

Malte Rehbein* und Marlene Ernst

Erschließung handschriftlicher Dokumente zwischen Fachwissen, Citizen Science und KI

<https://doi.org/10.1515/bfp-2023-0043>

Zusammenfassung: Der Beitrag erörtert Spannungsfelder, die sich aus der fortwährenden digitalen Transformation der Gesellschaft und dem „ChatGPT-Schock“ derzeit neu formieren. Anhand des Beispiels der Erschließung handschriftlicher Dokumente werden diese Felder analysiert und es wird versucht, die verschiedenen sich ändernden Rollen und Verständnisse auszuloten. Dies umfasst die Anteilseigner aus Bibliothek als traditionellem Informationsdienstleister, Plattformbetreiber als Mediatoren, Wissenschaft wie auch Gesellschaft, wobei sich die Technik zunehmend vom passiven Werkzeug zum handelnden Akteur wandelt. Aus diesen Überlegungen heraus wird eine Zukunftsperspektive einer prozess-orientierten Wissensschaffung abgeleitet, zu der alle genannten Akteure als „Co-Produzenten“ beitragen.

Schlüsselwörter: Big Data, Partizipatives Forschen, künstliche Intelligenz, Metadaten, Verdatung

Manuscript Exploration between Expert Knowledge, Citizen Science, and AI

Abstract: The paper discusses fields of tension that are currently reshaping themselves from the ongoing digital transformation of society and the “ChatGPT shock”. Using the example of manuscript exploration, these fields are analysed, and an attempt is made to explore the various changing roles and understandings. This includes the stakeholders from library as traditional information service provider, platform operators as mediators, academia as well as society, where technology is increasingly changing from a passive tool to an acting participant. Based on these considerations, a future perspective of process-oriented knowledge creation is derived, to which all the actors contribute as “co-producers”.

Keywords: big data, participatory research, artificial intelligence, meta data, datafication

1 Hoffnungsvolle Zukunft?

Trotz Einhaltung aller nur erdenklichen gängigen Standards ist das Überleben digitaler Produkte im Umfeld von Institutionen des kulturellen Gedächtnisses eine weiterhin viel diskutierte Frage. Die jüngere Vergangenheit hat gezeigt, dass noch häufig große Abhängigkeiten der Verfügbarkeit solcher Produkte von Projektlaufzeiten und -mitteln wie auch dem Engagement und der Verfügbarkeit einzelner Forscher:innen bestehen. Um dem entgegenzuwirken, sind projektübergreifende Lösungen dann besonders zielführend und langlebig, wenn klare Rahmenbedingungen inklusive einheitlicher Standards und persistenter Referenzierungen angeboten werden. Sind zusätzlich noch Beteiligungsmöglichkeiten für Bürger:innen wie auch Open-Access-Strategien integriert, lassen sich Anforderungen von institutioneller Seite, Wissenschaft und Öffentlichkeit miteinander verbinden.

Dieser Beitrag¹ diskutiert am Beispiel der bibliothekarischen und fachwissenschaftlichen Erschließung von handschriftlichen Dokumenten als elementare Quellen historischer Arbeit das Wechselspiel und Spannungsfeld zwischen den Gedächtnisinstitutionen einerseits, die diese Handschriften bewahren und (digital) zugänglich machen (v. a. Archive, Bibliotheken, Museen) und andererseits deren Nutzer:innen aus der Forschung. Als zentrale Schnittstelle zwischen diesen beiden Akteuren werden Erschließungsdaten betrachtet, die Objekte überhaupt erst auffindbar und damit nutzbar machen. Stellvertretend für die Perspektive der (historischen) Forschung sei dabei der französische Historiker Marc Bloch (1886–1944) zitiert, dessen Werkzeug-Metaphorik besonders anschaulich ist. Für Bloch sind Quellen „die Dokumente, die er für nötig hält“, ² und er bezeichnet diese als seine Arbeitsinstrumente. Er beklagt dabei um 1940 einen Mangel an solchen Instrumenten, den er als „Lücke [s]eines Rüstzeugs“ bezeichnet und folgert, dass diese Lücke „unnötige Arbeit [...] kosten wird“. ³ Er führt dies – in seiner Zeit – auf die „Willkür einzelner oder

*Kontaktperson: Malte Rehbein, malte.rehbein@uni-passau.de
Marlene Ernst, marlene.ernst@stadt-salzburg.at

1 Der Beitrag basiert auf dem Festvortrag „Auf dem Weg zu Big Data“ von Malte Rehbein anlässlich des 10-jährigen Jubiläums der Plattform e-manuscripta.ch am 13.3.2023 an der Universität Zürich.

2 Bloch (2008) 79.

3 Ebd.

Knausrigkeit einiger schlecht informierter Verlagshäuser“,⁴ denen man Entscheidungen über die Erschließung bzw. deren Publikation überließe, zurück. Bloch fordert daraufhin: „Eine Gesellschaft, die vorgibt, die Wissenschaften hochzuhalten, dürfte deren Arbeitsinstrumenten nicht gleichgültig gegenüberstehen.“⁵ Deutlich wird hier eine Sichtweise, die denjenigen, die diese Arbeitsinstrumente benötigen (Forscher:innen), eine andere Rolle und geringere Einflussmöglichkeiten einräumt als denjenigen, die die Arbeitsinstrumente zur Verfügung stellen (Infrastrukturanbieter).

Hintergrund für einen positiveren Ausblick als die Momentaufnahme Blochs hinsichtlich diesbezüglicher Zukunftsperspektiven sind entsprechende, wie eingangs bereits beschriebene, Rahmenbedingungen. Sie bilden das Fundament, um auch ein digitalisiertes kulturelles Erbe auf Dauer zu erhalten und permanent zugänglich zu machen. Auch wenn Nachhaltigkeit digitaler Angebote meist eher unter technischen Gesichtspunkten betrachtet wird, scheinen die entscheidenden Kriterien für den langfristigen Erhalt der Angebote zu sein: die Attraktivität ihrer Inhalte sowie eine konstante wissenschaftliche Nutzung nebst gesellschaftlicher Aufmerksamkeit. Denn was gebraucht wird, geht weniger schnell verloren als Unbenutztes. Dieses Phänomen lässt sich auch historisch herleiten, denn viele Archivalien aus der Vergangenheit sind heute nur überliefert, weil sie um des Gebrauchen-Willens gut aufbewahrt wurden; Unzähliges der Vergangenheit hingegen verschwand aus der Aufmerksamkeit und ist heute verloren.

Für die historische bzw. allgemein geisteswissenschaftliche Forschung sind viele gut erschlossene wie auch gut zugängliche handschriftliche Dokumente ein Pull-Faktor: Wo ein Angebot vorhanden ist, lässt sich auch die Wissenschaft nieder. Gleichzeitig sollten die Wünsche und Anforderungen der Wissenschaft aber für die Infrastrukturbetreiber ein Push-Faktor sein, ebendiesen Anforderungen auch gerecht zu werden. Und diese Anforderungen und Erwartungshaltungen nehmen derzeit in dem Maße zu, in dem durch die Digitalisierung neue Forschungsmöglichkeiten geschaffen werden, die gerne unter dem Schlagwort Digital Humanities oder Digital History zusammengefasst werden.

2 Erschließung von handschriftlichen Dokumenten im digitalen Raum

Was charakterisiert nun die „digitale Welt“ in Bezug auf die (historische) Wissenschaft? Donig und Rehbein haben das Phänomen der Digitalität durch einen politischen Bericht zu erschließen versucht.⁶ In dem Gutachten des Wissenschaftlichen Beirats der Bundesregierung Globale Umweltveränderungen (WBGU) der deutschen Bundesregierung über „Unsere gemeinsame digitale Zukunft“ aus dem Jahr 2019 wird die große digitale Transformation unter einer klaren Nachhaltigkeits-Agenda als „in allen wirtschaftlichen, sozialen und gesellschaftlichen Systemen tiefgreifend“ wirkend beschrieben. Die Digitalisierung „entfalte eine immer größere transformatorische Wucht, die [...] fundamental beeinflusst und daher gestaltet werden muss“.⁷ In diesem Zusammenhang arbeitete der WBGU in seinem Bericht fünf zentrale Merkmale der digitalen Transformation heraus: Virtualität, Vernetzung, Kognition, Wissensexplosion und Autonomie.

