

# Zum Verhältnis von klassischen Formen der Archiverschließung und den Zugängen der Digital Humanities zum Information Retrieval

Georg Vogeler

Was sind die digitalen Geisteswissenschaften? Das Forschungsgebiet entwickelt sich zunehmend zu einer akademischen Disziplin, die etabliertes Wissen in Studiengängen vermittelt und einen angeregten Forschungsdiskurs in Tagungen und Publikationen führt. Dabei ist die inhaltliche Ausrichtung immer noch sehr vielfältig, denn die Schwerpunkte können in der Verarbeitung von Texten und natürlicher Sprache ebenso liegen wie in der digitalen Repräsentation und formalen Erschließung von (nicht nur) schriftlichen Objekten des Kulturerbes. Eine mögliche Beschreibung des Tätigkeitsbereichs der Digitalen Geisteswissenschaften ist die Untersuchung des digitalen und digitalisierten Kulturerbes in seiner menschlichen Aneignung, seiner geisteswissenschaftlichen Reflexion, seiner hermeneutischen Kontextualisierung, seiner informationslogischen Formalisierung, seiner technischen Realisierung und seiner statistischen Berechnung. Pragmatischer können die digitalen Geisteswissenschaften als Auseinandersetzung von Geisteswissenschaftlerinnen mit Computern verstanden werden – auch wenn in der Fachdiskussion die Bezeichnung „Humanities Computing“ seit etwa zehn Jahren durch die „Digital Humanities“ ersetzt worden ist<sup>1</sup>. Im Folgenden soll ein Blick der digitalen Geisteswissenschaften auf das Problem der archivischen Erschließung geworfen werden. Es geht für die digitalen Geisteswissenschaften darum, sich zu fragen, was spezifisch an den Informationen in den Archiven und dem Informationsbedürfnis des Benutzers von Archiven ist, was die Kommunikation zwischen Archiv und Benutzer von Bibliotheken, Banken, Meteorologen oder Gendatenbanken unterscheidet, und wie diese Kommunikation vom Computer geformt wird und werden könnte.

Es kann also zum Beispiel aus dem allgemeinen Wissen über Informationsvisualisierungen auf den Spezialfall von Archiverschließung geschlossen und die Frage gestellt werden, ob Techniken wie Word Clouds, Statistiken, geographische Karten, hierarchische Bäume oder Netzwerke in der Kommunikation zwischen Archiv und Archivbenutzer

---

<sup>1</sup> Zur langfristigen Entwicklung vom „Humanities Computing“ zu den „Digital Humanities“ vgl. Patrik SVENSSON, Humanities Computing as Digital Humanities. *Digital Humanities Quarterly* 3/3 (2009), <http://www.digitalhumanities.org/dhq/vol/3/3/000065/000065.html> [1. 3. 2018]. Der Beitrag ist Teil einer längeren Serie zur Geschichte der Digital Humanities: DERS., The Landscape of Digital Humanities. *Digital Humanities Quarterly* 4/1 (2010), <http://www.digitalhumanities.org/dhq/vol/4/1/000080/000080.html> [1. 3. 2018]; DERS., From Optical Fiber To Conceptual Cyberinfrastructure. *Digital Humanities Quarterly* 5/1 (2011), <http://www.digitalhumanities.org/dhq/vol/5/1/000090/000090.html> [1. 3. 2018]; DERS., Envisioning the Digital Humanities. *Digital Humanities Quarterly* 6/1 (2012), <http://www.digitalhumanities.org/dhq/vol/6/1/000112/000112.html> [1. 3. 2018].

sinnvoll eingesetzt werden könnten. Ein Forschungsergebnis der digitalen Geisteswissenschaften ist dabei, dass ein Blick auf die Objekte des Kulturerbes – im Fall der Archive also Archivalien, im reicher erforschten Fall literaturwissenschaftlicher Analysen Texte – nicht mehr ein detailliertes Studium des Einzelobjekts ist, sondern ein in statistischen und graphischen Darstellungen aggregiertes „distant reading“<sup>2</sup>.

