

Von Menschen und Maschinen: Transdisziplinäre Workflows im Münsteraner Editionsprojekt Heinrich Scholz

Dietz, Katharina

kdietz@uni-muenster.de
Universität Münster, Deutschland
ORCID: 0000-0001-7405-3656

Dinger, Patrick

patrick.dinger@uni-muenster.de
Universität Münster, Deutschland
ORCID: 0000-0002-2649-4737

Horstmann, Jan

jan.horstmann@uni-muenster.de
Universität Münster, Deutschland
ORCID: 0000-0001-8047-2232

Normann, Immanuel

immanuel.normann@uni-muenster.de
Universität Münster, Deutschland
ORCID: 0000-0003-4702-1282

Schäfftlein, Vitus

vitus.schaefftlein@uni-muenster.de
Universität Münster, Deutschland
ORCID: 0009-0003-3620-5884

Das Akademieprojekt Heinrich Scholz: Schnittstellen zwischen Menschen und Maschinen

Das Projekt „Heinrich Scholz und die Schule von Münster – Mathematische Logik und Grundlagenforschung“ ist ein 2023 gestartetes und auf 13 Jahre angelegtes Akademieprojekt, das den umfangreichen Nachlass des Theologen, Philosophiehistorikers, Metaphysikers, Mathematikers und Logikers Heinrich Scholz (1884–1956) den FAIR-Prinzipien entsprechend erforschbar machen wird.¹ Unter der Leitung von Prof. Dr. Niko Strobach (Philosophisches Seminar, Universität Münster) ist auch die ULB Münster

mit mehreren Teilprojekten involviert. Ziel ist eine vollständige digitale Ausgabe zu Heinrich Scholz einschließlich seiner nicht veröffentlichten Manuskripte, eine digitale Edition seiner Korrespondenz sowie die semantische Verknüpfung, Annotation und dynamisch-interaktive Bereitstellung mittels Semantic-Web-Technologie in der Linked Open Data Cloud (vgl. etwa Wettlaufer, 2018). Zum Einsatz kommen u. a. Kalliope, Visual Library, verschiedene OCR/HTR-Tools, oXygen zur TEI-Edition sowie die Modellierung in RDF zur Publikation und Visualisierung des Nachlasses als Wissensgraph.

Noch im Jahr von Alan Turings einschlägiger Arbeit „On Computable Numbers, with an Application to the Entscheidungsproblem“ (Turing, 1937), die den Grundstein für die heutige Informatik legte, organisierte Heinrich Scholz das erste Seminar über das Turingmaschinenmodell. Er schrieb an Turing: „Die Methode, die Sie verwendet haben, um die Unlösbarkeit des Entscheidungsproblems schon für den Hilbertschen Prädikatenkalkül der ersten Stufe zu zeigen, ist so fein und originell, dass ich mir vorgenommen habe, über Ihre Arbeit in unserer logistischen Arbeitsgemeinschaft vortragen zu lassen.“

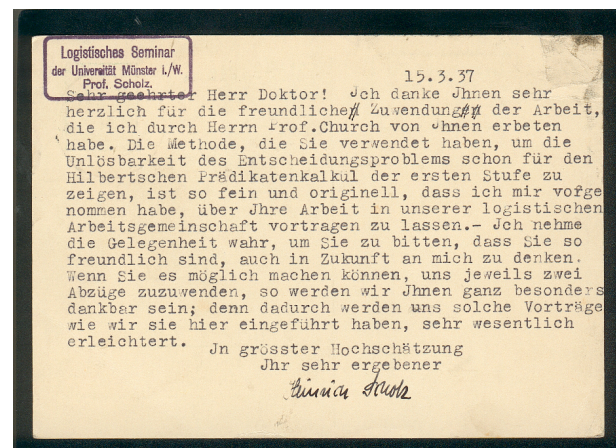


Abbildung 1: Postkarte von Heinrich Scholz an Alan Turing vom 15.03.1937, The Turing Digital Archive, Signatur d5.17a (<https://turingarchive.kings.cam.ac.uk/correspondence-amtd/amt-d-5-04.07.2023>)

Die Mitglieder des von Scholz ins Leben gerufenen Zentrums für mathematische Logik und Grundlagenforschung wurden auch bekannt als „Schule von Münster“. 1943 entstand in Münster der erste Lehrstuhl in Deutschland für mathematische Logik und Grundlagenforschung. Bis heute gibt es keine Gesamtausgabe der Werke von Scholz. Seine umfangreiche Rezensionstätigkeit und Briefkorrespondenz wurden nicht systematisch erschlossen (vgl. Molendijk, 2022).

