

Digitale Editionen und Auszeichnungssprachen. dhmuc-Workshop

Veranstalter: Eckhart Arnold, Bayerische Akademie der Wissenschaften; Mark Hengerer, Ludwig-Maximilians-Universität München; Julian Schulz, Ludwig-Maximilians-Universität München

Datum, Ort: 21.11.2016–22.11.2016, München

Bericht von: Isabella Hödl-Notter / Pia Zachary, Ludwig-Maximilians-Universität München

Am 21. und 22. November 2016 fand in der Bayerischen Akademie der Wissenschaften der vom Münchner Arbeitskreis für Digitale Geisteswissenschaften (dhmuc) organisierte Workshop „Digitale Editionen und Auszeichnungssprachen“ statt. Ziel der Organisatoren Eckhart Arnold (München), Mark Hengerer (München) und Julian Schulz (München) war, sich über die Chancen von digitalen Editionen interdisziplinär auszutauschen, editions-erprobte Auszeichnungssprachen und Datenformate kennen zu lernen, die institutionellen Voraussetzungen digitaler Editionsprojekte zu diskutieren und die Möglichkeiten und Herausforderungen zur Integration der Digital Humanities in die Lehre zu thematisieren.

Im ersten Block des Workshops ging es darum, die Anforderungen von Seiten der Fachwissenschaften, also denjenigen, die digitale Editionen nutzen, zu bestimmen. Den Auftakt machte KAREN RADNER (München) mit einem persönlichen Erfahrungsbericht zur digitalen Edition von Keilschriften, einem Langzeitprojekt, das seinen Anfang in den 1960er-Jahren nahm. Angesichts der langen Dauer derartiger Projekte sei langfristige Planung unerlässlich, genauso die nötige Infrastruktur, die über diesen Zeitraum und danach zur Verfügung stehen müsse. Karen Radner kam zu der Schlussfolgerung, dass angesichts der Zusammenarbeit mit Informatikern und wegen der Arbeit mit anderen Wissenschaftlern die Fähigkeit zur Teamarbeit fundamental sei. Schließlich wies sie auf die Kurzlebigkeit von Computerprogrammen hin und plädierte für die Verwendung möglichst einfacher Dateiformate (wenn es ausreicht am besten einfacher ASCII-Text), um relativ unabhängig von

technischen Neuerungen und möglicherweise kurzlebigen kommerziellen Angeboten zu agieren.

HELENA LEITHE-JASPER (München) stellte das Kommissionsprojekt „Mittelateinisches Wörterbuch“ (MLW) der Bayerischen Akademie der Wissenschaften vor. Als Korpus-basiertes Wörterbuch nutzt das MLW Texteditionen als Quellen. Die Suche nach Wortformen statt Wörtern, nach Syntagmen sowie nach Textabweichungen im kritischen Apparat sei ein Desiderat. Weiterhin müsse die Zitierfähigkeit und konkrete Überprüfbarkeit am Originaltext via Seitenbild gewährleistet werden, welche bei führenden Fachdatenbanken lateinischer Texte noch fehle.

Eine Vorstellung von innovativen Suchfunktionen gab MAX HADERSBECK (München) anhand der FinderApps „WittFind“ und „GoetheFind“. Er rückte die domänenspezifische Suche – deswegen sprach er von Apps – in den Vordergrund, da allgemeine Suchmaschinen für die speziellen Fragestellungen und die Komplexität von Editionen nicht geeignet seien. Grundlage für die vorgestellten Apps bildeten ein elektronisches, multilinguales Vollformen- und Phrasenlexikon. Lokale Grammatiken helfen, komplexe sprachliche Strukturen zu erkennen und somit zu einem besseren Suchergebnis zu gelangen. Zum Abschluss seines Vortrages stellte Max Hadersbeck schließlich noch die Videofunktion im Faust-Projekt vor. Die Einbindung von Filmsequenzen und weiteren ergänzenden Medien sei auch ein Mittel, insbesondere Schüler an vertiefte wissenschaftliche Auseinandersetzung heranzuführen.

