittelalter, einzelüberlieferte Frühdrucke usw. – entsteht hier die Möglichkeit, diese einem größeren Kreis verfügbar zu machen, ohne das Objekt zu großer Belastung durch häufige Benutzung auszusetzen, und ohne dass die Benutzer:innen lange Reisen auf sich nehmen müssten. Für mittelalterliche und frühneuzeitliche Handschriften und Drucke beispielsweise existieren mittlerweile mehrere (meist nationale) Portale, die eine zentrale Suche über alle Bestände ermöglichen; eine Auswahl findet sich unter Projekte und Ressourcen.



Figure 2.1: Randall Munroe, History Department, xkcd.com (17.12.2018).

Neben der Digitalisierung vorhandener Quellen (Retrodigitalisierung) steht die unaufhörliche Entstehung neuer Quellen in rein digitaler Form (born digital data). Der relativen Knappheit von Quellen – und damit Daten –, die Vormodernehistoriker:innen oftmals zu beklagen haben, steht eine Überfülle an zeitgenössischem Material gegenüber, und beide Situationen – zu wenig/zuvu unvollständige und zu viele/zuvu unübersichtliche Datenmengen – bergen methodische Probleme: Wie stellt man ein Korpus zusammen, das ausreichend Quellen beinhaltet, um Fragestellungen zu beantworten, Thesen zu stützen, neue Erkenntnisse zu erhalten, das aber gleichzeitig in einem Forscher:innenleben bewältigbar bleibt? Historiker:innen müssen neue Kompetenzen erwerben, um mit solchen Fragen reflektiert umzugehen. Zur klassischen Quellenkritik kommt die digitale Quellenkritik, zur Fähigkeit, analoge Quellen zu lesen und zu verstehen, ein Äquivalent für den digitalen Bereich. Etwas ausführlicher geht es in Kapitel 3 um Digital Literacy und Digital Criticism.

2.1 Digitale Tools zur Analyse

Die hier bereits zitierte Definition, die aktive und kritische Nutzung digitaler Werkzeuge in Forschung, Lehre oder Studium sei es, was digital humanities ausmachen, wirft die Frage auf, was genau unter “digital tools” zu verstehen ist, und zu welchem Zweck man sie einsetzt. Allein schon das Lesen dieses Guides ist ohne digitale Hilfsmittel nicht möglich – es existiert kein gedrucktes Exemplar davon. Lesen am Bildschirm allein macht noch keinen digital humanist, aber man muss nicht erst eine Programmiersprache lernen, um den Computer für die eigene Arbeit zu nutzen und zu Ergebnissen zu kommen, die mit klassischen Methoden – im Bereich der Geschichtswissenschaften etwa papierbasiertes *close reading* von Quellen und Forschungsliteratur – nicht im selben Ausmaß erzielt werden könnten.

Untersuchungen, die digitale Methoden einsetzen, sind im Normalfall skalierbar – wenn man eine Software benutzt, die die Häufigkeit von Begriffen in einem Dokument zählt, sollte es keinen Unterschied in der Anwendung machen, ob man eines oder einhundert Dokumente auswerten will. Würde man dasselbe per Hand tun, wäre man analog zum Anwachsen des Untersuchungskorpus mit der Auszählung beschäftigt. Digitale Werkzeuge ermöglichen es also unter anderem, Untersuchungen auf größere Mengen von Daten auszuweiten. Sie ermöglichen es auch, an einen solchen erweiterten Datensatz andere Fragen zu stellen, als dies mit einer kleineren Quellen-/Datengrundlage möglich wäre. Die vorherrschende Überlieferung historischer Quellen besteht aus Text, handgeschrieben, gemeißelt oder gedruckt – und durch die Möglichkeit, diesen mittels Texterkennung in computerlesbare Daten umzuwandeln, ergeben sich neue Perspektiven für die Arbeit von Historiker:innen: Wenn Texte als Daten verstanden werden, lassen sich aus Textquellen Datenbestände erstellen, die mithilfe quantitativer Methoden untersucht und ausgewertet werden können.¹