Bereits zu Beginn des Projekts waren hinsichtlich notwendiger Absprachen der beteiligten Personen, Abteilungen und Systeme zahlreiche Hürden zu nehmen. Es galt, sowohl eine gemeinsame Sprache trotz stark divergierenden fachlichen Hintergründen (Digitalisierung, Katalogisierung, Digital Humanities, Sammlungsforschung und Philosophie) zu finden, als auch Workflows so zu strukturi-

rieren, dass die Anforderungen der jeweiligen Teilschritte bereits in den anderen Prozessschritten antizipiert werden und für Reibungslosigkeit an den Schnittstellen gesorgt ist. Eine Unterteilung von Schnittstellen, die auch unseren Vortrag strukturieren wird, kann bezüglich der Akteure *Mensch* und *Maschine* vorgenommen werden, wodurch sich die folgenden Typen von Schnittstellen ergeben:

1. Mensch-Mensch
2. Mensch-Maschine
3. Maschine-Maschine

1) Aus einer systemtheoretischen Perspektive zerfällt der Akteur *Mensch* in verschiedene Abteilungen und Rollen, die jeweils ihre eigenen Praktiken und Sprachen entwickeln, sodass es bei dieser Schnittstelle vor allem um Kommunikation geht: Als Dokumentationsmedium spielt die wikiartige Software *Confluence* im Projekt eine zentrale Rolle, auf die alle Projektbeteiligten Zugriff haben und die eine sorgfältige Planung und Pflege der Informationsstruktur erfordert. Regelmäßige, abteilungsübergreifende Arbeitstreffen organisieren Arbeitsabläufe. In diesen Arbeitstreffen und ihrer Dokumentation treffen verschiedene Fachjargons aufeinander, die zuvor selten den Weg über Abteilungen hinweg fanden. Etwa die Worte „Ausheben“, „Entmetallisieren“, „Sigel“, „Signatur“, „URN“, „DOI“, „Faksimile“, „Scan“ oder „Digitalisat“ können je nach Äußerungskontext in verschiedenen Abteilungen etwas anderes bedeuten.

2) *Mensch-Maschine-Schnittstellen* führen dann zu Reibungen, wenn der Mensch mit unvertrauter Software oder kontraintuitiven grafischen Interfaces konfrontiert wird, deren Benutzung u. U. erst ermöglicht oder erlernt werden muss. Analog zu abteilungsspezifischen Fachjargons ist die Expertise abteilungsspezifischer Formate und Datenmodelle zur (Meta-)Datenerfassung: Während bei der Katalogisierung von Archivalien das EAD-Format² bekannt ist, wird bei der Digitalisierung durch Bibliotheken METS/MODS (vgl. Altenhöner u.a., 2023) verwendet, bei der Transkription TEI und bei der semantischen Anreicherung schließlich RDF. Da man die Arbeiten der im Workflow jeweils vorhergehenden Abteilung nachnutzen will, muss jeweils voneinander gelernt werden, was in einem bestimmten Schritt von wem erfasst wurde. Ein gemeinsames Verständnis muss hier an den Schnittstellen der Abteilungen aufgebaut werden. Die Formate dienen am Ende dazu, von Menschen geschaffene Information in eine maschinenlesbare Form zu gießen.

