

“Wissensorganisation ist ein Programm” - Workshopbericht

Zusammengestellt von

Claus-Michael Schlesinger (Ernst-Moritz-Arndt-Universität Greifswald, PhD-Net
Das Wissen der Literatur)

Björn Martin (BerlinOnline Stadtportal GmbH & Co. KG, iversity GmbH,
PhD-Net *Das Wissen der Literatur*)

Christian Thomas (BBAW/Deutsches Textarchiv, CLARIN-D, Einstein-Zirkel
Digital Humanities Berlin)

Mirus Fitzner (Universität der Künste, Berlin)

In den letzten Jahren hat in den Kulturwissenschaften die Nutzung persönlicher computergestützter Datenbanken zugenommen. Über die Verwaltung von bibliografischen Referenzen hinaus bieten Programme wie Citavi, verzetteln, Litlink usw. Funktionen der Wissensverwaltung, die sämtliche Phasen wissenschaftlicher Arbeit begleiten und strukturieren sollen. Neben Funktionen für Projekt- und Zeitmanagement zählen dazu vor allem die Datenbankfunktionen, die Exzerpte und Zitate, aber auch Bilder, Filme, eigene Analysen und Kommentare ordnen und miteinander verknüpfen. Auf diese Weise entsteht das, was die Anbieter der Programme selbst meist als *elektronischen Zettelkasten* bezeichnen. Dieser Zettelkasten versammelt dabei nicht nur Adressen, sondern holt zum Teil die referenzierten Objekte selbst als Volltexte, elektronische Bilder usw. in den Speicher. Schnittstellen ermöglichen darüber hinaus die Integration datenbankexterner Ressourcen und netzwerkgebundene Formen der Kollaboration. Im Workshop *Wissensorganisation ist ein Programm*, der vom 6.-8. Juni 2013 an der Humboldt-Universität zu Berlin stattfand, ging es darum, diese Aspekte alltäglicher Arbeitsgeräte in den Geistes- und Kulturwissenschaften vor dem Hintergrund der eigenen Arbeitspraxis systematisch zu untersuchen.

- Wie wird Wissen durch Programme strukturiert?
- Welche Konzepte von Wissen lassen sich an der Gestaltung der Programme (Funktionen, Nutzerführung usw.) ablesen?
- Welcher Wissensform oder welchen Wissensformen entspringen die Programme selbst?

Im Laufe des Workshops wurden diese Fragen nach dem Wissen der Programme anhand spezifischer Funktionen sowohl systematisch abstrakt wie fallbezogen diskutiert. Neben der Nutzung und Nutzbarkeit von Daten im eigenen Arbeitsprozess (Auffinden, Schreiben, Anpassung und Portierbarkeit) führten die Diskussionen zu disziplin- und wissenschaftsrelevanten Konzepten (Disziplinäre Sozialisation, Objektkonstitution).

Auffinden

Das Auffinden von Daten bezieht sich auf mehrere Dinge, erstens das Wiederfinden von Inhalten, zweitens eine Struktur zum Auffinden von Argumenten und passenden Notizen und Zitaten, schließlich drittens die überraschende Begegnung mit einem Zitat oder einer Notiz, die die eigene Gedankenkette in eine produktive, aber unerwartete Richtung lenkt (Luhmanns Gespräch mit dem Zettelkasten). Alle drei Punkte bzw. Probleme/Aufgaben werden durch Metadaten und durch darauf basierende Verknüpfungsstrategien gelöst, die aber von Programm zu Programm unterschiedlich sind und deshalb maßgeblich zu Gestalt und Gestaltung des in der Datenbank dargestellten Wissens zählen. Die Verknüpfungsalgorithmen bzw. die Verknüpfungstabellen bedingen also maßgeblich die Wege durch die eigene Datenbank.