Alle diese Eigenschaften der digitalen Transformation spielen im weiteren Verlauf des Beitrags noch eine Rolle. Die digitale Transformation wird, so die Übereinstimmung mit den Autor:innen des Berichts, die Wissenschaft, auch die Geschichtswissenschaft, grundlegend verändern. Unsere eigene Aufgabe und Verantwortung als Forschende dabei ist, dass sie sich im Positiven verändert, dass neue Möglichkeiten geschaffen werden, dass sich Forschungslandschaften öffnen, die vorher nicht bekannt waren, zugleich aber bewährte Praktiken bewahrt bleiben können. Dieses Versprechen der Digitalisierung ist zwar nicht gerade neu, aber seine Erfüllung beschleunigt sich kontinuierlich unter anderem unter den Schlüsselwörtern dieses Beitrags. Gleiches gilt für die Infrastrukturanbieter.

Hierbei steht „Big Data“ symbolisch für einen Teil der sich erweiternden Forschungslandschaft durch die schon vor einiger Zeit formulierten Kriterien *Volume* (große Datenmenge), *Velocity* (hohe Verarbeitungsgeschwindigkeit und Dynamik der Daten), *Variety* (vorhandene Varianz und Heterogenität der Daten und Datentypen) und *Veracity* (fragliche Qualität und Glaubwürdigkeit in Bezug auf Herkunft und Inhalt der Daten).⁸ Die vier „V“ korrespon-

⁴ Bloch (2008) 80.

⁵ Ebd.

⁶ Donig und Rehbein (2022). Die Autoren diskutieren die fünf Merkmale in Bezug auf Geschichtswissenschaften ausführlich.

⁷ WBGU (2019) 1.

⁸ Vgl. Lagoze (2014), Strasser und Edwards (2017), Schöch (2013). In manchen Fällen werden die Merkmale noch um die Punkte *Validity* und *Value* auf sechs ergänzt, vgl. Fickers (2022).

dieren dabei in gewisser Weise mit den Charakteristika der Digitalität allgemein. Die Big Data-Metapher steht überall im Raum, wo Digitalisierung wirkt, also nahezu überall; diesen Wirkungsbereich daher auch den digitalen Raum zu nennen, liegt nahe.

Digitalisierung im Bereich der handschriftlichen Dokumente schafft umfassenderes Volumen, wodurch mehr Quellen einem breiteren Publikum zugänglich gemacht werden. Im klassischen Nutzungsszenario der Historiker:innen – finden, lesen, bewerten – ist das bereits sehr hilfreich, aber noch nicht das Ziel, das man in diesem Kontext eigentlich erreichen möchte. Dies betrifft auch die Infrastruktur-anbieter, denn die Erwartungen der Forschung gehen über das primäre Auffinden von Material hinaus. Hierfür wäre der Prozess der Digitalisierung und digitalen Erschließung vielmehr als der einer „Verdatung“ zu verstehen, wobei sowohl *Variety* als auch *Veracity* zu typischen Merkmalen speziell geisteswissenschaftlicher Daten gehören.⁹

Was bedeutet aber „Verdatung“?¹⁰ Wir verstehen darunter die mehr oder weniger stark formalisierte Abbildung von in den Quellen enthaltenen, häufig nicht explizit sichtbaren Informationen auf Datenstrukturen. Das umfasst zunächst sowohl die klassischen Erschließungsdaten – also Metadaten zum Ordnen und Finden, im Bereich des *Discovery*, als auch die Volltexte der Quellen. Hinzu kommen aber auch aus den Texten heraus extrahierte und ggf. durch Annotation angereicherte Daten, wie zum Beispiel das Erkennen von *Named Entities*, Personen- Orts- und Institutionsnamen, das Identifizieren der dahinterliegenden Realobjekte und das Verlinken der Namenstoken wie Toponymen mit entsprechenden Normdaten mittels Annotation. In dem durch diese Prozesse der Verdatung, die durch forschende Tätigkeit noch deutlich erweitert werden können, aus historischen Objekten digitale, und vor allem digital verarbeitbare Daten entstehen, erwachsen neue Nutzungsszenarien, neue Forschungsperspektiven, die dann auch eine erweiterte historische Methodik nach sich ziehen bzw. voraussetzen. Bevor dieser Aspekt näher beleuchtet wird, möchten wir allerdings zunächst durch einige Bemerkungen die Aufmerksamkeit auf den Begriff „erschließen“ lenken.

Was wir hier als Verdatung bezeichnen, ist eine Weiterführung des klassischen Prozesses der Erschließung bzw. Verzeichnung von Objekten, wie sie in den Gedächtnisinstitutionen vorgenommen werden, seien diese Objekte Bücher, archivalische Quellen oder museale Artefakte. Dabei setzt das Erschließen stets das Bewahren voraus: Was nicht bewahrt wurde, kann nicht erschlossen werden. Und Er-

schließung ist wiederum Voraussetzung für die Zugänglichmachung: Was nicht erschlossen ist, wird kaum gefunden und genutzt. Dabei können zusätzlich mehrere Stufen der Erschließung angenommen und differenziert werden: Was nicht tief erschlossen ist, ist nicht für eine hermeneutische Tiefenanalyse zugänglich, was nicht breit erschlossen ist, erlaubt keine Raum und Zeit übergreifende Betrachtung, was allgemein nicht verdatet ist, verschließt sich den Ansätzen der Datenanalyse zunächst grundsätzlich.

Das Wort „erschließen“ ist häufig mit „kultivieren“, „urbanisieren“ und „zivilisieren“ und somit stark landwirtschaftlich konnotiert.¹¹ Damit erscheint das Erschließen als ein wesentlicher Aspekt der Entwicklung hin zum modernen, „zivilisierten“ Menschen. Unweigerlich rückt dabei der Übergang von der Altsteinzeit der Jäger und Sammler zur Jungsteinzeit ins Bild – der von da an fest siedelnden und Land- wie auch Tierwirtschaft betreibenden Menschen. Dies lässt eine provokante metaphorische Analogie zweier wissenschaftlicher Arbeitsweisen zu, die des „Jagens und Sammels“ und die des durch eine viel stärkere und aktive Systematik geprägte Arbeitsweise, die auf einem „urbanisierten“ Handschriftenwesen beruht. Für letztere werden 1. gemeinsame digitale Infrastrukturen geschaffen – um im Bild zu bleiben: Aus den einzelnen Höfen werden Städte geformt – und 2. indem Quellen in großem Umfang, idealiter vollzählig und vollständig, „kultiviert“, also digital erschlossen werden.¹²

Die Erschließung im digitalen Raum beginnt entweder mit digitalen Findmitteln¹³, die hier nicht weiter betrachtet werden sollen, oder mit der Digitalisierung des Objekts. Eine traditionelle Endform solcher wissenschaftlicher Erschließung von handschriftlichen Dokumenten ist die historisch-kritische Edition. Zwischen einfachem Findmittel und komplexer und aufwendiger Edition stehen verschiedene Abstufungen in der Erschließungstiefe. Dies beginnt bereits mit der Definition der zu erschließenden (archivalischen) Einheit, die etwa zwischen Briefkonvolut und einzelнем Brief, zwischen einer Archivbox „Verschiedenes“ und jedem darin enthaltenen Stück oder zwischen einem Fotoalbum

⁹ Vgl. Schöch (2013).

¹⁰ Vgl. hierzu Donig und Rehbein (2022).