3) Eine ebenso wichtige Rolle spielen Formate für die Interoperabilität von Softwaresystemen, womit sie als eine wichtige *Maschine-Maschine-Schnittstelle* aufgefasst werden können. Auf basaler Ebene stehen sich XML und JSON als Formatparadigmen gegenüber: Während XML-basierte Formate zu dokumentarischen Zwecken bevorzugt eingesetzt werden, überwiegt JSON als Datenaustauschformat für Webschnittstellen. Prominent sind METS/MODS und TEI für das XML-Paradigma und IIIF³ oder die Formate von Suchmaschinen-APIs für das JSON-Paradigma. Eine

Aufgabe besteht darin, Daten zwischen den Paradigmen automatisch zu konvertieren. So müssen etwa die bei der Digitalisierung in METS/MODS enthaltenen Bildmetadaten in ein IIIF-konformes JSON-Manifest konvertiert werden, um die Digitalisate für die Edition über einen IIIF-Server bereitzustellen. Auch innerhalb eines Paradigmas stehen verschiedene Konvertierungen (etwa von TEI-XML nach HTML).

Im Folgenden wollen wir beispielhaft zwei große Aspekte des Projekts anhand der in ihnen vorkommenden unterschiedlichen Schnittstellen thematisieren: zum einen die Katalogisierung/Digitalisierung des Scholz-Nachlasses, zum anderen die Texterkennung verschiedener Nachlassmaterialien.

Katalogisierung und Digitalisierung

Die Katalogisierung und Digitalisierung des umfangreichen Scholz-Nachlasses stellt eine Herausforderung dar. Einerseits wurden neue Werkzeuge eingeführt und bestehende Workflows der Bereitstellung digitalisierter Kulturgüter erweitert, andererseits mussten die Kommunikationswege für den interdisziplinären Austausch des Projektteams gefunden werden. Da der Nachlass zu Projektbeginn nur grob strukturiert war, wurde eine im Vorfeld erstellte Findliste (vgl. Heitfeld-Rydzik u.a., 2022) herangezogen. In regelmäßigen Abstimmungstreffen konnte so ein gemeinsamer Startpunkt für die Feinsortierung und Umsystematisierung von Dokumenten und Objekten einerseits, die Festlegung eines Startpunkts für die zeitgleiche Katalogisierung und Digitalisierung andererseits festgelegt werden. In der Folge wurde der Nachlass in Abschnitte geteilt und zyklisch feinsortiert. Damit die erschlossenen Materialien mit DH-Methoden prozessiert werden können, wurde zunächst eine Mensch-Mensch-Schnittstelle geschaffen. Damit war es möglich, Arbeitsschritte in den Bereichen Aushebung, Katalogisierung und Digitalisierung zu parallelisieren, was eine frühe Bereitstellung erster digitalisierter Dokumente ermöglichte.

Der weitere Erschließungsprozess in der ULB ist durch zwei voneinander abhängige Mensch-Maschine-Interaktionen geprägt: die Katalogisierung und die Digitalisierung. Durch die Katalogisierung in der Kalliope-Verbunddatenbank⁴ wird das Wissen der Aushebung und Feinsortierung konsolidiert und standardisiert in Katalogaufnahmen und Strukturbäumen festgeschrieben. Aufbauend auf den erzeugten Metadaten der Katalogisierung werden die Nachlassdokumente digitalisiert (vgl. Altenhöner u.a., 2023). Während die Digitalisierung selbst als Interaktion der Scanoperator*innen mit physischen Nachlassobjekten und dem Digitalisierungssystem geprägt ist (Mensch-Maschine-Schnittstelle), findet im Hintergrund ein Informationsaustausch der Systeme über verschiedene maschinelle Schnittstellen statt. Granulare Identifier⁵ werden vergeben, Metadaten und Strukturinformationen ausgetauscht. Diese Maschine-Maschine-Interaktion ermöglicht, das Wissen über die physikalischen Dokumente (Katalogaufnahmen),

die festgelegte Nachlass-Tektonik und die Digitalisate in einem System, hier der Visual Library⁶, zusammenzuführen. Die Visual Library ist dabei jedoch mehr als das Zielsystem des Teilworkflows. Über die definierten Maschine-Maschine-Schnittstellen wie OAI-PMH für die METS/MODS-Daten oder IIIF wird das Portal *Kulturgut Digital*⁷ der ULB Münster zur Datenmanagementplattform für die automatisierte Bereitstellung der erschlossenen Nachlassdokumente der digitalen Edition.