Zu den Verknüpfungsmöglichkeiten einer Datenbank zählt auch die Suche. Durch Suchen werden ad-hoc-Verknüpfungen hergestellt, die sich in der Regel in sucheabhängigen Listen ausdrücken. Auch hier haben die Suchmöglichkeiten einen maßgeblichen Einfluss auf die jeweilige Zusammenstellung von Inhalten, und zwar auf doppelte Weise: erstens über die konkreten Suchprozesse und -ergebnisse, zweitens aber auch schon auf Ebene der Dateneingabe, insofern bei der Eingabe von Inhalten Texte, Notizen, Bilder usw. bereits im Hinblick auf die von der Suche und von den Verknüpfungsfunktionen erfassbaren Eigenschaften überprüft werden bzw. eine entsprechende Aufbereitung erfolgt, sodass von vornherein nur die Daten ihren Weg in die Datenbank finden, die für die je eigenen Suchroutinen relevant sind.

Literatur- und Wissensverwaltung produzieren also mit den codierbaren Eigenschaften, automatischen Verknüpfungen und verfügbaren Suchalgorithmen mögliche Ordnungen von Gegenständen, Beschreibungen, Analysen, Textbausteinen usw. und bedingen damit nicht zuletzt die Umgangsformen, die gegenüber den bearbeiteten Gegenständen ausgebildet werden, sowie die Konzeptionen der Gegenstände selbst (siehe auch [Objektkonstitution](#))

Schreiben

Bei allem universellen Anspruch auf die Begleitung des Arbeitsprozesses vom Entwurf über die Recherche bis zur Textproduktion, der von den Programmen und ihren Herstellern formuliert wird, bleibt der Fokus der Programme auf der

Verwaltung von Gegenständen bzw. von Adressen (in Form von bibliografischen Daten, Exzerpten etc.). Primär geht es darum, Information zu rezipieren und das heißt zu analysieren und zu katalogisieren. Manche Programme (z.B. Litlink, Citavi) bieten Funktionen zur freien Anordnung und Strukturierung der Daten an, z.B. im Rahmen von ergebnisorientierten Projekten und erweitern die Katalogisierungsfunktionen hin auf den Schritt der Gliederung im Schreibprozess. Es bleibt aber deutlich, dass die Programme von der Literaturverwaltung und nicht vom fertigen Text her gedacht sind. Das hängt vermutlich mit ihrer eigenen Genese zusammen. Die Funktionen zur Verwaltung von Wissen in Zettelkastensystemen entwickeln sich historisch aus den Bedürfnissen der Literaturverwaltung (Krajewski). Das gleiche gilt für die elektronischen Zettelkästen, die auch zunächst reine Literaturverwaltungsprogramme sind und erst später mit Funktionen zur Verwaltung von Wissen in Form von Zitaten, Kommentaren, Notizen, Schlagwörtern und Tags sowie mit Funktionen fürs Projektmanagement (Litlink, Citavi) ausgestattet werden. Wissen wird also zunächst in die Form einer Literaturdatenbank gebracht, und die entsprechende Poetik ist erstmal eine Fußnotenpoetik. Das Problem dabei: Fußnoten machen noch keinen Text.

Dieser Umstand führt zu zwei Überlegungen: Erstens gibt es zwischen der Sammlung und Organisation der Daten und der Textproduktion eine Lücke. Diese Lücke zeichnet sich nicht nur durch den Übergang von einem Programm (Datenbank) zu einem anderen (Editor) aus, sondern auch durch die unterschiedliche Art und Weise, mit der Informationen in Datenbanken und in Fließtexten miteinander verknüpft sind. Datenbanken besitzen eine anders gelagerte Logik, Grammatik, Rhetorik und Poetik als Fließtexte (s. auch [Auffinden](#)), die gleichwohl aufeinander bezogen sind, nicht zuletzt durch die beständige Umarbeitung von Fließtexten in Exzerpte und bibliografische Daten, Tags, Schlagworte etc. und die daran anschließende Umarbeitung dieser Formate in Fließtext. Die Differenzen zwischen beiden Gattungen ist in dieser Hinsicht also nicht nur Hindernis auf dem Weg zum eigenen Text, sondern produktiv.