¹¹ erschließen – Schreibung, Definition, Bedeutung, Synonyme, Beispiele (2022).

¹² *Nota bene*, dass solche Überlegungen in den Geisteswissenschaften keineswegs neu sind, sondern ihre Wurzeln durchaus etwa bei den Bolandisten, spätestens aber bei den Ansätzen einer „Groß-Wissenschaft“ etwa bei Theodor Mommsen haben, vgl. Mommsen (1905).

¹³ Archivinformationssysteme gibt es zahlreiche am Markt, von freien Lösungen wie beispielsweise AtoM (Access to Memory) über weit verbreitete Produkte wie die Schweizer Software scopeArchiv bis hin zu – aufgrund von Nachhaltigkeits- und Vernetzungsaspekten weniger zu bevorzugenden – Individuallösungen. Siehe auch eine etwas ältere Bestandsaufnahme Maier und Fritz (2010).

als Ganzes und den einzelnen Fotos, die im Album einsortiert sind, schwankt. Erschließung ist somit auch eine Frage von Aggregation und folglich von Komplexitätsreduktion – ein für die gesamte Wissenschaft wichtiges Prinzip. Auch in Bezug auf die zu erschließende Einheit beobachten wir verschiedene Ebenen, teilweise abhängig voneinander bzw. aufeinander aufbauend.

So reichen die verschiedenen Erschließungstiefen etwa bei einem kürzlich abgeschlossenen Digitalisierungsprojekt an der Universität Passau von fünf Minuten für die Erfassung eines einzelnen Dokuments in Form von Pflichtfeldern (ID, Kurzbeschreibung) und einem Mindestmaß an Metadaten (wie Umfang und Laufzeit) durch eine gut qualifizierte Hilfskraft bis hin zu 60 Minuten für die Erfassung aller definierten sachbezogenen Metadatenfelder einschließlich möglichst gründlicher Verschlagwortung inklusive (sofern möglich) GND-Verlinkung durch eine:n Fachwissenschaftler:in. Bei Erschließungsprojekten mit großem Volumen (hier: ca. 14 000 Digitalisate) wird klar, welch hohe Bandbreite hinsichtlich (Zeit-)Ressourceneinsatz daraus resultiert. Um mit vertretbarem Zeitaufwand eine nutzbare Forschungsdatenbank aufzubauen, ist es daher nötig, flexibel und individuell unterschiedliche Erschließungsgrade bzw. -tiefen für die einzelnen Sammlungsbestandteile einzuplanen. In der oftmals zu entscheidenden Frage zwischen Breite und Tiefe kann ein allgemeiner Trend in der Wissenschaft beobachtet werden, der doch deutlich in Richtung Breite statt Tiefe zeigt. Viel möglichst schnell online zu stellen, heißt meist die Devise: *Volume* und *Velocity* stehen im Zentrum.¹⁴ Dennoch, die Erschließung von Volltexten und/oder Generierung ist eine wichtige Perspektive für die Forschung, und das muss sich nicht gegenseitig ausschließen.

3 Perspektiven für die Forschung

Im Wortfeld von „erschließen“ finden wir neben vielen wirtschaftlichen Begriffen auch die Wörter „Neuland“ und „Panorama“, die sich ebenso treffend in die Thematik einfügen. Schlussendlich ist es das Ziel der Forschung, Neuland zu entdecken und es auch zu betreten, die *terra incognita* zu erforschen und bekannt zu machen. Dazu passt das Bild des Panoramas, die Aussicht auf einen nun erweiterten Horizont.

In diesem Sinne hat die Erschließung handschriftlicher Dokumente für die historische Forschung, die ja letztlich

immer auf Quellen zurückzuführen ist, eine fundamentale Funktion inne. Gleichzeitig wirkt sie deterministisch auf die Forschung zurück durch die Auswahl, die sie tätigt bzw. die diejenigen tätigen, die die Erschließung handschriftlicher Dokumente vornehmen: Die gedruckte Quelle war schon immer eher die Grundlage der historischen Forschung als das kaum beschriebene Archival selbst. Das Ungedruckte ist häufig unerreichbar; eher wenige Student:innen wagen den unbequemen Gang ins Archiv. Das war früher so und scheint im digitalen Raum erst recht so zu sein. Weiterhin werden mit der zunehmenden Verdattung und der Etablierung der Data Science in der historischen Wissenschaft neben Quellen im klassischen Sinne auch Datengrundlagen geschaffen, die andere methodische Ansätze neben dem *close reading* bieten: Distanzverfahren (*distant reading*), quantitative Analysen, Raum-Zeit-Visualisierungen, Netzwerkanalysen und dergleichen wie die folgende Fallstudie illustriert, die sowohl auf archivalische Quellen als auch auf bibliothekarische Editions-Bestände Bezug nimmt.

Grundlage der Betrachtung zur päpstlichen delegierten Gerichtsbarkeit im 12. Jahrhundert bildet ein rund 1250 Papsturkunden umfassendes, mit Metadaten erschlossenes Urkundenkorpus, das die Rombeziehungen der Iberischen Halbinsel in ihrer Gesamtheit abbildet und als Folie dient, vor der sich die delegierte Gerichtspraxis abhebt und in den größeren Kontext der Ausrichtung auf das Papsttum gestellt werden kann. Dazu wird der gesamte Quellenbestand im Rahmen einer statistischen Kartografierung räumlich, zeitlich und formal beschrieben und für vertiefende Analysen aufbereitet.¹⁵ Mit der Untersuchung von Akteursgeflechten und formelhaftem Schreiben erkundet das Projekt zwei zentrale Dimensionen des Verbreitungsprozesses, die durch die Anwendung computergestützter Verfahren, hier: Netzwerkanalyse sowie Natural Language Processing, überhaupt erst durchführbar werden.¹⁶ Als Datengrundlage wird auf Edi-

¹⁵ Dabei werden die Praxis und Verbreitung der päpstlichen delegierten Gerichtsbarkeit auf der Iberischen Halbinsel im 12. Jahrhundert analysiert. Das Projekt lenkt damit den Blick auf eine Region, für die in den letzten Jahren größere Fortschritte in der Quellenerschließung erzielt wurden, so dass eine großflächige Untersuchung des Phänomens vor dem Hintergrund der zunehmenden Verrechtlichung des kirchlichen Lebens möglich erscheint.

¹⁶ Mit den Methoden der Historischen Netzwerkforschung werden die an den Gerichtsverfahren beteiligten Personen (Streitparteien und Richter) in ihren Beziehungen untereinander und zum Papsttum betrachtet, um raumzeitliche Cluster aufzudecken, die historisch zu bewerten und zu interpretieren sind. So kann danach gefragt werden, ob sich für bestimmte Personen(gruppen) eine Vorreiterrolle abzeichnet oder ob gemeinsame Merkmale hervortreten, die es erlauben, die Verbreitung dieser neuen Form von Rechtsprechung mit bestimmten klerikalen Gruppen oder Milieus in Verbindung zu bringen. In Verbindung damit untersucht das Projekt auch die Urkundensprache, speziell die

¹⁴ Vgl. große Infrastrukturanbieter wie Fachinformationsdienste (FID).