Erzeugte Digitalisate und Metadaten durchlaufen von dem Zeitpunkt ihrer Entstehung bis zu ihrer Repräsentation als Linked Open Data verschiedene Systeme und Anwendungen. In jedem Bearbeitungs- und Anreicherungsschritt entstehen unterschiedliche Nutzformate, die die vorhandenen Informationen aufgreifen, erweitern oder zusammenführen. Um diese Abhängigkeiten des Workflows von spezifischen Eingangs- und Ausgangsformaten einerseits, die Flexibilität des wissenschaftlichen Erkenntnisgewinns andererseits aufrecht zu halten, werden Mechanismen der Versionierung von Anfang an mitgedacht und verankert (s.u.). Mithilfe gezielter Maschine-Maschine-Interaktionen, aber auch über definierte Schnittstellen (SRU, OAI-PMH, IIIF), Standardisierung und dem Zusammenführen zentraler Informationen aus verschiedenen Nutzformaten wird eine einheitliche und reproduzierbare Datengrundlage für den Editionsprozess geschaffen, die im Folgenden näher erläutert wird.

Texterkennung

Texterkennung im Scholzprojekt geschieht sowohl in der Erschließung der Werke Heinrich Scholz', die bereits in gedruckter Form vorliegen, als auch bei der Erschließung des knapp 100.000 Einzelseiten umfassenden handschriftlich wie maschinenschriftlich verfassten Nachlasses. Bei den gedruckten Werken konnte sofort mit dem Scannen begonnen werden. Zur Distribution der Scans innerhalb des Projekts wird die Datenaustauschplattform Sciebo genutzt, das eine Cloud-Speicherung auch größerer Datenmengen sowie kollaborative Bearbeitung ermöglicht. Auch das Scannen des Nachlasses konnte nach umfangreichen Vorbereitungen mittlerweile beginnen.

Die Scans der veröffentlichten Scholz-Rezensionen wurden zur Qualitätsverbesserung zunächst mit dem Tool ScanTailor Advanced⁸ nachbearbeitet und schließlich mit der OCR-Software tesseract⁹ weiterverarbeitet. In der OCR-Erkennung und den dafür notwendigen Workflows zeigen sich entscheidende Mensch-Maschine-Schnittstellen: ScanTailor Advanced und tesseract wurden beide als Kommandozeilenprogramme entwickelt. Um die Zugänglichkeit für studentische Hilfskräfte zu verbessern, haben wir zwei Skripte geschrieben, die eine Bedienung via grafischer Nutzeroberfläche ermöglichen.¹⁰

Das Skript für ScanTailor fordert die User zunächst zur Auswahl einer TIF-Datei auf, falls diese nicht bereits als Kommandozeilenargument vergeben wurde. Das Skript erstellt einen Ordner mit einer Kopie der zu bearbeiten-

den Datei und öffnet ScanTailor Advanced. Der ScanTailor-Output (mehrere einseitige TIF-Dateien) wird in eine einzelne PDF-Datei umgewandelt. Eine vorkonfigurierte Maschine-Maschine-Schnittstelle im Post-Processing ermöglicht eine anschließende Ausführung der PDF-Datei im zweiten Skript: der Texterkennung. Eine grafische Benutzeroberfläche (vgl. Abb. 2) erlaubt die Angabe von Trainingsdatensätzen, die je nach Schriftart und Sprache variieren, und die Auswahl von dateispezifischen Optionen wie automatisches Entzerren oder Überschriften eines bereits vorhandenen OCR-Layers. Ein Klick auf den Startbutton erzeugt aus den Angaben den entsprechenden Kommandozeilenbefehl für ocrmypdf und führt ihn im Hintergrund aus.