Im folgenden Teil des Workshops wurden digitale Werkzeuge und Plattformen zur Erstellung von Editionen vorgestellt. MARK HENGERER (München) und GERHARD SCHÖN (München) stellten den Online-Editor des Projektes „Die Briefe Erzherzog Leopold Wilhelms an Kaiser Ferdinand III. aus dem Reichsarchiv Stockholm“ vor. Die besondere historische Relevanz der Briefe liege an der spärlichen Überlieferung kaiserlicher Privatkorrespondenz sowie der fachdisziplinübergreifenden Informationsdichte zur frühneuzeitlichen Emotions-, Kultur-, Kunst- und Mentalitätsgeschichte.

Für die Transkription der Briefe wurde ein webbasierter Online-Editor auf Grundlage von XML und MySQL entwickelt. Aufgrund der hinterlegten Datenmenge erarbeitet Gerhard Schön eine vektorbasierte Software zur Erkennung von Handschriften. Eine weitere innovative Funktion ist die beim Transkribieren aufscheinende Anzeige des Schriftbilds von bereits transkribierten Wörtern mit gleichen und ähnlichen Buchstabenfolgen.

Das Programm Transkribus wurde von BARBARA DENICOLO (Innsbruck) vorgestellt. Dieses Programm will die Computerwissenschaften, Archive, Geisteswissenschaften und die interessierte Öffentlichkeit zusammenbringen. Mit Hilfe der Software READ (Recognition and Enrichment of Archival Documents) ist es möglich, nach etwa 30 bis 50 transkribierten Seiten ein Training laufen zu lassen, das eine automatische Transkription des weiteren Textes mit einer Fehlerquote von nur etwa 10 Prozent erlaube. Das erfordert zwar immer noch händische Nacharbeit, die aber zeitsparender ausfällt als eine komplette manuelle Transkription. Transkribus kann bereits Schrift in mehreren Sprachen nach Training automatisiert erkennen. Künftig soll unter anderem ein vereinfachtes Webinterface den Anwendern zur Verfügung stehen.

SUSANNE HAAF (Berlin) und ALEXANDER GEYKEN (Berlin) stellten eingangs das Deutsche Textarchiv an der Berlin-Brandenburgischen Akademie der Wissenschaften vor, das sich wegen seiner sehr sorgfältigen Transkriptionen und automatischer linguistischer Analysen besonders an Sprachwissenschaftler richtet. Anschließend ging es den Referenten vorrangig um Interoperabilität, was TEI-XML per se nicht liefert und deswegen DTABF zur Auszeichnung historischer Texte entwickelt wurde. Es legt formale Dinge wie Fettdruck, inhaltliche (wie indirekte Rede), Metadaten und Transkriptionsrichtlinien fest. Dadurch werden die Elemente von TEI reduziert und spezifiziert. DTABF ist inzwischen ein von der DFG empfohlenes Format.

Etwas technischer wurde die Diskussion im anschließenden Themenblock des Workshops, in welchem Auszeichnungssprachen für die Eingabe von Editionen und geeignete Datenbanktechnologien zur ihrer Speicher-

ung erörtert wurden. ECKHART ARNOLD (München) und STEFAN MÜLLER (München) loteten die Möglichkeiten und Grenzen domänenspezifischer Sprachen aus. Im Gegensatz zu XML und TEI bestehen die domänenspezifischen Sprachen (DSL) aus selbstgesetzten Regeln, die auf den Anwendungszweck hin zugeschnitten werden. Der Vorteil von DSL liege insbesondere darin, dass sie pragmatisch, syntagmatisch und sprachökonomisch seien. Auszeichnungen in DSL seien so gehalten, dass eine Konvertierung in XML automatisch erfolgen könne. Anders als XML lieferten DSLs als Eingabetechnologie übersichtliche Quelltexte, die vergleichsweise wenig Tipparbeit erfordern. Die technische Verarbeitung sei allerdings komplizierter als bei XML. Müller und Arnold empfahlen bei der Auswahl der Eingabetechnologie auch DSLs als mögliche Option mit einzubeziehen.