¹Kommandozeile/command line, bash, shell, prompt finden sich oft als synonym

Für die Literaturwissenschaften beispielsweise ist ein wichtiges Anwendungsfeld die Überprüfung von Autor:innenschaft: Ob ein anonym überliefertes Werk einem:namentlich bekannten Autor:in zugeschrieben werden kann, lässt sich entweder durch *close reading* von Literaturwissenschaftler:innen überprüfen, oder durch die Suche nach Mustern, nach quantifizierbare Eigenschaften eines Textes, wie beispielsweise die Häufigkeit von Funktionswörtern, Partikeln, Satzzeichen usw. Der unter dem Pseudonym Robert Galbraith veröffentlichte Kriminalroman *The Cuckoo's Calling* konnte mit entsprechender Software Joanne K. Rowling zugeschrieben werden – mit dieser dauerte das dreißig Minuten, was etwa dem Lesen von zwanzig Romanseiten entspricht. Zu einem Artikel, der diesen Fall thematisiert und in das Feld der linguistischen Forensik einbettet, die Straftäter:innen mithilfe quantitativer Textanalyse ermittelt, geht es hier. Ein Video zur Entwicklung und Anwendung von Software zur Zuschreibung von Autor:innenschaft finden Sie hier. Die genutzte Software, JGAP (Java Graphical Authorship Attribution Program), lässt sich offensichtlich auch für historische Analysen nutzen – man denke nur an Herrschaftssysteme, in denen Zensurpolitik herrscht(e) und viele Autor:innen daher nicht unter ihrem Klarnamen publizier(t)en. Durch eine Identifikation anonymer Schreiber:innen lassen sich weitere Aspekte rund um die Thematik Zensur untersuchen – welche Akteur:innen waren öffentlich bekannt, wer publizierten gleichzeitig anonym und unter Klarnamen, welche Autor:innen schrieben aus dem Exil, welche Netzwerke lassen sich rekonstruieren, usw. Dadurch, dass ein Programm durch quantitative Auswertungen die Kärnerarbeit der Identifikation abnehmen *kann* – um einen reflektierten Umgang mit Daten und Algorithmen geht es in Kapitel 3](03_digital_literacy_criticism.qmd) –, bleibt mehr Zeit für die qualitative Arbeit; gleichzeitig fußt die Analyse auf einem aussagekräftigen Datensatz, anstatt nur Einzelbeispiele beleuchten zu können.

Quantitative und qualitative Methoden sollen hier keinesfalls gegeneinander ausgespielt werden; vielmehr soll verdeutlicht werden, dass beide Herangehensweisen Vor- und Nachteile haben, und dass sie im besten Fall gewinnbringend miteinander kombiniert werden können – quantitative Auswertungen nur um ihrer selbst willen und ohne eine spezifische historische Fragestellung generieren kaum je einen Mehrwert.

Je nach Datengrundlage, Analysezzweck und Forschungsfrage bieten sich unterschiedliche Tools zur Nutzung an; für die meisten Forschungsvorhaben bis zum Ende des Studiums dürfte existierende Software ausreichen, sei es für die Akquise und Aufbereitung von Daten(-sätzen), für verschiedene Arten von Textanalyse, statistische Auswertungen, Netzwerkanalysen, Geomapping oder Visualisierungen. Eine Auswahl an Tools – alle free/open source – für spezifische

genutzte Begriffe für Command Line Interfaces. Auf UNIX-basierten Betriebssystemen wie Mac OS und Linux ist das Terminal als Interface weit verbreitet; für Details: https://en.wikipedia.org/wiki/Command-line_interface#History. Windows-nutzer:innen kommen mit der Powershell ganz gut zurecht, es empfiehlt sich eventuell die Installation von Cygwin oder MinGW, um mit einem UNIX-basierten Interface arbeiten zu können.

Analysen findet sich unter Further Ressources. Für gewisse Analysen bietet es sich an, Programmierkenntnisse zu erwerben – das Erstellen eigener Skripts beinhaltet die umfassende Kontrolle darüber, wie Daten eingelesen, aufbereitet, angereichert, analysiert und visualisiert werden; bei wiederkehrenden Prozessen, die händisch einige Arbeitszeit in Anspruch nehmen würden, lässt sich so zusätzlich Zeit sparen. Für geisteswissenschaftliche Projekte werden zurzeit vor allem zwei Programmiersprachen genutzt, R und Python. Da sich beide großer Beliebtheit in den Humanities erfreuen, existieren mittlerweile zahlreiche Packages, die Data und Text Mining sehr einfach machen. Solche Packages für Programmiersprachen kann man sich wie Plug-Ins für Programme vorstellen, beispielsweise ein AdBlocker für den Browser. So etwas ist nativ von den Entwickler:innen nicht vorgesehen, aber jemand hatte Bedarf, Anzeigen zu blockieren, hat hierzu ein Programm geschrieben und es der Community zur Verfügung gestellt. Der Unterschied zu einem Package ist, dass dieses verschiedenen Funktionen zur Verfügung stellt - auswählen und ausführen müssen die Anwender:innen. Wer in Schule und Studium keine Berührungspunkte mit Programmieren hatte, wird zu Beginn vielleicht größere Berührungängste haben – und nochmal: Sie müssen nicht programmieren können, um quantitativ zu arbeiten. Speziell für Historiker:innen ohne Programmier-Vorkenntnisse ist das Projekt The Programming Historian, das seit 2008 zahlreiche Tutorials veröffentlicht, um verschiedene Tools, Techniken und Workflows für Forschung und Lehre vorzustellen.