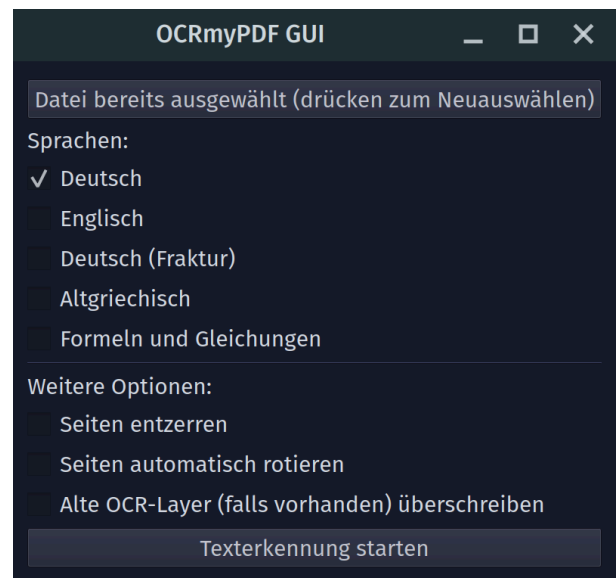


Abbildung 2: GUI für OCRmyPDF

Auch der Ablauf der Nachlassaufbereitung hält herausfordernde Schnittstellen bereit, etwa zwischen Maschine und Maschine. Da OCR4all (vgl. Reul et al. 2019) vorab Informationen benötigt, um welchen Schrifttyp es sich bei dem jeweils zu erkennenden Scan handelt, sind zusätzliche Vorverarbeitungsschritte notwendig. Insgesamt gestaltet sich der schriftliche Nachlass in Bezug auf Art, Größe und Schrifttyp ausgesprochen heterogen. Angesichts des Nachlassumfangs ist eine automatisierte Lösung zu finden. Da OCR und Handschriftenerkennung an den Schrifttyp angepasste Modelle erfordern, muss eine entsprechende Vorselektion erfolgen. Diese Aufgabe umfasst ggf. auch das Zergliedern einer einzelnen Seite in mehrere Einheiten (vgl. Abb. 3).

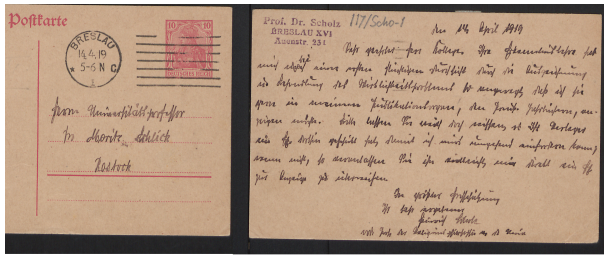


Abbildung 3: Postkarte vom 14.04.1919, Heinrich Scholz an Moritz Schlick, Testscan aus dem Nachlass Scholz, Universitäts- und Landesbibliothek Münster

Zur Segmentierung der Seiten bietet OCR4all zwei Möglichkeiten: Zum einen die Einbindung der selbstlernenden Layout-Erkennungssoftware LaReX, zum anderen ein Interface zur händischen Segmentierung nach Layout. Diese händisch markierten Seiten dienen außerdem als Trainingsmaterial für die nachfolgende maschinelle Layout-Erkennung. Während die grafische Benutzeroberfläche die Ausführung der händischen Segmentierung erleichtert und beschleunigt, waren Vorschläge zur automatischen Selektion unzureichend. Ein Teil wird daher manuell von Hilfskräften als Goldstandard und Trainingsgrundlage bearbeitet, auf dessen Grundlage geprüft werden kann, ob sich die Vorschläge von LaReX im Laufe der Arbeit verbessern. Die erkannten Texte sowie Metadaten werden anschließend in TEI-XML codiert und mit dem dazugehörigen Scan verknüpft. Eine statistische Bewertung der OCR-Qualität wird für verschiedene Schrift- und Dokumenttypen vorgenommen (vgl. Neudecker et al., 2021, 138–165).

Eine besondere Herausforderung besteht in der Erkennung der zahlreichen im Material befindlichen mathematischen Formeln. Die genaue Art und der Detailgrad der zu erfassenden Informationen muss bestimmt werden. Ob Formeln in der Textfassung nur als solche abgebildet oder auch weitere semantische Informationen hinterlegt werden sollen, hat im weiteren Forschungsverlauf Konsequenzen für sämtliche Typen von Schnittstellen.