Anpassung, Konversion, Portierung

Anpassungen von Datenbanken setzen umfangreiche technische Kenntnisse voraus und sind deshalb individuell nur schwer zu verwirklichen. gleichzeitig bedingen Design und Formate der Datenbank die Möglichkeiten der Konvertierung und Portierung auf andere Systeme. Hier spielen nicht zuletzt ökonomische Überlegungen der Nutzerbindung als wichtiger Grund für proprietäre Formate eine Rolle. Spätestens damit wird aber klar, dass Arbeitsweisen und Wissensstrukturen in einer Weise technisch und ökonomisch begründet sind, die in den wissenschaftlichen Ergebnissen, also Texten, selten explizit reflektiert werden.

Problematisch für die Nutzerin wird dieser Umstand, wenn das Arbeitsinstrument gewechselt werden soll. Proprietäre Formate fördern die Nutzerbindung zwangsläufig, aber auch die unterschiedliche Ausgestaltung von Datenstrukturen

aufgrund der Komplexität der Wissensverwaltungsfunktionen führt selbst im gleichen Datenbankformat in der Regel zu einem Datenverlust.

Möglichkeiten der Portierung gibt es trotzdem: Da die bibliografischen und anderen Referenzdaten der Datenbank zumindest so ansprechbar sein müssen, dass unterschiedliche Zitierstile verwendet und auch selbst programmiert werden können, lässt sich diese Schnittstelle auch zum Export der Inhalte von anderen Feldern der Datenbank nutzen (z.B. bei Litlink), sodass die Zitierstile Funktionen bei der Datenkonvertierung übernehmen können.

Individuelle und disziplinäre Sozialisation

Die Programme bedingen bestimmte Verhaltensweisen der Nutzerin und privilegieren bestimmte Umgangsformen mit dem Gegenstand. Durch die Verbreitung der Programme lässt sich argumentieren, dass damit auch disziplinär relevante Sozialisationsformen verbunden sind. Universitäten statten über Campuslizenzen und mit Einführungskursen ihre Studierenden und Mitarbeitenden mit spezifischen Programmen aus und fördern damit eine sowohl disziplinäre wie auch allgemein wissenschaftliche Sozialisation im Hinblick auf die Wissensverwaltung.

Software zur Verwaltung von Literatur und Wissen führt positiv zu Strukturen des eigenen Archivs, und unterschiedliche Software führt zu unterschiedlichen Strukturen und auch zu unterschiedlichen Umgangsformen. Die wissenschaftliche Reflexion dieser Strukturen und Umgangsformen ist dabei längst nicht soweit wie die Verbreitung der entsprechenden Programme.

Gerade Nachwuchswissenschaftler_innen werden mit einem schwer überschaubaren Angebot konfrontiert, das vor allem drei Probleme aufweist: Erstens sind zu Beginn der wissenschaftlichen Arbeit im Studium die eigenen Anforderungen, die an das jeweilige Programm gestellt werden können, nur schwer formulierbar. Zweitens steht deshalb das Problem mangelhafter Umstiegs- und Portierungsmöglichkeiten der Wahl eines Programms entgegen, weil sowohl zukünftige eigene Anforderungen wie auch die zukünftige Entwicklung von Features nicht absehbar sind. Und drittens führt dieser Lock-in-Effekt zu einer Hemmung methodischer Innovation, weil die eigenen Arbeitsroutinen im Anschluss an die Software entwickelt werden, während das Ausprobieren und Umgestalten von Umgangsformen erschwert wird.

Objektkonstitution

Eine vieldiskutierte Frage des Workshops war die nach der Konstitution von Objekten in digitalen Datenbanken und in zugehörigen Prozessen der Digitalisierung. Im Raum stand bei den Diskussionen dieses Zusammenhangs der universale Anspruch digitaler Texte: Die Digitalisierung ermöglicht eine umfangreiche und lückenlose Erfassung aller Eigenschaften eines Gegenstands in Form von Daten und Metadaten. Diese Übertragung von konkreten Eigenschaften des Objekts in