2.2 Digitale Tools zur Kommunikation

tbd

2.3 Digitale Elemente in der Hochschullehre

tbd

2.4 Projekte und Ressourcen

2.4.1 Alte Geschichte

Projekte:

Ressourcen/Portale:

2.4.2 Mittelalter und Frühe Neuzeit

Projekte:

Ressourcen/Portale:

- dMGH: Monumenta Germaniae Historica online (Beta-Version)

- e-codices: Virtuelle Handschriftenbibliothek der Schweiz
- Fragmentarium: Laboratory for Medieval Manuscript Fragments
- Handschriftenportal: Zentraler nationaler Nachweis für Buchhandschriften in deutschen Bibliotheken und in deutscher Sprache (Entwicklungsstadium)
- e-manuscripta: Digitalisierte handschriftliche Quellen aus Schweizer Bibliotheken und Archiven
- e-rara: Plattform für digitalisierte Drucke aus Schweizer Institutionen
- Gallica: Digitalisierte Quellen aus französischen Bibliotheken

2.4.3 Moderne und Zeitgeschichte

Projekte:

- Refugee History: Wissenschaftliches Blog und interaktives Netzwerk zu aktuellen Debatten um das Thema “Flüchtlinge”

Ressourcen/Portale:

- Datenbank Bild + Ton zur Geschichte (Schweizer) sozialer Bewegungen
- Dodis: Wissenschaftliche Edition von Dokumenten zur Schweizer Außenpolitik
- e-newspaperarchives.ch: Schweizer Zeitungen online
- e-periodica: Schweizer Zeitschriften online
- Historische Statistik der Schweiz (HSSO)
- histat: Zeitreihen zur Historischen Statistik

2.4.4 Jüdische Geschichte

Projekte:

- Digital Jewish Studies Online, Stroum Center for Jewish Studies, University of Washington

Ressourcen/Portale:

- Menny, Anna; Rürup, Miriam; Siegel, Björn: Jüdische Geschichte im deutschsprachigen Raum, in: Busse, Laura u. a. (Hg.): Clio-Guide. Ein Handbuch zu digitalen Ressourcen für die Geschichtswissenschaften, Berlin 2018, S. E.2-1–E.2-56. Online: <https://doi.org/10.18452/19244>.

2.4.5 Geschichte Afrikas

Projekte:

- Emandulo: A digital archive which convenes and re-assembles archival/museal collections and presentations on southern African history before colonialism (the last 500 years)

Ressourcen/Portale:

- FHYA: Experimental digital research platform on southern African history before colonialism (the last 500 years)

2.4.6 Osteuropäische Geschichte

Projekte:

Ressourcen/Portale:

2.4.7 Epochen-/Areaübergreifend:

Projekte:

Ressourcen/Portale:

- Around DH in 80 days

Chapter 3

Digital Literacy, Digital Criticism

Unter Data Literacy wird die Kompetenz verstanden, Daten zu sammeln, zu managen, zu evaluieren und zu nutzen,¹ eine Kompetenz, die jeder für den mittlerweile unvermeidlichen Umgang mit Daten verschiedenster Art im eigenen Alltag entwickeln sollte. Je nach Forschungsdisziplin ergeben sich weitere gewisse Spezifika, wobei Studierenden der Geisteswissenschaften ein Thema wie Algorithmenkritik nicht als erstes in den Sinn kommt, wenn es um die im Studium zu erwerbenden Kompetenzen geht.² Aber auch ohne den Quellcode von machine-learning-Software im Detail zu verstehen, ermöglicht ein grundlegendes Verständnis von und ein Wissen über die Funktionsweisen solcher Anwendungen einen reflektierten Umgang mit diesen; eine solche Art von digital bzw. data literacy ist vor allem dann relevant, wenn es um die Interpretation von Ergebnissen geht, die scheinbar objektiv sind, bzw. objektiv entstanden. Ein gutes Beispiel hierfür sind die Ergebnislisten bei Suchanfragen in einer Suchmaschine. Je nachdem, welchen Anbieter Sie nutzen, spielen verschiedene Umstände in die Generierung von Trefferlisten hinein, beispielsweise Ihre Suchhistorie, sodass search neutrality nicht mehr gewährleistet ist.³

¹Ridsdale, Chantel; Rothwell, James; Smit, Mike u. a.: Strategies and Best Practices for Data Literacy Education Knowledge Synthesis Report, 2015, S. 8. Online: <<https://doi.org/10.13140/RG.2.1.1922.5044>>.

²Kommandozeile/command line, bash, shell, prompt finden sich oft als synonym genutzte Begriffe für Command Line Interfaces. Auf UNIX-basierten Betriebssystemen wie Mac OS und Linux ist das Terminal als Interface weit verbreitet; für Details: https://en.wikipedia.org/wiki/Command-line_interface#History. Windowsnutzer:innen kommen mit der Powershell ganz gut zurecht, es empfiehlt sich eventuell die Installation von Cygwin oder MinGW, um mit einem UNIX-basierten Interface arbeiten zu können.

³Alternative Suchmaschinen sind zum Beispiel [Startpage](<https://www.startpage.com/>) oder [DuckDuckGo](<https://duckduckgo.com/>).

Gehen Sie auf die Bilder-Suche von Google und suchen Sie nach “historian”. Was sehen Sie?