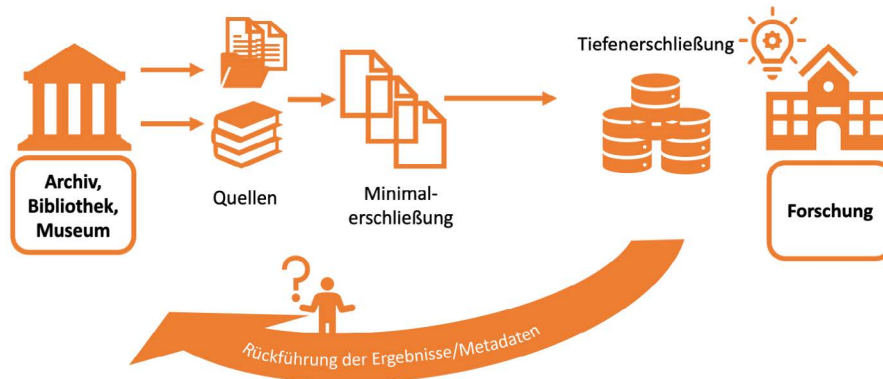


Abb. 1: Spannungsfeld zwischen bewahrenden Institutionen und Forschung hinsichtlich Erschließungstiefe; CC-BY-SA Rehbein/Ernst

tions- bzw. Regestenbände, aus denen bereits vorhandene Metadaten übernommen werden, bestehende Volltexte, aus denen durch die digitale Bereitstellung der darin enthaltenen Informationen und Daten ein nachhaltiges „Up-cycling“¹⁷ bzw. eine Nachnutzung ermöglicht und neue Erschließungsdaten gewonnen werden, zurückgegriffen. Teilweise geschieht dies in beiden Fällen als Ersterfassung und -erschließung, in jedem Fall ist aber der Ursprung derselbe: handschriftliche Dokumente (Urkunden). Es sind hier die makroskopischen „large-scale“-Untersuchungen über den gesamten Bestand, und zwar sowohl auf der deskriptiv-quantitativen Ebene (Verteilungen etc.) als auch auf der qualitativ-inhaltlichen Ebene, die interessieren.

Vor allem bei solchen Ansätzen der Digital History greift der Aspekt „Kognition“ aus den oben diskutierten Kernmerkmalen der Digitalisierung im Wissenschaftsbereich: Menschliche kognitive Fähigkeiten werden durch Technologie insofern erweitert, als diese Werkzeuge bereitstellt, mit deren Hilfe größere Datenmengen intellektuell verarbeitet werden können, die in diesem Falle durch Erschließung handschriftlicher Dokumente bereitgestellt werden. Zugängliche Daten und Werkzeuge induzieren gemeinsam also neue Forschung, sowohl bei einer in die Breite als auch bei einer in die Tiefe angelegten Erschließung.

Verbreitung formelhafter Elemente in den päpstlichen Justizbriefen. Hier stellt sich unter anderem die Frage, inwieweit sich Genese und Verfestigung solcher Formeln an konkrete Streitverläufe rückkoppeln lassen und ob sich bei deren Verbreitung ebenfalls Cluster abzeichnen, die zu den personalen Verflechtungen in Beziehung gesetzt werden könnten. Neben der händischen Kollationierung bekannter Formeln und Klauseln erprobt das Projekt hierzu eine datenzentrierte Identifizierung formelhafter Elemente mittels Text-Mining-Verfahren.

¹⁷ Donig und Rehbein (2022). Der Begriff „Up-cycling“ wird jüngst auch von Scheltjens (2023) verwendet und ausführlich diskutiert.

4 Herausforderungen der Erschließung handschriftlicher Dokumente: zwischen Überzeitlichkeit, nachhaltiger Generalisierung und notwendiger Spezialisierung

Im Wechselspiel zwischen Bereitstellung und Erschließung einerseits und Nutzung durch die Forschung andererseits ist festzustellen, dass die Erschließung zunächst eine unmittelbare Konsequenz aus dem Auftrag von Bibliotheken und Archiven ist: Zugänglichkeit erfordert Auffindbarkeit, wofür eine zumindest minimale Erschließung notwendig ist, weswegen auch die Auffindbarkeit und somit minimale Erschließung im natürlichen Interesse der Gedächtnisinstitutionen liegt.

Aus der Erschließung des Materials an der bewahrenden Stelle ergibt sich aber auch ein Spannungsfeld mit der Forschung (siehe Abb. 1). Dies deshalb, weil mit der Erschließung den Objekten Metadaten zugefügt und diese somit kategorisiert werden. Der gewählte Begriff der Kategorie ist an dieser Stelle besonders hilfreich, um auf das Spannungsfeld hinzuweisen, da er sowohl in der Geschichtstheorie als auch der Informationswissenschaft und in der allgemeinen Modelltheorie hilfreiche Impulse setzt, ebenso ist er für die hermeneutische Methodik zentral. Wer erschließt, bildet Kategorien bzw. ordnet Einheiten solchen Kategorien zu. In seinem noch heute viel beachteten Einführungsbuch „Apologie der Geschichtswissenschaft oder Der Beruf des Historikers“ weist auch Marc Bloch schon in den 1940er-Jahren auf die zentrale Bedeutung von Klassifikation und Kategorisierung für die Geschichtswissenschaft hin; ja man mag sogar meinen, er setzt sie mit dieser gleich.

Herausforderungen des Verarbeitungsprozesses

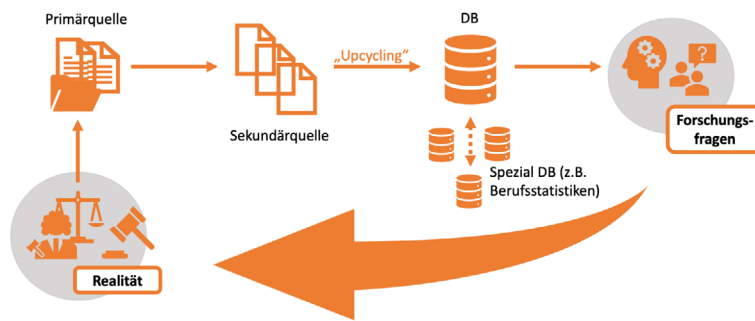


Abb. 2: Übersicht zu den Verarbeitungsschritten bei der Aufarbeitung der Verfahren vor dem NS-Sondergericht München; CC-BY-SA Rehbein/Ernst

Damit wird sogleich deutlich, wie sensibel die Aufgabe der Kategorisierung und mit ihr die Erschließung von historischen Handschriften mittels Metadaten ist.¹⁸

Bevor auf konkrete Ansätze der Fachwissenschaft, der Citizen Science und der künstlichen Intelligenz eingegangen wird, sei zunächst eine kurze Diskussion von Klassifizierungsproblemen bzw. -erfordernissen ausgeführt, die anhand 1. einer projektspezifischen Perspektive illustriert und 2. mit wissenschaftstheoretischen Überlegungen, insbesondere der historischen Semantik wie wir sie bei Marc Bloch ebenso finden wie bei Hermeneutiker:innen, belegt werden soll. Dabei illustriert das folgende Projektbeispiel das Spannungsfeld der Erschließungsdaten zwischen deren Bereitstellung durch Dritte und dem, was für einen spezifischen Forschungszweck benötigt wird und kontextualisiert dieses im Konzept des „Upcyclings“.

Ein Fallbeispiel zu diesem Thema liefert ein Projekt zur NS-Justiz, bei dem die digitale Aufarbeitung von ca. 10 000 Verfahren, die vor dem Sondergericht München zwischen 1933 und 1945 verhandelt wurden, im Fokus steht. Die teils handschriftlichen Akten der Fälle wurden in den 1970er-Jahren erschlossen und in Form von „Aktenregesten“ analog aufbereitet.¹⁹ Die modellierten Informationen, einheitlich in Maschinenschrift dokumentiert, eignen sich besser für eine systematische und vor allem automatische Analyse durch digitale Methoden als die heterogenen Originaldokumente.²⁰ Forschungsfragen und -ziele der aktuellen Bearbeitung liegen in der Untersuchung der Urteilspraxis inklusive einer Suche nach Mustern und Korrelationen, z. B. sozialen Merkmalen, die eine härtere/mildere Strafe wahrscheinlich machten, oder nach den individuellen Ur-

teils-Spielräumen der Richter. Allerdings gibt es auch Probleme bei der Verwendung der vorverarbeiteten (Sekundär-)Quelle. Die Beschreibungen der Straftaten sind nicht nach strafrechtlicher Terminologie klassifiziert, durch den langen Bearbeitungszeitraum und die Vielzahl an Beteiligten sind Uneinheitlichkeiten im Inhalt zu erkennen und es fehlen für die umfassende Analyse wichtige Informationen hinsichtlich der Ankläger und Richter.