seine digitale Repräsentation ermöglicht deshalb sowohl Fragen, die zuvor an das Objekt oder über das Objekt gestellt werden konnten, als auch Auswertungen anderer Art, insbesondere statistischer Natur. In diesem Argument erscheint der digitale Text als ein transzendenter Gegenstand, der qua Codierung und Recodierung alle möglichen Erscheinungsformen annehmen kann. Ein Buch kann wieder zum Buch werden, aber auch zu einem pdf-Dokument, einer Schriftrolle, einer tag-Wolke usw. Dagegen wurde argumentiert, dass diese Verfasstheit des Texts selbst nicht unabhängig von einem Medium und einem materialen Träger ist. Der Text als codierter ist selbst ein spezifisches Objekt und ein spezifischer Gegenstand, der bestimmte Fragen an oder über sich ermöglicht und bestimmte Frageweisen privilegiert.

Definiert man nun Objektklassen oder Objektkonstitutionen nicht über ontologische Eigenschaften, sondern über Fragen, die sich an die je gegebenen Objekte stellen lassen, werden die Schwierigkeiten, die sich aus der materiellen Gebundenheit und der Virtualisierung von Eigenschaften mit weitgehenden Reproduktionsmöglichkeiten ergeben, zwar nicht geringer, sie lassen sich aber nochmal in ein anderes Verhältnis zueinander setzen und markieren sowohl die medialen Differenzen als auch die mediale Gebundenheit digitaler Transzendenz. Digitale Texte unterliegen anderen Produktionsmechanismen, entwickeln andere Poetiken, provozieren andere Lektüretechniken, stellen andere Fragen und geben andere Antworten als gedruckte Bücher.

Gleichzeitig lassen sich Eigenschaften digitaler Datenbanken und Datenverarbeitung aus papiergebundenen Techniken ableiten. Die Tabelle, Grundform aller relationalen Datenbanken, hat eine Wissensgeschichte, die sehr weit in die Geschichte reicht, gleiches gilt für Indices, Konkordanzen und Verschlagwortung, Karteikarten und Zettelkästen. Alles Techniken, die bei der Entwicklung von rechnergestützten Systemen der Wissensverwaltung relevant sind und von den Hintergrundtechniken (Tabelle) bis zum Nutzerinterface (Zettelkasten) noch wirksam sind.

Fazit

Der Workshop hat ein thematisches Spektrum zur Formatierung von Wissen durch Programme der Literatur- und Wissensverwaltung geöffnet. Die beiden Keynotes von Philipp Sarasin und Andrea Rapp setzten mit den zugehörigen Panels Schwerpunkte bei der kulturwissenschaftlich orientierten Einzelplatznutzung der Programme, Möglichkeiten der projektgebundenen Kollaboration in digitalen Infrastrukturen wie TextGrid und DARIAH sowie der Herausforderungen geistes- und kulturwissenschaftlicher Arbeit an und mit Texten durch die digitalen Verfasstheiten der Gegenstände.

Die Diskussionen im Rahmen des Workshops haben gezeigt, dass Software für die Literatur- und Wissensverwaltung neben einer Einschätzung der Nutzbarkeit für definierte Zwecke weitergehend auf ihre Implikationen im Hinblick auf die

Ordnung des Archivs (Datentypen, Verknüpfungsmöglichkeiten, Schnittstellen für die Einbindung anderer Datenbanken, für Vernetzung und für Kollaboration) untersucht werden müssen. Relevant dafür sind erstens die bei der Konzeption der Software zugrunde gelegten Wissenskonzepte. Antizipiert eine Software eine kategoriale, trennscharfe Ordnung der bearbeiteten Texte und Gegenstände und damit ein stärker hierarchisches Ordnungsmodell, oder stehen eher assoziative Ordnungsmöglichkeiten über Schlagworte und Tagging im Vordergrund des User Interface und/oder im Hintergrund der Datenbank? Diese Frage führt, zweitens, neben den konzeptuellen Entscheidungen seitens der Softwareentwicklung zu den Bedingungen und Konsequenzen der eingesetzten Hintergrundtechniken, die als grundlegende Kulturtechniken dabei nicht nur die Wissensverwaltung durch Software, sondern das Wissen der Software selbst begründen. Die tabellarische Strukturierung von Information in relationalen Datenbanken und das neuere Paradigma einer objektorientierten Programmierung sind Beispiele für diese Ebene und bereits Gegenstand mediengeschichtlicher Studien (Markus Krajewski). Im Workshop wurde versucht, beide Ebenen gemeinsam zu reflektieren, um einerseits ausgehend von arbeitspraktischen Regelmäßigkeiten den Grundlagen weiter nachzugehen, andererseits die Routinen über die verwendete Technik zu reflektieren.