Im Bereich der Maschine-Maschine-Interaktion stellt die Datenhaltung einen besonderen Aufgabenbereich dar, der im Hintergrund an jeder Stelle des Projekts über den gesamten Projektzeitraum von essentieller Bedeutung ist. Zu Beginn der Projektarbeiten mussten bereits die zentralen Pfeiler dieser Architektur bestimmt und implementiert werden. Die Daten, die im Projekt geschaffen werden, werden von zahlreichen verschiedenen Zugangspunkten eingespeist und später zusammengeführt. Alle IT-Subsysteme im Projekt sind durch die gemeinsame Versionierung verbunden. Neben der punktgenauen Zusammenführung der verschiedenen Dateien gibt es einige weitere spezifische Anforderungen, die an ein Versionierungssystem zu stellen sind, um den Anforderungen des Projekts gerecht zu werden. Zum einen ist eine Datenmenge von mehreren Terabyte zu erwarten, zum anderen verteilt sich diese Datenmenge auf viele einzelne Dateien. Eine Langzeitsicherung muss von Anfang an mitbedacht werden, da die Projektlaufzeit mit 13 Jahren den Zeitraum, den Projektdaten allgemein verfügbar vorgehalten werden, bereits überschreitet.

Aus diesen Gründen fiel die Entscheidung auf das auf git-annex¹¹ basierende Versionierungs-Tool datalad¹², in dem der datenbasierte Projektworkflow vollständig abgebildet werden soll.

Diskussion und Ausblick: Workflows, Absprachen und Interoperabilität

Wenn man eine digitale Edition nicht auf dem leeren „Papier“, sondern innerhalb vorhandener Infrastrukturen und ihrer jeweiligen Vorgaben beginnt, entstehen zahlreiche potenzielle Reibungsflächen, die wir in unserem Beitrag als vielfältige Schnittstellen zwischen Menschen und Maschinen beschreiben. Es ist für den Projekterfolg relevant, sich bereits in der Konzeption von Arbeitsschritten und Workflows bewusst zu machen, dass Menschen mit unterschiedlichen disziplinären Hintergründen, auf verschiedenen Ausbildungsstufen und mit divergierenden Zugangsrechten nicht nur untereinander eine gemeinsame Sprache finden müssen (Mensch-Mensch-Schnittstelle), sondern auch unterschiedliche Voraussetzungen haben, um mit digitalen Tools umgehen zu können (Mensch-Maschine-Schnittstelle): Einige Nutzer*innen können etwa mit Kommandozeilentools umgehen, andere benötigen eine grafische Oberfläche. Diese Punkte einzukalkulieren und mit einer fundierten Planung insbesondere der Übergabepunkte zu begegnen, ist ebenso wichtig wie auf technischer Ebene über Schnittstellen eine Interoperabilität der eingesetzten (ggf. proprietären) Tools zu erreichen (Maschine-Maschine-Schnittstelle).

Beim vorgestellten Akademieprojekt zum Nachlass Scholz zeigten sich diese Herausforderungen von Beginn an. Auch nach der Digitalisierung, Katalogisierung und Texterkennung werden mit steigender Komplexität der erzeugten Daten und Aufgaben potenzielle Reibungsflächen einzuplanen sein: Die erkannten Drucke und Handschriften sollen in TEI-XML editiert (Schnittstelle: Page-XML zu TEI-XML) und schließlich als Linked-open-Data in RDF modelliert werden (vgl. Wettlaufer 2018). Während Austauschformate für die Interoperabilität auf Datenebene sorgen, wird bei einer API über Protokolle die Kommunikation der Daten geregelt. Wir gehen davon aus, dass in den meisten Fällen das im Web omnipräsente HTTP-Protokoll auch bei uns die größte Rolle spielen wird. Aber ebenso wie TEI auf XML aufsetzt, so setzt jede Web-API zwar auf HTTP auf, jedoch in jeweils spezieller Ausprägung.