Ein möglicher Lösungsansatz ist es, die Annotation der Daten so generisch wie notwendig und upcycling-fähig wie möglich zu gestalten, um die Nachnutzbarkeit wie auch Möglichkeiten der Fragestellungen an die Quellen zu erhöhen. Bei allen Kategorisierungs-, Modellierungs- wie auch Abstrahierungsschritten muss allerdings die Nachvollziehbarkeit nicht außer Acht gelassen und mitbedacht werden. Am Ende sind die Forschungsfragen (meist) an die historische Realität, die tatsächlichen Sachverhalte oder so gut sie eben in Quellen festgehalten und überliefert wurden, gerichtet (siehe Abb. 2).

Schlussendlich wird für alle Arbeitsabschnitte ein Reflektieren der Modellierungs- und Abstrahierungsschritte nötig, wie auch im gesamten Einsatzgebiet von digitalen Methoden – von der Recherche bis hin zur Dissemination – ein Plädoyer für eine digitale Hermeneutik besteht.²¹ Das zusätzliche Mitdenken der Rückführung der angereicherten Daten an die Verwahrungsinstitutionen gehört ebenso dazu (siehe Abb. 1). Nach wie vor kommt dieser Aspekt aufgrund fehlender oder mangelnder Infrastrukturen zu kurz.

Das zweite Spannungsfeld ist im Kontext der historischen Semantik auszuloten. Die den Erschließungsverfahren zugrundeliegende Klassifizierung kann – wie durch den vorherigen Abschnitt illustriert – durch spezifische Forschungsfragen und -ziele problematisch werden. Verallgemeinernd kommt hinzu, dass die für die Verdattung herangezogenen Kategorien, Begriffe und Konzepte kul-

¹⁸ Vgl. Bloch (2008).

¹⁹ Vgl. Widerstand und Verfolgung in Bayern 1933–1945. Hilfsmittel, im Auftrag des Bayerischen Staatsministeriums für Unterricht und Kultus (BStMUK) herausgegeben von der Generaldirektion der Staatlichen Archive Bayerns. Archivinventare, Bd. 3: Sondergericht München, 7 Teile (1975–1977).

²⁰ Vgl. Gerstmeier et al. (2022), Ernst et al. (2023).

²¹ Vgl. Fickers et al. (2022).

turell bedingte und weitere kontextuale Ambiguitäten und Unschärfen aufweisen. Sie können zudem selbst einem historischen Wandel unterliegen, was eine zentrale Herausforderung darstellt. Marc Bloch schreibt in seiner anschaulichen Art u. a. am Beispiel des Lehnwesens aber auch von moderneren technischen Entwicklungen: „Denn zum großen Leidwesen der Historiker pflegen die Menschen nicht immer ihren Wortschatz zu verändern, wenn sie ihre Gewohnheiten ändern.“²² Er setzt fort:

„Ein abstrakter Name ist immer nur ein Etikett, das der Klassifizierung dient. Alles, was man von ihm erwarten kann, ist, dass es die Tatsachen in eine Ordnung bringt, die ihrer Erkenntnis dienlich ist. Nur willkürliche Klassifikationen sind verhängnisvoll. Es ist daher Aufgabe des Historikers, die seinen ständig zu überprüfen, sie nötigenfalls zu korrigieren und vor allem flexibler zu machen. Im übrigen sind sie naturgemäß sehr verschiedenartig.“²³

Bloch weiter: Es

„gehen die Veränderungen nämlich fast immer so langsam vor sich, dass die Menschen, die von ihnen betroffen sind, sie nicht wahrnehmen können. Sie empfinden kein Bedürfnis, das Etikett auszutauschen, da ihnen die Veränderung des Inhalts entgangen ist. [...] Zunächst bringen Veränderungen der Dinge keineswegs immer parallele Veränderungen ihrer Bezeichnungen mit sich.“²⁴ Aber „umgekehrt kommt es [...] auch vor, dass die Bezeichnungen im Lauf der Zeit oder von Ort zu Ort variieren, ohne dass sich die Dinge verändert hätten.“²⁵

Nur wenige Begriffe sind überzeitlich,²⁶ wodurch es bei der Aufgabe der Klassifizierung zu Umformungen im Verständnis der kontemporären Perspektive im Vergleich zur Zeit des Erschließens kommen kann. Welche zukünftigen Bedeutungsebenen dem hinzukommen, lässt sich nur erahnen. Damit einher geht auch der Bedeutungswandel hinsichtlich Quellen- und Gegenwartssprache, was exemplarisch in besonderem Maße für den NS-Sprachegebrauch gilt. „Im Unterschied zu Mathematik oder Chemie“, so sagt Bloch, „verfügt unsere Wissenschaft über kein von allen Nationalsprachen unabhängiges Symbolsystem. Der Historiker drückt sich nur mit Wörtern aus – und zwar denen seiner Landessprache.“²⁷ Dies betrifft ebenso die Kategorientitel (wie Datum, Name, Schlagwort), aber insbesondere auch die Befüllung z. B. mit kontrollierten Vokabularen. Um hier objektiv eine „ground truth“ zu erreichen, mögen schwergewichtiger Ontologien eine Abhilfe bieten, da Raum-,

Zeit- und Kulturbezüge implementiert werden können. Sie sind allerdings ausgesprochen schwer zu erstellen und zu verhandeln. Kompatibilitäten bzw. Interoperabilitäten sind nach wie vor ein Problem bzw. nicht gegeben, auch wenn mit Standards und Normdatensätzen wie GND, Wikidata, DOI etc. gearbeitet wird. Wenn größere Informationsanbieter zudem auf mehreren Plattformen vertreten sind, ergibt sich ein komplexes n:m-Gefüge, das entsprechend gepflegt werden muss.

5 Citizen Science vs. Fachwissen (?)

Eingangs wurde das öffentlichkeitsnahe Arbeiten bereits angeschnitten. Mit Citizen Science wird dem aktiven Mitbinden von Laien in den Forschungsprozess ein Name gegeben.²⁸ *Crowdsourcing*, das Auslagern von (Teil-)Aufgaben, ist eine der häufigsten Beteiligungsmöglichkeiten für interessierte Bürger:innen, unabhängig von der Forschungsdisziplin. In den Geschichtswissenschaften fallen in diesen Bereich vor allem Transkriptionsvorhaben.²⁹ Trotz des gemeinsamen Themas sind die Herangehensweisen oft recht unterschiedlich und ebenso divers wie die Quellen, die bearbeitet werden. In vielen Fällen werden wenig bis gar kein Vorwissen und Lesekompetenz vorausgesetzt und die Laienforscher:innen durch mehr oder weniger intensive Schulungen bzw. Online-Tutorials auf die Mitarbeit vorbereitet. Ein Extrembeispiel hierfür ist das über die Zoo-universe-Plattform angebotene US-amerikanische Projekt *Scribes of the Cairo Geniza*³⁰, wo ohne jedwede Vorkenntnisse hinsichtlich Sprache, vormoderne jüdische wie auch arabische Textfragmente bearbeitet werden. In erster Linie geht es vor allem um die Unterstützung bei der Schriftidentifikation und damit einhergehenden Anreicherung der Digitalisate um rudimentäre Metadateninformationen. Erst als nachgereichte und noch etwas komplexere Aufgabenstellung wird an Transkriptionen gearbeitet bzw. deren Bearbeitung vorbereitet.

Ein (älteres) Beispiel für eine Einbindung der „crowd“ zur Metadaten-Erschließung bzw. eine Nutzung der „Schwarmintelligenz“, um weitere Konzepte innerhalb der Citizen Science zu verwenden, setzt auf Aspekte der Gamification, d. h. spielerische Elemente zur Motivation

²² Bloch (2008) 39.

²³ Bloch (2008) 164.