Durch den weitgehenden Einfluss auf die Gestaltung und Organisation wissenschaftlicher Arbeit, den die kaum als Medien eigenen Rechts wahrgenommenen Grundlagen der eigenen wissenschaftlichen Arbeit haben, wurde auch ein Bedarf der akademischen **Bildung** deutlich, insofern der Umgang mit Medien der Wissensverwaltung ein integraler Bestandteil wissenschaftlicher Methodenlehre ist, aber selten als solcher behandelt wird.

Offenheit ist deshalb eine wichtige Anforderung an Software. Sowohl die Analyse und wissensorientierte Rezensionen von Software als auch eine auf Sicherheit, Beständigkeit und Flexibilität ausgelegte Verwaltung von Daten sind darauf angewiesen, dass Daten portierbar und Algorithmen für die Analyse und bestenfalls für Weiterentwicklungen zugänglich sind.

Die Teilnehmerinnen und Teilnehmer des Workshops kamen aus ganz unterschiedlichen Bereichen, deren Beschäftigung mit Texten zum Teil eher technisch, zum Teil eher hermeneutisch und insgesamt sehr heterogen ausfiel, was insbesondere bei den Diskussionen zur **Objektkonstitution** deutlich wurde und sowohl den Reflexionsbedarf als auch das Reflexionspotential der Verbindung von Wissensgeschichte und Digital Humanities in den Kultur- und Geisteswissenschaften vor Augen (und auf die Rechner) führte. Deshalb sei an dieser Stelle allen Teilnehmerinnen und Teilnehmern noch einmal herzlich gedankt!

Liste der Teilnehmerinnen und Teilnehmer

Die Ergebnisse des Workshops basieren maßgeblich auf den Diskussionsbeiträgen und der Zusammenarbeit aller Teilnehmerinnen und Teilnehmer. Weil es,

abgesehen von zwei Keynotes, keine Vorträge gab, die im [Programm](#) und in den Ergebnissen auch namentlich erwähnt werden könnten, listen wir an dieser Stelle alle Teilnehmerinnen und Teilnehmer des Workshops auf, alphabetisch nach Vornamen sortiert:

Alexander Czmiel, Andrea Rapp, Björn Martin, Bob Göhler, Christian Thomas, Claus-Michael Schlesinger, Davor Löffler, Eugen Zentner, Karin Krauthausen, Kaspar Renner, Linda Treude, Marcel Nouvertne, Matti Stöhr, Mirco Limpinsel, Mirus Fitzner, Philipp Sarasin, Wilm Grunwaldt.

Kommentierte Linksammlung mit Seiten, die im Laufe des Workshops genannt wurden

Dieser Text ist auch als Ergebnissicherung für die Teilnehmerinnen und Teilnehmer des Workshops gedacht, daher sind die meisten der nachfolgend gelieferten Links in der obigen Zusammenfassung nicht erwähnt, werden hier aber trotzdem aufgeführt.

Culturomics <http://www.culturomics.org/>

Die Webseite ist ziemlich inaktiv. Daher der Artikel „Adventures of a Man of Science“ von Elif Batuman in n+1 über Franco Moretti. <http://nplusonemag.com/adventures-man-science>

Grundlagen für ein Referenzkorpus der neuhochdeutschen Sprache, vorgestellt am Samstag im Workshop Deutsches Textarchiv <http://www.deutschestextarchiv.de>

Digital Humanities im deutschsprachigen Raum <http://dhd-blog.org/>

Digital Humanities in der Geschichtswissenschaft, Überlegungen von Peter Haber http://docupedia.de/zg/Digital_Humanities