ANDREAS KUCZERA (Mainz) stellte graphdatenbankbasierte Editionen vor. Besonders bei mehreren Deutungsebenen, die in überlappenden Strukturen resultieren, kämen XML-Baumstrukturen an ihre Grenze. Als Grundlage für die Informationsspeicherung dienen Graphen, die aus „Knoten“ und „Kanten“ bestehen (so etwa bei der Graphdatenbank Neo4j). Ein „Text“ wird als Verkettung von Wortknoten modelliert, dem beliebig viele Eigenschaften zugeordnet werden könnten. Somit seien verschiedenen Hierarchie- und Deutungsebenen keine Grenzen mehr gesetzt. Der Wechsel zur Graphdatenbank stelle für Editionsprojekte eine innovative Alternative zur Programmierung in XML dar. Da jedoch XML als Format zur Langzeitarchivierung derzeit noch die beste Variante darstelle, ist die Möglichkeit vorhanden, die Graphdatenbanken in XML-Dateien umzuwandeln und damit nachhaltig zu sichern.

Nicht zuletzt wegen der oft als besser empfundenen Ergonomie¹ sind Druckausgaben bei aktuellen Editionsprojekten nach wie vor erwünscht. Deshalb widmeten sich zwei Vorträge unterschiedlichen technischen Ansätzen für die Transformation des digitalen Aus-

¹ Zum kognitionspsychologischen Hintergrund vgl. Ferris Jabr: The Reading Brain in the Digital Age: The Science of Paper versus Screens, in: Scientific American, 2013 (42), URL: <https://www.scientificamerican.com/article/reading-paper-screens/>

gangsmaterials in eine Druckvorlage. MAXIMILIAN WESTPHAL (München) referierte über den Arbeitsablauf vom digitalen Manuskript zur Druckvorlage der Edition mit InDesign. Gegenüber Textverarbeitungsprogrammen biete diese Software zahlreiche Vorteile, wie Werkzeuge zur Anpassung des Satzatzes, die Erstellung von komplexen Satzvorlagen und der Datenausgabe nach Standards der Druck- und Medienindustrie. So sei bei guter Vorbereitung der eingespeisten Dateien ein fast vollständig automatisierter Arbeitsablauf möglich. Je nach fachlicher Ausrichtung des zu setzenden Textes könnten noch Nachjustierungen im Bereich von Tabellenstrukturen oder Bildplatzierungen anfallen. Um eine hochwertige Typographie zu gewährleisten, müsse dabei der Grundsatz *form follows function* gelten und der individuelle Charakter jedes Projektes anerkannt werden.

OLIVER VON CRIEGERN (München) beschäftigte sich mit einer weiteren Möglichkeit zur Erzeugung des Drucksatzes, nämlich mit LaTeX in Verbindung mit XML. Die typographischen Anforderungen an kritische Editionen seien vielfach sehr hoch, unter anderem aufgrund der Verwendung mehrerer Fußnotenapparaten und Marginalien, können von LaTeX aber realisiert werden. LaTeX unterstützt aber weder semantische Auszeichnungen noch sei die Konvertierung nach HTML für eine Online-Edition trivial. Abhilfe könne durch die Erstellung einer Edition mit XML geschaffen werden, aus der die TeX-Datei automatisiert erstellt wird. Ausgehend davon könnten durch die Verarbeitung von XML mit XSLT eine Bandbreite unterschiedlicher Datenformate erzeugt werden, wie XML, HTML, LaTeX, pdf oder eine Textdatei.