²⁴ Bloch (2008) 176.

²⁵ Bloch (2008) 177.

²⁶ Vgl. zum Begriff Seiffert (1992) 148 ff.

²⁷ Bloch (2008) 179.

²⁸ Vgl. Haklay et al. (2021), für die Geistes- und Kulturwissenschaften siehe Smolarski und Oswald (2016), für die Digital Humanities siehe Carius et al. (2020). Siehe ebenso den Beitrag zu CS und Bibliotheken in diesem Band.

²⁹ Vgl. Oswald (2018), Brauer und Ernst (2023) 191–94.

³⁰ *Scribes of the Cairo Geniza* (o. J.).

in die Beteiligung einzubauen.³¹ Neben bekannten naturwissenschaftlichen Forschungsspielen wie *Eterna*³² oder *Stall Catchers* bzw. *EyesOnAlz*³³, gibt es auch in den Geisteswissenschaften ein Vorzeigbeispiel in diesem Kontext. Das kunsthistorische Projekt ARTigo, seit 2010 an der LMU München stetig (weiter-)entwickelt, veranschaulicht, was die Unterstützung durch Bürger:innen zu leisten vermag: Seit Projektbeginn wurden über 60 000 Kunstwerke mit beinahe einer Million Schlagwörter durch über 45 000 Freiwillige auf spielerische Art und Weise versehen.³⁴ Dabei werden die Ergebnisse durch keinerlei Vorgaben, weder eine Kategorie der Verschlagwortung noch kontrollierte Vokabulare, eingeschränkt. Es gibt keine explizite Frage nach dem Entstehungszeitraum des Gemäldes, nicht nach den Künstler:innen, nicht nach dem Kunststil – bei anderen entsprechend ausgerichteten Informationssystemen, wozu auch Handschriftenportale zu zählen sind, wird dies anders gehandhabt und Taxonomien zur Projektbeschreibung vorgegeben bzw. mit kontrollierten Vokabularen von Fachbegriffen gearbeitet. Dass die Herangehensweise bei ARTigo eine andere ist sowie die Tatsache, dass ein Objekt durch viele und nicht nur wenige Menschen beschrieben wird, führt offenbar zu deutlich mehr und zu anderen Schlagwörtern. Es kommen auch laienhafte Begriffe zum Einsatz, ebenso werden dem Objekt Eigenschaften zugeschrieben, die von Expert:innen vielleicht als trivial oder irrelevant angesehen werden könnten bzw. die in der Fachtaxonomie nicht vorkommen. So zeigt das Projekt ARTigo in der Verschlagwortung eines Objekts deutliche Unterschiede zu „Expertensystemen“, sowohl was die Quantität als auch die Art der Schlagwörter betrifft.³⁵

Diese andere Quantität wie Qualität der Verschlagwortung führt zu anderen Möglichkeiten der Nutzung solcher Informationssysteme. Man könnte hier provokant formulieren: Die Citizen Scientists annotieren Dinge, die auch intuitiv nachgefragt werden, während die kunsthistorischen Fachwissenschaftler:innen Dinge aufgreifen, die von früheren Autoritäten gelernt wurden. Sowohl auf Seiten der Verschlagwortung als auch der späteren Suche

scheint die Fachwissenschaft durch dieses Vokabular eingeschränkt und dadurch, dass diese Systematik immer wieder zugrunde gelegt wird, könnte das System autopoetisch erstarren. Freilich kann der Citizen-Science-Ansatz auf der anderen Seite zu einer gewissen Beliebigkeit und zur Ignoranz gegenüber bestehendem Fachwissen führen.

Wozu Beispiele von Citizen Science, von den es inzwischen vor allem im naturwissenschaftlichen Bereich zahlreiche gibt, auf jeden Fall führen, sind diese Diskussionen: über die Rollenverteilung in der Wissenschaft, über ein neues Verständnis von Autorität in der Wissensschaffung, letztlich zu der großen Frage: Was überhaupt ist Wissen?

6 ChatGPT-Schock

Aktuell in aller Munde und ein weiterer Player, der nun hinzukommt und die diversen Spannungsfelder erweitert, betrifft den Einsatz von künstlicher Intelligenz (KI) in allen Lebensbereichen, auch der Forschung. Wie auch schon in der Vergangenheit können neue, erstmalig eingesetzte Technologien zu einer gewissen Konsternation innerhalb der Gesellschaft führen. Der Start des ersten künstlichen Erdsatelliten Sputnik 1 am 4. Oktober 1957 durch die Sowjetunion ging etwa mit intensiven Reaktionen in der westlichen Welt einher – dem sogenannten „Sputnikschock“.³⁶ Diese Schockreaktion ist vergleichbar mit aktuellen Diskussionen rund um ChatGPT und ähnliche Systeme seit Ende 2022. Neben dem – manchmal spielerischen – Austesten der Möglichkeiten und Grenzen des sinnvollen Einsatzes dieser Tools sind es vor allem Abwehrreaktionen und Reglementierungsgedanken, die uns begegnen. Die Angst vor dem vermeintlich allwissenden Computer wird breit diskutiert, die Einflüsse auf Schule, Beruf und Alltag scheinen allgegenwärtig. Der bisherige Konsens: Die Vermittlung gewisser Grundkompetenzen in diesem Zusammenhang wird unumgänglich sein, eine Reglementierung bzw. rechtliche Rahmenbedingungen müssen geschaffen werden.

Für die Forschung ergeben sich dadurch ebenso neue Perspektiven wie gewisse Verschiebungen. Bereits seit Jahren wird intensiv an der automatisierten Handschriftenerkennung (HTR, Handwritten Text Recognition) gearbeitet, und die Fortschritte in diesem Sektor sind enorm.³⁷ Mit diesen wird es zukünftig noch eine stärkere Verlagerung der methodischen Herangehensweise bei der Quellenbear-

31 Vgl. Bowser et al. (2013), Kreitmair und Magnus (2019), Michelucci und Dickinson (2016).

32 Ein Spiel, bei dem es um RNA-Faltungen geht: *Eterna* (o. J.).

33 Hier wird die Alzheimer-Forschung unterstützt: *Stall Catchers* | *EyesOnAlz* (o. J.).

34 Vgl. Commare (2011), ARTigo – Social Image Tagging. Über das Projekt (o. J.).

35 Vgl. Bry et al. (2018). Wie sich auch schon in anderen Citizen-Science-Spielen gezeigt hat, werden besonders gute Ergebnisse erzielt, wenn die verschiedenen Bearbeitungs- und Beteiligungsmodi kombiniert und u. a. auch durch KI-Systeme ergänzt werden. Vgl. Michelucci und Dickinson (2016), Inkpen et al. (2022).

36 Insbesondere in den USA wurde als Ergebnis beispielsweise sehr viel Geld in die Forschung bzw. Gründung großer Forschungsinstitute (meist militärisch motiviert) investiert.