Vorhaben und Dauer des Projekts drängen eine genauere Auseinandersetzung mit den zahlreichen vergleichbaren DH-Editionsprojekten bezüglich angewandter Technologien, Standards und Tools geradezu auf. Im einschlägigen Akademien-Vorhaben „Alexander von Humboldt auf Reisen – Wissenschaft aus der Bewegung“ (vgl. Kraft und Dumont 2020) an der Berlin-Brandenburgischen Akademie der Wissenschaften entsteht die Edition *humboldt di-*

gital (Laufzeit 2015–2032).¹³ Mit Fokus auf X-Technologien werden u.a. Reisejournale, Tagebücher und auch Korrespondenzen in TEI-XML ediert. Als Editions-umgebung kommt hierbei *ediarum*¹⁴ zum Einsatz, in der *oXygen* als XML-Editor und die XML-Datenbank *existdb* als zentrales Repositorium fungiert. Das Münsteraner Scholz-Projekt setzt ebenfalls *oXygen* ein, aber als zentrales Repositorium die Versionsverwaltungssoftware Git, da die Versionskontrolle im Editionsprozess von größerer Relevanz ist als die Suchmöglichkeiten, wie sie primär von XML-Datenbanken bereitgestellt werden. Anders als bei *humboldt digital* werden der Semantic-Web-Technologie für die Erfassung von Daten in RDF sowie ihre Bereitstellung über einen SPARQL-Endpoint eine zentrale Rolle zugemessen.

Im ersten Teilprojekt – der Edition der Briefe aus dem Nachlass Heinrich Scholz – lassen sich in Materialität, Ziel und Methodik beispielsweise auch Parallelen zum an der Universität Hamburg durchgeführten Projekt *Dehmel Digital* ziehen (vgl. Nantke et al., 2022).¹⁵ Von der Nachnutzung bibliothekarischer Katalogdaten (HANS/Kalliope) über die automatische Auszeichnung (NER) bis hin zur Modellierung als Netzwerk setzt das Münsteraner Projekt ähnliche Akzente.

Möchte man gegenwärtig eine digitale Edition erstellen, die auch noch in 20 Jahren und länger funktioniert, lohnt sich ein Blick auf Projekte, die bereits sehr lange laufen und vor längerer Zeit zukunftsfähige Software- und Format-Entscheidungen treffen mussten. Das bereits seit 1956 in Bearbeitung befindliche Akademieprojekt zur Leibniz-Edition (Hannover, Münster, Potsdam und Berlin)¹⁶ entspricht hinsichtlich der Datenmodellierung zwar nicht den heutigen Standards, hat mit seiner Wahl der Software *TUSTEP*¹⁷ aber dennoch Weitsicht bewiesen. Neben der Tool-Wahl ist hinsichtlich der Langlebigkeit eines Datenprojektes wie der digitalen Edition immer auch die Beachtung von langzeitarchivierbaren Datenformaten entscheidend. Die HBZ gibt hier wertvolle Einschätzungen.¹⁸

Fußnoten

1. Vgl. www.heinrich-scholz.de (22.06.2023).
2. Encoded Archival Description; vgl. <https://www.loc.gov/ead/> (30.06.2023).
3. Vgl. in diesem Zusammenhang etwa die Diskussion von Mertens (2021).
4. Vgl. <https://kalliope-verbund.info/> (21.06.2023), vgl. Grothe 2006.
5. Für Digitalisate und Metadaten werden granulare URNs vergeben, Kalliope definiert für die Katalogaufnahmen eigene persistente Identifier (Kalliope-IDs). Vgl. Sommer u.a. 2008 sowie <https://kalliope-verbund.info/de/support/sru.html> (21.06.2023).
6. Vgl. https://www.semantics.de/visual_library/ (21.06.2023).
7. Vgl. <https://sammlungen.ulb.uni-muenster.de/> (29.06.2023).