Wüsste ich nichts über Geschichtswissenschaftler:innen, würde ich aufgrund der Ergebnisse meiner Suche davon ausgehen, “a historian” wäre meist ein alter, weißer Mann mit Brille, Bart und einem großen Bücherregal; wenn Sie sich am Departement Geschichte der Uni Basel umsehen, dürfte ein etwas anderer Eindruck entstehen. Die Ergebnisse von Suchmaschinen, die für ihr Funktionieren Algorithmen anwenden, sind biased: Sie beruhen auf vorangegangenen Suchen, Vorlieben, geographischem Standort – und auf von Menschen eingegebenen Metadaten. Ein Bewusstsein hierfür und das Hinterfragen von Datensätzen gehören also mit zur Arbeit in einer digitalisierten Welt.

3.1 Digital Criticism, Data Criticism

Digitalisierte Quellen ebenso wie rein digitale erfordern eine erweiterte Art von Quellenkritik – im Einführungskurs an der Universität Basel lernen Sie die Grundlagen:

Woher kommt eine Quelle, wer hat sie unter welchen Umständen und zu welchem Zweck erstellt? Welche Absichten können darin verborgen sein, und welche Verzerrungen können sich durch sie ergeben?

Welche Tendenzen könnten sich in hochmittelalterlichen Herrscherchroniken verstecken, wenn der Verfasser in direkter Abhängigkeit des Auftraggebers stand? Wie sind Zeug:innenaussagen in Hexenprozessen zu bewerten, die unter Anwendung körperlicher Strafe entstanden sind? Mit wie viel Vorsicht sind die Inhalte eines Tagebuchs zu bewerten, das allem Anschein nach mit Blick auf eine spätere Veröffentlichung verfasst wurde?

Neben der inneren Kritik geht es bei der Arbeit mit Quellen immer auch um Fragen der Korpusbildung: Wie kann eine Quellengrundlage erstellt werden, die für Beantwortung einer spezifischen historischen Fragestellung belastbar und aussagekräftig genug ist und gleichzeitig in angemessener Zeit bearbeitet werden kann? Hinzu kommen Spezifika bei der Arbeit mit unterschiedlichen Quellenformen bzw. -formaten: Bei analogen Quellen, die auch in digitaler Form zur Verfügung stehen, besteht die Gefahr, dass ein Thema, ein Bereich, ein Aspekt vernachlässigt wird, wenn nur die unmittelbar verfügbaren, digitalisierten Bestände zur Korpusbildung genutzt werden. Wenn Sie sich beispielsweise für die Schweizer Historikerin und Frauenrechtlerin Meta von Salis (1855–1929) und deren briefliche Korrespondenz – Friedrich Nietzsche war einer ihrer Brieffreunde – interessieren und über die Suchplattform für historische Schweizer Bestände, swisscollections, in nationalen Bibliotheken und Archiven nach entsprechenden Dokumenten suchen, erhalten Sie 361 Treffer:

Digital verfügbar waren hiervon im Oktober 2022 lediglich 3 Einträge, wobei der erste ein Brief von Nietzsche an Meta von Salis ist, der zweite Eintrag umfasst sieben Briefe von Caroline Farner, und der dritte Eintrag ist weder an noch von

Figure 3.1: Erweiterte Suchmaske von swisscollections

Figure 3.2: Suchergebnisse für “Meta von Salis” + “Brief”

Meta von Salis, sondern hat sie nur zum Thema:

Ihnen würde bei einer Korpuserstellung vom Schreibtisch aus also der Großteil der Überlieferung fehlen, und Ihre Untersuchungsergebnisse wären wohl sehr verzerrt, würden Sie statistische Aussagen treffen wollen: Meta von Salis unterhielt brieflichen Kontakt zu einem Mann und einer Frau, das Geschlechterverhältnis wäre also ausgeglichen; und Frauen schreiben im Schnitt mehr Briefe an Meta von Salis als Männer. Beim Blick auf alle Suchergebnisse würden sich Ihre Aussagen aber sehr ändern, und es würde sich lohnen, diese Verzerrung, diesen Bias aus Ihrer Datengrundlage zu entfernen.

Hinzu kommt natürlich immer das grundlegende Problem bei der Suche nach Quellen: swisscollections und ähnliche Portale können nur anzeigen, was die Kooperationspartner:innen zur Verfügung stellen. Hat eine Bibliothek Briefe von Meta von Salis in ihrem Bestand, diese aber noch nicht als Datensatz erfasst, wissen Sie im Gegensatz zum obigen Beispiel nicht einmal, dass Ihnen etwas entgehen würde, dass in Ihrem Korpus überhaupt ein Bias vorhanden ist.

Ähnliche Vorsicht zur Vermeidung von Verzerrungen in der Datengrundlage gilt bei der Arbeit mit rein digitalen Daten, beispielsweise bei der Auswertung von Datensätzen aus Befragungen. Wenn Sie sich am 27.10.2022 vor die Universitätsbibliothek in Basel stellen und einen Tag lang mithilfe eines kurzen Fragebogens und einer Tabellendatei erfassen, wie zufrieden die befragten Personen mit dem Essen in der Unimensa sind, werden Sie am Ende einen Datensatz erhalten, in


Stichwortsuche Indexsuche Durchstöbern Archivtektonik Bibliographien

swisscollections [Erweiterte Suche bearbeiten](#) [Neue erweiterte Suche starten](#) [Neue einfache Suche starten](#)


Digitalisat verfügbar ☒ Format Brief ☒

3 Suchergebnisse für: (Alle Felder:meta von salis)


Seite: [←](#) 1 von 1 [→](#) Sortieren nach Relevanz



Brief an Meta von Salis-Marschlins
von Friedrich Nietzsche
 Nietzsche, Friedrich
 1887.09.01-14
 Basel, UB, UBH NL 61 : N. II 10 IV



7 Briefe an Meta von Salis-Marschlins
von Caroline Farner
 Farner, Karoline
 1893.05.19-1895.04.07
 Basel, UB, UBH NL 61 : S. II C. 2 1a-1b u.a.