Im fünften und letzten Block der Veranstaltung konzentrierten sich die Vorträge auf den Bereich der Lehre, Ausbildung und institutionellen Verankerung der Digital Humanities in den Geisteswissenschaften. PEGGY BOCKWINKEL (Stuttgart) berichtete als Studiengangskoordinatorin von den ersten Erfahrungswerten des neu eingerichteten Digital Humanities Masterstudiengangs an der Universität Stuttgart. Der Studiengang zeichne sich durch drei Komponenten aus, die aus der Vertiefung eines geisteswissenschaftlichen Faches, der Digital Humanities sowie

der Informatik bestehen. Grundlage für die Herausbildung eines solchen Studiengangs sei die optimale Zusammenarbeit der Fakultäten gewesen. Schwierigkeiten bestünden im Umgang mit der heterogenen Gruppe an Studierenden, die unterschiedliches Vorwissen mitbrächten und den Programmierkursen, die für die Studierenden oftmals eine große Herausforderung darstellten. Dennoch gebe es bislang nur vereinzelt Studienabbrecher, was auch auf das sehr gute Betreuungsverhältnis zurückzuführen sei.

ANNE KLAMMT (Mainz) erläuterte anschließend das Konzept des hochschulübergreifenden Studiengangs „Digitale Methodik in den Geistes- und Kulturwissenschaften“. Die Schwerpunkte des Studiengangs lägen im Bereich der Informatik, der Digital Humanities und ebenso der digitalen Methodik. Letzteres erlaube den Studierenden eine Schwerpunktbildung in Bereichen wie beispielsweise der Musikinformatik oder Korpuslinguistik. Herausforderungen dieses Digital Humanities Studiengangs zeigten sich in der Herstellung von Prüfungsgerechtigkeit, dem Gewinnen von Informatikern für dieses Studium und verfügbarem Lehrpersonal aufgrund der dünnen Personaldecke im Bereich der Digital Humanities.

JULIAN SCHULZ (München) ging der Fragestellung nach, wie viel Technik Geisteswissenschaftler angesichts eines zunehmend auf digitale Fähigkeiten ausgelegten Arbeitsmarkts beherrschen müssten. Dabei zeigte er die Notwendigkeit einer IT-Grundlagenvermittlung für Geisteswissenschaftler auf, die sich nicht zuletzt aus der Kommunikationsfähigkeit mit IT-Fachleuten und der Fähigkeit zur Strukturierung von erhobenen Forschungsdaten ergäben. Um den Studierenden Grundlagen der IT nahe zu bringen, wurde im Rahmen der Förderlinie „Digitaler Campus Bayern“ mit dem DHV-Lab (Digital Humanities Virtual Laboratory) eine virtuelle Ausbildungsplattform geschaffen. Die enge Kooperation des Projekts mit den beteiligten Fachwissenschaften gewährleiste eine Heranführung der Studierenden an die vielfältigen Möglichkeiten, die sich den Fächern durch die Digital Humanities böten. Das DHVLab stellt für die praxisnahe Ausbildung eine virtuelle Rechenumgebung zur

Verfügung, die beispielsweise anwendungs-
bereite statistische Analyseprogramme bereit
hält.

MONIKA UEMMINGHAUS (München)
referierte zu finanziellen und ideellen Unter-
stützungsmöglichkeiten für gute Lehre und
stellte dabei das Multiplikatoren-Programm
als Teil des Projektes Lehre@LMU vor. In ih-
rem Vortrag legte sie den Schwerpunkt auf
die pädagogischen Chancen und Herausfor-
derungen von e-Learning in der Lehre. Ziel
sei es, eine Lehrkultur im Bereich der Lehr-
leistung, Innovation und Wertschätzung des
Individuums zu stärken und damit Lehre
nachhaltig zu verbessern. Aus den bisher-
igen e-Learning-Modellen seien unter ande-
rem die Erkenntnisse gewonnen worden, dass
eine genaue Zielgruppenanalyse seitens der
Lehrenden Voraussetzung für eine erfolgrei-
che Anwendung von e-Learning sei; insbe-
sondere müsse auf eine ausreichende inter-
aktive Anleitung zum Umgang mit den e-
Learning-Tools geachtet werden. Unerlässlich
sei nach wie vor die Kombination von e-
Learning-Einheiten mit Präsenzphasen.