Digitalisierter Zugriff auf Archive aus den USA. Digital Public Library <http://dp.la/>

Digital Research Infrastructure for the Arts and Humanities (DARIAH) <https://portal-de.dariah.eu>

Übersicht über die deutsche Digital Humanities-Landschaft mit fachwissenschaftlichen Diensten <https://portal-de.dariah.eu/web/guest/scholarly-services>

und Fachbibliografie zu Digital Humanities <https://portal-de.dariah.eu/web/guest/bibliography>

Einstein-Zirkel Digital Humanities Berlin

Wachsende Sammlung europäischer Archive mit Texten und Bildern sowie Ton- und Filmaufzeichnungen. Europeana <http://www.europeana.eu/>

Sammlung von geistes- und kulturwissenschaftlichen Blogs Hypotheses.org <http://de.hypotheses.org/>

Blog zur Literaturverwaltung, Aktuelles und Analysen zu Literaturverwaltungs- und Wissenssoftware <http://literaturverwaltung.wordpress.com>

MOOCS – Massive Open Online Courses: EdX <http://www.edx.org>

MOOCs – Massive Open Online Courses: Iversity <http://www.iversity.org>

Interaktive Vorlesungen im Netz. Zahlreiche Anbieter auf dem Markt. <http://www.zeit.de/studium/uni-leben/2013-06/mooc-deutschland-iversity-vorlesungen-internet>

Kritische Betrachtung von MOOCs <http://nplusonemag.com/can-venture-capital-deliver-on-the-promise-of-the-p>
von Bob Meister im n+1 Magazin

N-Gram Viewer <http://books.google.com/ngrams>

N-Gram Experimente: <http://www.informationisbeautiful.net/visualizations/google-ngram-experiments>

N-Gram Viewer für Musik – Peachnote <http://www.peachnote.com>

Schöne, witzige oder bemerkenswerte Fundstücke in den von Google digitalisierten Büchern – Art of Google Books <http://theartofgooglebooks.tumblr.com>

Tumblr eines Teils des Dissertationsprojekts (Geschichte des Webs anhand der Geocities-Seiten) von Olia Lialina (Professorin an der Merz-Akademie): One Terabyte of Kilobyte Age <http://oneterabyteofkilobyteage.tumblr.com>

PhD-Net *Das Wissen der Literatur* an der Humboldt-Universität zu Berlin, Kooperation mit den Universitäten Princeton, Harvard, Berkeley (alle USA) <http://www2.hu-berlin.de/wissen-literatur>

In diesem Artikel wird der Neologismus Culturomics erstmals erwähnt: Quantitative Analysis of Culture Using Millions of Digitized Books (Michel et al., Science 331, S. 176-182) <http://pinker.wjh.harvard.edu/articles/papers/Michel%20et%20al%20Quantitative%20analysis%20of%20culture%20Science%202011.pdf>

Projekte der BBAW im Bereich Digital Humanities (Auswahl):

- Telota – The Electronic Life of the Academy <http://www.bbaw.de/telota/>
- DH-Aktivitäten am Zentrum Sprache (Korpusaufbau, Texttechnologie, Lexikographie, Computerlinguistik u. a.): z. B. Deutsches Textarchiv (siehe oben), Digitales Wörterbuch der deutschen Sprache (DWDS, <http://www.dwds.de/>)
- Interdisziplinärer Forschungsverbund Digital Humanities in Berlin <http://www.ifdhberlin.de>
- Digitale Editionen <http://www.bbaw.de/telota/projekte/digitale-editionen>
- Dokumentenserver für Veröffentlichungen von Akademie-Mitgliedern <http://edoc.bbaw.de/>

TextGrid – Virtuelle Forschungsumgebung für die Geisteswissenschaften, v.a. für digitale Editionen <http://www.textgrid.de>

Ulysses auf Rap Genius. Der Volltext als durch die Crowd annotierbare Edition <http://rapgenius.com/albums/James-joyce/Ulysses>

Vergleich von Literaturverwaltungssoftware in der engl. Wikipedia http://en.wikipedia.org/wiki/Comparison_of_reference_management_software