Brief an Prof. Dr. Richard Oehler
von der Universitätsbibliothek Basel
 Universitätsbibliothek Basel
 25. Mai 1937
 Dokument=Item=Pièce
 Basel, UB, UBH NL 53 : B III 1, Beil. 3

Seite: [←](#) 1 von 1 [→](#)

☒ Digitalisat verfügbar

Ressourcentyp ^

+ ☐ Archivmaterial (3) x

+ ☐ Handschrift (3) x

Bibliothek ^

☐ Basel, UB (3) x

Person / Körperschaft ^

+ ☐ Salis-Marschlins, Meta von (18... (2) x

+ ☐ Farner, Karoline (1842-1913) (1) x

+ ☐ Nietzsche, Friedrich (1844-1900) (1) x

+ ☐ Oehler, Richard (1878-1948) (1) x

+ ☐ Universitätsbibliothek Basel (1) x

Sprache ^

☐ Deutsch (3) x

Ort ^

☐ Basel (1) x

Figure 3.3: Suchergebnisse für “Meta von Salis” + “Brief” + “Digitalisat verfügbar”

dem sich vermutlich über 80% der Befragten für besseres und nahezu 100% für günstigeres Essen in der Mensa aussprechen – eine gute Schlagzeile für die BZ, die sich auf die neuesten Ergebnisse einer wissenschaftlichen Studie berufen kann. Führen Sie die gleiche Umfrage eine Woche später, mitten während der Herbstmesse durch, werden die Ergebnisse wohl erheblich anders aussehen. Die Wahrscheinlichkeit, dass die Mensa infolge der BZ-Schlagzeile innerhalb weniger Tage den Menüplan überarbeitet und die Preise herabgesetzt hat, ist dabei wohl geringer als diejenige, dass sich Ihr Sample, die Auswahl an Datenpunkten, also befragten Personen, durch die Messe stark verändert hat: Im Umkreis der Bibliothek treffen Sie nun nicht mehr vor allem Studierende und andere Uni-Angehörige an, sondern auch Messebesucher:innen vom Petersplatz. Auch hier sind Verzerrungen entstanden, ähnlich wie beim vorherigen Beispiel mit den Briefen: Wenn aus einer Gesamtheit nur eine spezifische Untermenge beobachtet wird, die sich durch ein gemeinsames Merkmal von der Gesamtheit unterscheidet – digitalisierte Quelle oder Besucher:in der Universitätsbibliothek –, ist die Datengrundlage und damit die Untersuchungsergebnisse biased. Um bei Daten, die Sie nachnutzen, eventuell vorhandene Verzerrungen nicht weiterzutransportieren, ist das Üben von Datenkritik eine essentielle Kompetenz.

Zur Tatsache, dass Daten eben nicht gegeben sind (lat. dare, datum: gebene, gegeben), sondern gemacht und daher entsprechend interpretiert werden müssen, finden Sie ein gutes Interview von Roopika Risam;⁴ zur Zementierung von Klischees durch Übersetzungsalgorithmen gibt es einen Artikel in der Republik von Marie-José Kolly und Simon Schmid;⁵ und über die Macht von Data Science und dem Änderungspotential von Data Feminism haben Cathherine D'Ignazio und Lauren F. Klein ein ganzes Buch veröffentlicht.⁶

Zur Frage, wie sich die digitale Wende, der digital turn, auf die Quellenkritik auswirkt, sehen Sie sich dieses kurze Video des Projekts Ranke.2 – Quellenkritik im digitalen Zeitalter an⁷:

Eine Handreichung zum Umgang mit digitalisierten und digitalen Daten, das im selben Projekt erarbeitet wurde, finden Sie hier.

⁴Risam, Roopika: "It's Data, Not Reality": On Situated Data With Jill Walker Rettberg, 06.2020. Online: <<https://medium.com/nightingale/its-data-not-reality-on-situated-data-with-jill-walker-rettberg-d27c71b0b451>>, Stand: 16.08.2022.

⁵Kolly, Marie-José; Schmid, Simon: Sie ist hübsch. Er ist stark. Er ist Lehrer. Sie ist Kindergärtnerin, in: Republik, 04.2021. Online: <<https://www.republik.ch/2021/04/19/sie-ist-huebsch-er-ist-stark-er-ist-lehrer-sie-ist-kindergaertnerin>>, Stand: 23.08.2022.

⁶D'Ignazio, Catherine; Klein, Lauren F.: Data feminism, 2020. Online: <<https://direct.mit.edu/books/book/4660/Data-Feminism>>.