37 Vgl. Dietrich (2022).

beitung geben. Aktuell sind mehrere Systeme am Markt,³⁸ die eine Vielzahl an Trainingsmodellen zur Bearbeitung der unterschiedlichsten Sprachen und Schriftarten bieten. Ein Problem, neben der hohen Fehlerquote bei der Verwendung von nicht auf das eigenen Korpus trainierten Modellen, ist aktuell vor allem die korrekte Layout-Erkennung. Nach wie vor ungelöst ist beispielsweise die Problematik der korrekten Tabellenübertragung von digitalisiertem Original in die maschinenproduzierte Transkription – ein Manko, dass nicht nur, aber äußerst häufig bei wirtschafts- und sozialgeschichtlichem Quellenmaterial zu unzureichenden Ergebnissen bei der Schrifterkennung führt.³⁹

Neben der reinen Texterkennung gibt es zahlreiche weitere Einsatzgebiete bei der KI-unterstützten (Geschichts-) Forschung. Vor allem mittels Textmining-Verfahren und *Machine- bzw. Deep-learning*-Ansätzen lassen sich große Bestände auch inhaltlich systematisch erschließen. Analog wie bei den HTR-Verfahren sind es wiederum die Wahl der Trainingsmodelle bzw. -daten, die Einfluss nehmen auf die Ergebnisse. Unter Nutzung des oben diskutierten Kategorien-Begriffs formuliert, erfolgt „[d]ie Sammlung und speziell auch die Strukturierung von Daten [...] zwar ohne bössartige Absicht, widerspiegelt aber im Kern nicht nur die Vorurteile, sondern bewegt sich in kulturell geprägten Kategorienkatalogen, die reproduziert und durch die Algorithmen verstärkt werden.“⁴⁰ Die kritische Reflexion und Bewertung von digitalen Methoden im Allgemeinen und bei der Erschließung handschriftlicher Dokumente im Speziellen werden somit unumgänglich, um valide Aussagen über große Datenbestände hinweg zu treffen. Wie uns auch ChatGPT wissen lässt:

„Wichtig ist hierbei, die Vorteile der verschiedenen Ansätze zu nutzen und die Nachteile zu minimieren. Die Arbeit der Fachleute bleibt dabei unverzichtbar, da sie über ein profundes Wissen und eine spezielle Expertise verfügen, die für die Handschriftenerschließung von großer Bedeutung sind. Nur durch eine enge Zusammenarbeit von Experten, KI und Citizen Science kann die Handschriftenerschließung erfolgreich bewältigt werden und somit dazu beitragen, unser kulturelles Erbe zu bewahren und zugänglich zu machen. Es ist wichtig, dass die verschiedenen Ansätze sinnvoll kombiniert werden, um optimale Ergebnisse zu erzielen. Dabei ist auch eine klare Kommunikation mit der Öffentlichkeit wichtig, um das Bewusstsein für die Bedeutung der Handschriftenerschließung zu erhöhen und um eine breite Beteiligung zu erreichen.“⁴¹

38 Vgl. Transkribus (o. J.), eScriptorium: Digital Palaeography (o. J.).

39 Vgl. und weitere Details zum HTR-Verfahren siehe Hodel (2023).

40 Hodel (2023) 175.

41 Auszug aus der ChatGPT-Antwort zum Prompt „Bewerte die verschiedenen Ansätze: Fachwissen, Citizen Science und KI für die Handschriftenerschließung“. Sprachmodell: GPT-3.5.

7 Fazit: Prozess und Co-Produktion

Digitalisierung und Erschließung von handschriftlichen Dokumenten sind weiterhin elementar. Sie bewahren diese Form des Kultur- und Erinnerungsguts und machen es (auch) für die Forschung zugänglich. Sowohl das Volumen an Material als auch die neuen methodischen Herangehensweisen unter dem Schlagwort Digital History versprechen einen Katalysator der so wichtigen historischen und kulturellen Forschung. Es ist daher von großer Bedeutung, dass die verschiedenen Ansätze weiterentwickelt und verbessert werden, um die Qualität der Arbeit zu steigern und um neue Erkenntnisse zu gewinnen. Es ist dabei ebenso wichtig zu beachten, dass ein zu starres Vorgehen vermieden und Inklusion bedeutsam gemacht werden. Vor allem letzteres bedeutet, dass keine (potenziellen) Mitwirkenden ausgegrenzt werden sollen; im Gegenteil sollte die Arbeitsweise derart gestaltet sein, dass gemeinsam Wissen und Wissensgrundlagen geschaffen werden. Wir sollten nicht auf das Laienwissen und die Laienarbeit verzichten, genauso wenig auf die maschinelle Unterstützung bis hin zur KI. Bibliotheken und Archive sollten nicht mit ihrer Arbeit stoppen, wenn das handschriftliche Dokument erschlossen ist, und Historiker:innen sollten sich nicht zu schade sein, ihre Erschließungs- und Forschungsdaten, Transkriptionen, Annotationen usw. in das System zurückzuspielen.

Diese beiden Punkte gehen Hand in Hand. Modularität und Flexibilität sind Grundvoraussetzung für die Prozesshaftigkeit, die Co-Produktion und das Überkommen der geschilderten Probleme etwa der der historischen Semantik und der Vergänglichkeit von Wissen. Zugleich aber kann ohne Beständigkeit eine Bestandserschließung nicht funktionieren. Es kann nicht immer alles neu gemacht werden. Zu suchen sind also ein Trade-off zwischen Beständigkeit und Wandel, Produkt und Prozess, zwischen den Bedürfnissen des Erhalts (auch der digitalen Daten) und den Forschungshypes, zwischen Effizienz und Qualität, zwischen Mensch und Maschine, zwischen Anspruch und Wirklichkeit.

Literaturverzeichnis

- ARTigo – Social Image Tagging. Über das Projekt (o. J.): Verfügbar unter <https://www.artigo.org/de/about>, zugegriffen am 10.05.2023.
- Bloch, Marc (2008): Apologie der Geschichtswissenschaft oder der Beruf des Historikers. 2. Auflage. Stuttgart: Klett-Cotta.
- Bowser, Anne; Hansen, Derek; He, Yurong; Boston, Carol et al. (2013): Using gamification to inspire new citizen science volunteers. In: *Proceedings of the First International Conference on Gameful Design, Research, and Applications*, 18–25. Toronto Ontario Canada: ACM. DOI:10.1145/2583008.2583011.