In der ABSCHLUSSDISKUSSION stellten
CHRISTIAN RIEPL (München) und MARK
HENDERER (München) die institutionelle
Entwicklung der IT-Gruppe Geisteswissen-
schaften an der LMU vor und betonten die
Bedeutung der Nachhaltigkeit struktureller
Entwicklung im Bereich der Digital Huma-
nities. Der zweite Teil der Diskussion drehte
sich darum, wie Informatiker für die Koope-
ration mit Digital Humanists gewonnen wer-
den könnten; die Teilnehmer waren sich ein-
ig, dass Begeisterung und Kommunikation
auf Augenhöhe wichtig seien.

Konferenzübersicht:

Erwartungen der Fachwissenschaften an (di- gitale) Editionen

Karen Radner (Alte Geschichte des Nahen
und Mittleren Ostens, LMU München)

Helena Leithe-Jasper (Mittelateinisches Wör-
terbuch, Bayerische Akademie der Wissen-
schaften, München)

Max Hadersbeck (Centrum für Informations-
und Sprachverarbeitung, LMU München)

Digitale Editionswerkzeuge im Einsatz

Mark Hengerer (Geschichte der Frühen Neu-
zeit, LMU München) / Gerhard Schön (IT-
Gruppe Geisteswissenschaften, LMU Mün-
chen): Handschriften-Transkription und Edi-
tion im interdisziplinären Kontext: Das rela-
tionale Datenmodell und die Werkzeugkette
der Leopold-Wilhelm-Edition

Barbara Denicolo (Universität Innsbruck): Die
Erstellung von Handschriften-Editionen mit
„Transkribus“

Auszeichnungssprachen und Datenformate

Susanne Haaf / Alexander Geyken (Deut-
sches Textarchiv, Berlin-Brandenburgische
Akademie der Wissenschaften): Deutsches
Textarchiv: TEI-XML anhand des DTA-
Basisformats

Eckhart Arnold / Stefan Müller (Digital Hu-
manities, Bayerische Akademie der Wissen-
schaften): Jenseits von TEI und XML: Mög-
lichkeiten und Grenzen domänenspezifischer
Sprachen?

Andreas Kuczera (Mainzer Akademie der
Wissenschaften): Graphdatenbankbasierte
Editionen

Druckausgaben aus digitalen Quellen

Maximilian Westphal (Universitätsbibliothek,
LMU München): Vom digitalen Manuskript
zur Druckvorlage mit InDesign

Oliver von Criegern (Digital Humanities,
Bayerische Akademie der Wissenschaften):
Automatisierte Erzeugung des (LaTeX-)
Drucksatzes aus XML-Dokumenten

Lehre und Ausbildung und institutionelle Verankerung

Peggy Bockwinkel (Universität Stuttgart):
Erste Erfahrungen mit dem Digital Huma-
nities Studiengang in Stuttgart

Anne Klammt (mainzed – Mainzer Zentrum
für Digitalität in den Geistes- und Kulturwis-
enschaften): Digitale Methoden studieren in
Mainz – ein neues kooperatives Modell für
Studiengänge im Feld DH

Julian Schulz (IT-Gruppe Geisteswissen-
schaften, LMU München): Wieviel Technik müssen
Geisteswissenschaftler/innen beherrschen?
Das Digital Humanities-Curriculum im

„Digitalen Campus Bayern“

Monika Uemminghaus (Center for Leadership and People Management, LMU München): Unterstützungsmöglichkeiten für gute Lehre: Das Multiplikatorenprojekt

Offene Diskussion: Langfristige institutionelle Verankerung von DH-Großprojekten

Tagungsbericht *Digitale Editionen und Auszeichnungssprachen. dhmuc-Workshop*. 21.11.2016–22.11.2016, München, in: H-Soz-Kult 19.01.2017.