⁷Eine Selbstlerneinheit zum Thema finden Sie hier: <https://ranke2.uni.lu/u/archival-digital-turn/>.

Chapter 4

Datenerhebung und -analyse

1. Daten aufbereiten und sichern digitalisieren übersetzen und transkribieren aufarbeiten anonymisieren
2. dokumentieren

Wie wurden die Daten erhoben? Wie wurden die Daten annotiert? In welchem Format sind die Daten vorhanden? Welche Software wurde benutzt? In welchem Zusammenhang stehen diese Daten zu anderen Daten? Was stellen die Daten dar? Hintergrundgeschichte der Daten (Zweck, Methode, etc.)?

3. Daten auswerten

Sie interpretieren und visualisieren die Daten, halten Ihre Forschungsergebnisse fest und zitieren die Datenquellen. Zudem bereiten Sie die Daten auf ihre Archivierung vor

Chapter 5

FAIR und CARE

Bereits beim Beginn eines Projekts, sei es eine Proseminararbeit oder ein kollaboratives Großprojekt, sollten Fragen nach Interoperabilität, Sicherung und Nutzbarkeit von Forschungsdaten gestellt werden. Denn oftmals enden Projekte, ohne dass erstellte Daten für anschließende Forschungen verfügbar gemacht werden, sei es, weil nicht rechtzeitig nach Lösungen zur langfristigen Speicherung gesucht wurde, sei es, weil Daten in einer Form erhoben und gespeichert wurden, die eine Nachnutzung erschwert oder auch unmöglich macht.

Zu Beginn des Studiums stehen solche Fragen weniger im Fokus; dennoch sollen diese hier kurz thematisiert werden, um dafür zu sensibilisieren; auch, weil sie den Prozess der Datenerhebung beeinflussen.

Die Prinzipien FAIRer Daten wurden 2016 von einem Konsortium aus Wissenschaftler:innen und Organisationen wie folgt definiert:¹ **F**indability, **A**ccessibility, **I**nteroperability, **R**euse of digital assets.

¹Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, IJsbrand Jan u. a.: The FAIR Guiding Principles for scientific data management and stewardship, in: Scientific Data 3 (1), 03.2016, S. 160018. Online: <<https://doi.org/10.1038/sdata.2016.18>>, Stand: 09.11.2022.

Chapter 6

shell

Shell 101

Es gibt zwei Arten, um mit dem Computer zu interagieren bzw. ihn zu nutzen: über ein **G**raphical **U**ser **I**nterface (GUI) oder, etwas direkter, über die Kommandozeile¹. Um eine Datei “Bild1.jpg” im Ordner “Bilder” zu löschen, öffnet man den Explorer (Windows) oder den Finder (Mac), klickt sich zum Ordner “Bilder”, macht einen Rechtsklick auf die zu löschende Datei “Bild1.jpg”, klickt “In den Papierkorb legen” oder zieht die Datei mit der Maus direkt dorthin. Man kann dieselbe Aktion als Kommando eintippen: Man öffnet eine Powershell (Windows) oder das Terminal (Mac), navigiert mit Texteingabe zum entsprechenden Ordner, bspw. `cd Dokumente/Bilder` und gibt dort das Kommando `rm "Datei.jpg"` ein, das mit der Entertaste ausgeführt wird.

```
(base) serina00@dg-19-mac-02 Bilder % rm "Datei.jpg"
```

Die beiden Vorgehensweisen unterscheiden sich dabei in drei Punkten:

1. Das Kommando `rm` ist endgültig, die Datei ist ohne Übergangszeit im Papierkorb gelöscht.
2. Das Kommando lässt sich relativ simpel auf eine Vielzahl von Dokumenten anwenden, wobei ganz unterschiedliche Bedingungen beachtet werden können, und es lässt sich mit anderen Kommandos verbinden.
3. Terminal sieht k3wl aus.

Bevor wir den zweiten – und für unsere Arbeit hilfreichsten – Unterschied genauer anschauen, kurz zur Kommandozeile:

¹Kommandozeile/command line, bash, shell, prompt finden sich oft als synonym genutzte Begriffe für Command Line Interfaces. Auf UNIX-basierten Betriebssystemen wie Mac OS und Linux ist das Terminal als Interface weit verbreitet; für Details: https://en.wikipedia.org/wiki/Command-line_interface#History. Windows-nutzer:innen kommen mit der Powershell ganz gut zurecht, es empfiehlt sich eventuell die Installation von Cygwin oder MinGW, um mit einem UNIX-basierten Interface arbeiten zu können.

In einem Terminal/Shell können Befehle/Programme ausgeführt werden, die auf der Datenstrukturebene stattfinden – wie beispielsweise das Löschen einer Datei, `rm Dateiname.xyz`, oder das Erstellen eines Ordners, `mkdir NeuerOrdner`. Oder aber Operationen auf Dateninhaltsebene – wie beispielsweise das Suchen eines bestimmten Begriffs in einer Textdatei, `grep Begriff Textdatei.txt`, oder das Auszählen mehrerer Begriffe und das Speichern des Ergebnisses in einer neuen Datei, `grep -Ec "Begriff1|Begriff2" Textdatei.txt > Ergebnisse.txt`.