- Brauer, Michael; Ernst, Marlene (2023): Verderben viele Köche den Brei? Citizen Science im Spannungsfeld von Food Studies und öffentlicher Erwartungshaltung. In: *Citizen Science in Historical Science*, hg. von Hendrikje Carius, Martin Prell und René Smolarski, 203–20. Göttingen: V&R Unipress (DH&CS: 3).
- Bry, François; Schefels, Clemens; Schemainda, Corina (2018): Eine qualitative Analyse der ARTigo-Annotationen: Serious Games, Games With A Purpose, Suchmaschine für Kunstwerke, Kunstgeschichte, Human Computation, semantische Tiefe. In: *Computing Art Reader*. DOI:10.11588/ARTHISTORICUM.413.C5771.
- Carius, Hendrikje; Prell, Martin; Smolarski, René (Hrsg.) (2020): Kooperationen in den digitalen Geisteswissenschaften gestalten: Herausforderungen, Erfahrungen und Perspektiven. Göttingen: Vandenhoeck & Ruprecht (DH&CS: 1).
- Commare, Laura (2011): Social Tagging als Methode zur Optimierung kunsthistorischer Bilddatenbanken – Eine empirische Analyse des Artigo-Projekts. In: *Kunstgeschichte. Open Peer Reviewed Journal*. Acquafredda di Maratea, Italien. Verfügbar unter <https://www.kunstgeschichte-ejournal.net/160/>, zugegriffen am 10.05.2023.
- Dietrich, Felix (2022): The road to Handwritten Text Recognition – Part 1. *READ-COOP*. Verfügbar unter <https://readcoop.eu/insights/road-to-htr-1/>, zugegriffen am 14.11.2022.
- Donig, Simon; Rehbein, Malte (2022): Für eine „gemeinsame digitale Zukunft“. Eine kritische Verortung der Digital History. In: *Geschichte in Wissenschaft und Unterricht*, (9/10), 527–45. Verfügbar unter https://elibrary.utb.de/doi/10.5555/gwu-9%2B10-2022_04, zugegriffen am 07.11.2022.
- e-manuscripta / Über e-manuscripta.ch (o. J.): Verfügbar unter <https://www.e-manuscripta.ch/wiki/aboutEmanuscripta>, zugegriffen am 09.05.2023.
- Ernst, Marlene; Gassner, Sebastian; Gerstmeier, Markus; Rehbein, Malte (2023): Categorising Legal Records – Deductive, Pragmatic, and Computational Strategies. In: *Digital Humanities Quarterly*, 17 (3). Verfügbar unter <http://www.digitalhumanities.org/dhq/vol/17/3/000708/000708.html>, zugegriffen am 14.10.2023.
- erschließen – Schreibung, Definition, Bedeutung, Synonyme, Beispiele (2022): *DWDS*. Verfügbar unter <https://www.dwds.de/wb/erschlie%C3%9Fen>, zugegriffen am 04.05.2023.
- eScriptorium: Digital Palaeography (o. J.): eScriptorium: Digital Palaeography. Verfügbar unter <https://www.escriptorium.uk/>, zugegriffen am 09.05.2023.
- Eterna (o. J.): Eterna. Verfügbar unter <https://eternagame.org/>, zugegriffen am 11.05.2023.
- Fickers, Andreas (2022): What the D does to history: Das digitale Zeitalter als neues historisches Zeitregime? In: *Digital History. Konzepte, Methoden und Kritiken Digitaler Geschichtswissenschaft*, hg. von Karoline Dominika Döring, Stefan Haas, Mareike König und Jörg Wettlaufer, 45–64. Berlin, Boston: De Gruyter. DOI:10.1515/9783110757101-003.
- Fickers, Andreas; Tatarinov, Juliane; Heijden, Tim van der (2022): Digital history and hermeneutics – between theory and practice: An introduction. In: *Digital history and hermeneutics: between theory and practice*, hg. von Andreas Fickers und Juliane Tatarinov, 1–22. Boston: De Gruyter Oldenbourg.
- Gerstmeier, Markus; Donig, Simon; Gassner, Sebastian; Rehbein, Malte (2022): Die Archivinventare zum Sondergericht München (1933–1945) digital. Quellenwert – Verdattung – Erkenntnisperspektiven. In: *Archivalische Zeitschrift*, 99 (1), 215–51.
- Haklay, Mordechai (Muki); Dörler, Daniel; Heigl, Florian; Manzoni, Marina; Hecker, Susanne; Vohland, Katrin (2021): What Is Citizen Science? The Challenges of Definition. In: *The Science of Citizen Science*, hg. von Katrin Vohland, Anne Land-Zandstra, Luigi Ceccaroni, Rob Lemmens, Josep Perelló, Marisa Ponti, Roeland Samson und Katherin Wagenknecht, 13–33. Cham: Springer International Publishing. DOI:10.1007/978-3-030-58278-4_2.
- Hodel, Tobias (2023): Konsequenzen der Handschriftenerkennung und des maschinellen Lernens für die Geschichtswissenschaft. Anwendung, Einordnung und Methodenkritik: In: *Historische Zeitschrift*, 316 (1), 151–80. DOI:10.1515/hzhz-2023-0006.
- Inkpen, Kori; Chappidi, Shreya; Mallari, Keri; Nushi, Besmira et al. (2022): Advancing Human-AI Complementarity: The Impact of User Expertise and Algorithmic Tuning on Joint Decision Making. In: *Paper accepted and to be published in Transactions on Computer Human Interaction*. DOI:10.48550/ARXIV.2208.07960.
- Kreitmaier, Karola V.; Magnus, David C. (2019): Citizen Science and Gamification. In: *Hastings Center Report*, 49 (2), 40–46. DOI:10.1002/hast.992.
- Lagoze, Carl (2014): Big Data, data integrity, and the fracturing of the control zone. In: *Big Data & Society*, 1 (2), DOI:10.1177/2053951714558281.
- Maier, Gerald; Fritz, Thomas (Hrsg.) (2010): Archivische Informationssysteme in der digitalen Welt: Aktuelle Entwicklungen und Perspektiven. Stuttgart: Kohlhammer (Werkhefte der Staatlichen Archivverwaltung Baden-Württemberg Serie A, Landesarchivdirektion: 23).
- Michelucci, Pietro; Dickinson, Janis L. (2016): The power of crowds. In: *Science*, 351 (6268), 32–33. DOI:10.1126/science.aad6499.
- Mommsen, Theodor (1905): »Antwort an Harnack, 3. Juli 1890«, in: Reden und Aufsätze, verfasst von Theodor Mommsen, Berlin: Weidmann.
- Oswald, Kristin (2018): Citizen Science-Formate: Transkriptionsprojekte. Bürger Künste Wissenschaft. Verfügbar unter <https://bkw.hypotheses.org/671>, veröffentlicht am: 06.07.2018, zugegriffen am 29.05.2022.
- Scheltjens, Walter (2023): Upcycling historical data collections. A paradigm for digital history? In: *Journal of Documentation*. DOI:10.1108/JD-12-2022-0271
- Schöch, Christof (2013): Big? Smart? Clean? Messy? Data in the Humanities. In: *Journal of Digital Humanities*, 2 (3). Verfügbar unter <http://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/>, zugegriffen am 13.12.2022.
- Scribes of the Cairo Geniza (o. J.): Scribes of the Cairo Geniza. Verfügbar unter <https://www.scribsofthecairogeniza.org/>, zugegriffen am 10.05.2023.
- Seiffert, Helmut (1992): Einführung in die Hermeneutik: die Lehre von der Interpretation in den Fachwissenschaften. Tübingen: Francke.
- Smolarski, René; Oswald, Kristin (2016): Einführung. Citizen Science in Kultur und Geisteswissenschaften. In: *Bürger Künste Wissenschaft. Citizen Science in Kultur und Geisteswissenschaften*, hg. von Kristin Oswald und René Smolarski, 9–27. Gutenberg: Computus Druck Satz & Verlag.
- Stall Catchers | EyesOnAlz (o. J.): Stall Catchers | EyesOnAlz. Verfügbar unter <https://stallcatchers.com>, zugegriffen am 11.05.2023.
- Strasser, Bruno J.; Edwards, Paul N. (2017): Big Data Is the Answer ... But What Is the Question? In: *Osiris*, 32, 328–45. Verfügbar unter <https://www.jstor.org/stable/26562778>, zugegriffen am 10.05.2023.
- Transkribus (o. J.): READ-COOP. Verfügbar unter <https://readcoop.eu/>, zugegriffen am 09.05.2023.
- von Keyserlingk-Rehbein, Linda; Gerstmeier, Markus; Kunze, Reiner (2019): Projektbeginn zur Digitalisierung und Erschließung des Archivs der Reiner und Elisabeth Kunze-Stiftung. Verfügbar unter

<https://www.geku.uni-passau.de/dh/projekte/reiner-kunze-projekt>, veröffentlicht am 01.10.2019, zugegriffen am 09.05.2023.

WBGU (2019): Wissenschaftlicher Beirat der Bundesregierung Globale Umweltveränderungen. WBGU-Hauptgutachten: Unsere gemeinsame digitale Zukunft. Berlin: WBGU. Verfügbar unter https://issuu.com/wbgu/docs/wbgu_hg2019/1, zugegriffen am 09.05.2023.

Widerstand und Verfolgung in Bayern 1933–1945. Hilfsmittel, im Auftrag des Bayerischen Staatsministeriums für Unterricht und Kultus (BStMUK) herausgegeben von der Generaldirektion der Staatlichen Archive Bayerns. Archivinventare, Bd. 3: Sondergericht München, 7 Teile (1975): München: Eigendruck des Staatsarchivs München / Buchbindung Schmidkonz Regensburg.

Zauner, Mirja-Leena (2021): Reiner Kunzes Archiv gibt es jetzt digital. Lehrstuhl für Digital Humanities an der Uni Passau übergibt Festplatte an die Kunze-Stiftung. In: Passauer Neue Presse, 6. November 2021, 23.



Malte Rehbein

Lehrstuhl für Digital Humanities
Universität Passau
Dr.-Hans-Kapfinger-Str. 16
D-94032 Passau
malte.rehbein@uni-passau.de



Marlene Ernst

Stadtarchiv Salzburg
Historisches Archiv
Glockengasse 8
A-5020 Salzburg
marlene.ernst@stadt-salzburg.at