8. Vgl. <https://github.com/4lex4/scantailor-advanced> (30.06.2023).
9. Vgl. <https://github.com/tesseract-ocr/tesseract> und <https://ocrmypdf.readthedocs.io> (30.06.2023).
10. Vgl. <https://codeberg.org/opensource-philosophy/PDFPostProcessingGUI> (30.06.2023).
11. <https://git-annex.branchable.com/> (29.11.2023).
12. <https://www.datalad.org/> (29.11.2023).
13. Vgl. <https://edition-humboldt.de/> (14.11.2023).
14. Vgl. <https://www.ediarum.org/> (14.11.2023).
15. Vgl. <https://dehmel-digital.de/> (14.11.2023).
16. Vgl. <https://www.uni-muenster.de/Leibniz> (14.11.2023).
17. Vgl. <http://www.tustep.uni-tuebingen.de/> (14.11.2023).
18. Vgl. <https://www.hbz-nrw.de/produkte/langzeitverfuegbarkeit/langzeitverfuegbarkeit-fuer-hochschulen-v1/lzv-dateiformatliste> (14.11.2023).

Bibliographie

- Altenhöner, Reinhard, Andreas Berger, Christian Bracht, Paul Klimpel, Sebastian Meyer, Andreas Neuburger, Thomas Stäcker und Regine Stein.** 2023. *DFG-Praxisregeln 'Digitalisierung'*. Aktualisierte Fassung 2022. Zenodo. DOI: 10.5281/zenodo.7435724.
- Grothe, Ewald.** 2006. „Die kooperative Erschließung von Autographen und Nachlässen im digitalen Zeitalter. Probleme und Perspektiven“. In *Bibliothek. Forschung und Praxis* 30.3: 283–289. DOI: 10.1515/BFUP.2006.283.
- Heitfeld-Rydzik, Birgit, Ingeburg Abdul Wahed und Jens Brumann.** 2022. *Nachlass Scholz / Sammlung Frege: Findlisten*. [Electronic ed.]. URN: urn:nbn:de:hbz:6-93009597846.
- Kraft, Tobias und Stefan Dumont.** 2020. „The Humboldt Code“. In *Wiener Digitale Revue* 1: Tagebuch. DOI: 10.25365/WDR-01-03-02.
- Mertens, Ina.** 2021. „Zwei Seiten einer Medaille – IIF und die Arbeit mit digitalen Bildbeständen“. In *Zeitschrift für digitale Geisteswissenschaften*. Wolfenbüttel. DOI: 10.17175/2021_002.
- Molendijk, Arie L.** 2022. „The troubled Life of Heinrich Scholz“. In *Journal for the History of Modern Theology / Zeitschrift für Neuere Theologiegeschichte* 29.2: 316–349. DOI: 10.1515/znth-2022-0016.
- Nantke, Julia, Sandra Bläß und Marie Flüh.** 2022. „Literatur als Praxis: Neue Perspektiven auf Brief-Korrespondenzen durch digitale Verfahren“. In *Digitale Verfahren in der Literaturwissenschaft*, hg. von Jan Horstmann und Frank Fischer. Sonderausgabe #6 von Textpraxis. Digitales Journal für Philologie. DOI: 10.17879/64059432335.
- Reul, Christian, Dennis Christ, Alexander Hartelt, Nico Balbach, Maximilian Wehner, Uwe Springmann, Christoph Wick, Christine Grundig, Andreas Büttner und Frank Puppe.** 2019. „OCR4all — An open-source tool providing a (semi-) automatic OCR workflow for

historical printings“. In *Applied Sciences* 9.22. DOI: 10.3390/app9224853.

Sommer, Dorothea, Christa Schöning-Walter und Kay Heiligenhaus. 2008. „URN Granular: Persistente Identifizierung und Adressierung von Einzelseiten digitalisierter Drucke. Ein Projekt der Deutschen Nationalbibliothek und der Universitäts- und Landesbibliothek Sachsen-Anhalt“. In *ABI Technik* 28.2: 106–114. DOI: 10.1515/ABITECH.2008.28.2.106.

Turing, Alan M. 1937. „On Computable Numbers, with an Application to the Entscheidungsproblem“. In *Proceedings of the London Mathematical Society* s2-42.1: 230–265. DOI: 10.1112/plms/s2-42.1.230.

Wettlaufer, Jörg. 2018. „Der nächste Schritt? Semantic Web und digitale Editionen“. In *Digitale Metamorphose: Digital Humanities und Editionswissenschaft*, hg. von Roland S. Kamzelak und Timo Steyer. DOI: 10.17175/sb002_007.