Woher weiss Ihre Shell aber, was sie ausführen soll, wenn Sie `rm` oder `grep` eintippen? Es gibt zahlreiche Shell-Programme, die bereits auf Ihrem System vorinstalliert sind, und mit denen Sie vieles tun können – öffnen Sie Ihre Shell und tippen Sie `date` ein: Das aktuelle Datum mit Uhrzeit erscheint. Ihre Shell sucht nach dem ersten Argument, dem Befehl, im Filesystem des Computers, und wenn sie fündig wird, führt sie eine Aktion mit den entsprechenden Parametern aus.

tmi: Wenn Sie `echo $PATH` im Terminal eingeben, sehen Sie eine Auflistung der Orte, an denen nach Befehlen gesucht wird. Tippen Sie `which date` ein und drücken Sie ‘Enter’, um zu sehen, wo das Programm “date” in Ihrem Computer liegt.

Wenn Sie einen Befehl eintippen, den es nicht gibt bzw. für den es noch kein installiertes Programm auf Ihrem Computer gibt, bekommen Sie eine simple Fehlermeldung:

```
(base) serina00@dg-19-mac-02 EK-dh % nonsense
command not found: nonsense
```

Das aktuelle Datum wird Ihnen wahrscheinlich auch in Ihrer Toolbar angezeigt, und einen neuen Ordner können Sie per Rechtsklick erstellen, dazu brauchen Sie das Terminal nicht unbedingt. Um einen Begriff in einem Textdokument zu finden und alle Vorkommen zu zählen, können Sie das Dokument öffnen, `Strg-F` drücken, den Begriff eingeben und das Ergebnis sehen. Wenn Sie nach mehreren Begriffen suchen wollen, müssen Sie dieselbe Aktion zweimal ausführen: `Strg-F`, Begriff 2.

Wenn Sie wissen möchten, wie häufig Arthur Dent, Zaphod Beeblebrox, Slartibartfast und Mrs Enid Kapelsen in den sechs Bänden von “The Hitchhiker’s Guide to the Galaxy” genannt werden, müssen Sie, wenn Sie den Volltext heruntergeladen haben, für jeden Namen eine Suche ausführen, mit `Strg-F`. Bei der Suche nach Personen mit Vor- und Nachnamen wie “Arthur Dent” suchen Sie nach “Arthur”, nach “Dent” und nach “Arthur Dent” und ziehen alle “Arthur Dent”-Treffer von den übrigen Treffern ab, um nichts doppelt zu zählen und ihre Suchergebnisse zu verfälschen. Die Trefferzahl all Ihrer Suchanfragen schreiben Sie in ein neues Dokument.

Sie können dasselbe auch mit dem Terminal machen und einige der Built-in-Programme nutzen: Sie bewegen sich mit `cd`, `change directory`, in den Ordner

(directory), in dem Ihre Textdatei liegt:

```
(base) serina00@dg-19-mac-02 serina00 % cd Documents/progr/EK_dh/Hitchhiker
```

(Um zu prüfen, was dort liegt, können Sie im Terminal `ls` (für `list`) eingeben, bzw. in der Powershell `dir` (für `directory`))

```
(base) serina00@dg-19-mac-02 Hitchhiker % ls hitchhiker_fulltext
```

Mit einer Zeile können Sie die in einem Texteditor ausgeführten Suchvorgänge mit `grep` (Global Regular Expression Print) ausführen und die Ergebnisse mit `>` in eine neue Datei schreiben:

Appendix A

Appendix

API	A pplication P rogramming I nterface: a facility offered by a web resource which allows search queries independent of a GUI , often performed using scripts
bash	default program that runs in the command line
big data	huge amount of data, identifiable through repeated freezing of your standard program when opening a file
born digital data	data which originated in a digital form
CLI	C ommand L ine I nterface, text interface that allows interaction with the computer; see also bash
CMS	C ontent M anagement S ystem
Console	See CLI
Crowdsourcing	projects that include the active participation of the public to generate content, transcribe sources etc.
csv	c omma s eparated v alues, a structured text format, using commas as separators between columns
distant reading	quantitative approach to huge amounts of texts, using computational methods to search for interpretable patterns
GUI	G raphical U ser I nterface
HTML	H ypertext M arkup L anguage, a structured text format, like the format this guide is written in, to render documents in a browser
Jupyter notebook	web application/interactive coding environment that runs in a browser; let's you create and share code (https://jupyter.org)
machine read-able	transformation of, for example, text into a data format that is processable by a computer

OCR	O ptical C haracter R ecognition, process of transforming text on an image into a data format
OS	O perating S ystem
OSS	O pen S ource S oftware
Regular Expres- sion	syntax for search and replace text using patterns (instead of exact matches)
terminal	See CLI

Appendix B

Further Ressources

- Brennan, Sheila A.: Digital History, in: The Inclusive Historian's Handbook, <https://inclusivehistorian.com/digital-history/>, 04.06.2019.
- Cohen, Daniel J.; Rosenzweig, Roy: Digital History. A Guide to Gathering, Preserving, and Presenting the Past on the Web, Philadelphia 2006. Online: <https://chnm.gmu.edu/digitalhistory/>.
- Hohls, Rüdiger: Digital Humanities und digitale Geschichtswissenschaften, in: Busse, Laura u. a. (Hg.): Clio-Guide. Ein Handbuch zu digitalen Ressourcen für die Geschichtswissenschaften, Berlin 2018, S. A.1-1–B.1-34. Online: <https://doi.org/10.18452/19244>.
- Winters, Jane: Digital history, in: Tamm, Marek; Burke, Peter (Hg.): Debating New Approaches to History, London 2019, S. 277–300.
- Digital history, in: Wikipedia, 07.09.2022. Online: https://en.wikipedia.org/w/index.php?title=Digital_history&oldid=1109027465, Stand: 02.11.2022.

B.1 Digital Literacy, Digital Criticism

- Ekström, Andreas: The Moral Bias behind your Search Results, TED talk 7.12.2015 (9:18), Online: https://www.youtube.com/watch?v=__vBggxCNNno.
- Gibbs, Frederick W.: New Forms of History: Critiquing Data and Its Representations, in: The American Historian, February 2016. Online: <http://tah.oah.org/february-2016/new-forms-of-history-critiquing-data-and-its-representations/>.
- Tavani, Herman; Zimmer, Michael Zimmer: Search Engines and Ethics, in: Edward N. Zalta (Hg.): The Stanford Encyclopedia of Philosophy (Fall

2020 Edition), Online: <https://plato.stanford.edu/archives/fall2020/entries/ethics-search/>, Kap. 3.1.

B.2 Tools für digital history (n.b.: free/open source)

B.2.1 Allgemein

- Programming Historian: Tutorials zu verschiedenen Tools und Methoden für historische Forschung und Lehre

B.2.2 Text-/Korpusanalyse

- AntConc: Korpusanalyse-Toolkit
- Voyant-Tools: Sammlung von Tools zur Textanalyse, browserbasiert oder standalone

B.2.3 Visualisierung

- FID Romanistik: Sammlung von Tools zur Datenvisualisierung

B.3 Terminal/Command Line

- Dawson, Ted: Introduction to the Windows Command Line with PowerShell, Programming Historian 5 (2016), <https://doi.org/10.46430/phen0054>. (self-learning lesson)
- MIT Computer Science Department: 1-hour-lecture on the Shell (video)
- Milligan, Ian; Baker, James: Introduction to the Bash Command Line, Programming Historian 3 (2014), <https://doi.org/10.46430/phen0037>. (self-learning lesson)
- datacamp course: Introduction to Shell (interactive self-learning lesson)
- Jeroen Janssens: Data Science at the command line (book)

B.4 Regular Expressions

- Knox, Doug: Understanding Regular Expressions, Programming Historian 2 (2013), <https://doi.org/10.46430/phen0033>. (self-learning lesson)
- RegexOne: Learn Regular Expressions with simple, interactive exercises. (interactive self-learning tutorial)

Blaney, Jonathan; Winters, Jane; Milligan, Sarah u. a.: Doing digital history: a beginner's guide to working with text as data, Manchester 2021 (IHR research guides).

D'Ignazio, Catherine; Klein, Lauren F.: Data feminism, 2020. Online: <<https://direct.mit.edu/books/book/4660/Data-Feminism>>.

Kolly, Marie-José; Schmid, Simon: Sie ist hübsch. Er ist stark. Er ist Lehrer. Sie ist Kindergärtnerin, in: Republik, 04.2021. Online: <<https://www.republik.ch/2021/04/19/sie-ist-huebsch-er-ist-stark-er-ist-lehrer-sie-ist-kindergaertnerin>>, Stand: 23.08.2022.

Le Roy Ladurie, Emmanuel: La fin des érudits, in: Le Nouvel Observateur, 08.1968.

Ridsdale, Chantel; Rothwell, James; Smit, Mike u. a.: Strategies and Best Practices for Data Literacy Education Knowledge Synthesis Report, 2015. Online: <<https://doi.org/10.13140/RG.2.1.1922.5044>>.

Risam, Roopika: "It's Data, Not Reality": On Situated Data With Jill Walker Rettberg, 06.2020. Online: <<https://medium.com/nightingale/its-data-not-reality-on-situated-data-with-jill-walker-rettberg-d27c71b0b451>>, Stand: 16.08.2022.

Romein, C. Annemieke; Kemman, Max; Birkholz, Julie M. u. a.: State of the Field: Digital History, in: History 105 (365), 04.2020, S. 291–312. Online: <<https://doi.org/10.1111/1468-229X.12969>>, Stand: 15.09.2022.

Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, IJsbrand Jan u. a.: The FAIR Guiding Principles for scientific data management and stewardship, in: Scientific Data 3 (1), 03.2016, S. 160018. Online: <<https://doi.org/10.1038/sdata.2016.18>>, Stand: 09.11.2022.