Für die Erschließung von Kulturerbeobjekten beziehen sich die Digitalen Geisteswissenschaften besonders auf das Forschungsfeld der „Informationsrückgewinnung“, in dem Informatiker, Informationswissenschaftler und Computerlinguisten daran arbeiten, wie man den Computer benutzen kann, um textuelle Inhalte, die sich nicht in einer einfachen Zeichenkette eindeutig abbilden lassen können, in großen Datenmengen finden zu können. Das Forschungsfeld existiert schon seit der Mitte des 20. Jahrhunderts und findet insbesondere im Feld der Websuchmaschinen, aber auch für digitale Bibliotheken Anwendung<sup>3</sup>. Obwohl die von den Archivaren im Zuge der Erschließung erstellten „Metadaten“ für Archivalien für diese Methoden der Informationsrückgewinnung einschlägiges Material bilden, sind diese Methoden im Archivkontext bislang noch nicht systematisch diskutiert worden<sup>4</sup>.

Im Folgenden möchte ich ein paar Ideen und Methoden aus den digitalen Geisteswissenschaften für die Kommunikation zwischen Archiv und Benutzer in der Archiverschließung vorstellen. Es wird dabei ebenso um eher kreative Experimente gehen wie um etablierte Verfahren der Informationsrückgewinnung, die in der Archiverschließung noch wenig Verbreitung gefunden haben. Ich stelle dabei den fachwissenschaftlichen Benutzer in den Vordergrund, der das Archiv als Teil einer historischen Forschungsarbeit besucht. Welche der etablierten Verfahren für die Kommunikation zwischen diesem und den Archiven sinnvollerweise berücksichtigt werden können, ergibt sich aus dem Verhältnis zwischen vorhandenen Technologien, allgemeinen Nutzererwartungen und grundsätzlichen Erkenntnissen über die Theorie der Geschichtswissenschaft und des Archivwesens. In der Zusammenschau stellt sich heraus, dass sich die etablierten Verfahren gut mit archivischen Konzepten in der Gestaltung der Systeme zur Informationsrückgewinnung verbinden lassen, obwohl sie in den existierenden Archivinformationssystemen erstaunlich wenig verbunden werden.

Beginnen möchte ich mit zwei Beispielen, die der klassischen Form der Archiverschließung zuwiderzulaufen scheinen. In etablierten Formen der Archiverschließung ord-

<sup>2</sup> Der Begriff ist von Franco Moretti geprägt, der es als computergestützte literaturwissenschaftliche Methode dem klassischen „close reading“ gegenüberstellte: Franco MORETTI, *Graphs, Maps, Trees. Abstract Models for a Literary History* (London–New York 2005); DERS., *Distant Reading* (London–New York 2013).

<sup>3</sup> Zum Forschungsfeld, das auch unter der englischen Bezeichnung „Information Retrieval“ geläufig ist, vgl. Christopher D. MANNING–Prabhakar RAGHAVAN–Hinrich SCHÜTZE, *Introduction to Information Retrieval* (Cambridge u. a. 2008); Winfried GÖDERT–Klaus LEPSKY–Matthias NAGELSCHEIDT, *Informationserschließung und Automatisches Indexieren. Ein Lehr- und Arbeitsbuch* (Berlin–Heidelberg 2012); Qiaozhu MEI–Dragomir RADEV, *Information Retrieval*, in: *The Oxford Handbook of Computational Linguistics*, hg. von Ruslan MITKOV [Online-Vorabpublikation der 2. Auflage], <http://doi.org/10.1093/oxfordhb/9780199573691.013.022> [1. 3. 2018]; Marijn KOOLEN–JAAP KAMPS–Vincent DE KEIJZER, *Information Retrieval in Cultural Heritage. Interdisciplinary Science Reviews* 34 (2009) 268–284.

<sup>4</sup> Eine Fallstudie bieten Silvia SCHENKOLEWSKI-KROLL–Assaf TRACTINSKY, *Archival Description, Information Retrieval, and the Construction of Thesauri in Israeli Archives. Archival Science. International Journal on Recorded Information* 6 (2006) 69–107. In letzter Zeit sind dagegen die Forschungen zur automatischen Extraktion von Text aus Bildern von handschriftlichen Dokumenten in den Vordergrund gerückt; siehe insbesondere das READ-Projekt: READ. Recognition and Enrichment of Archival Documents, <https://read.transkribus.eu/> [1. 3. 2018].

nen Archivarinnen das Archivmaterial mit ihrem historischen Fachwissen sinnhaft und organisieren es in einer Hierarchie. Eine hierarchische Organisation bietet sich insbesondere deshalb an, weil sich die Archivalien im Laufe ihrer Entstehung als Gruppen organisiert haben: Aus Einzelschriftstücken werden Akten, Akten werden in Schrankfächern gelagert, die in Räumen stehen und so die physische Registratur bilden<sup>5</sup>. Die Archiverschließung kann nun diese Vorordnungen benutzen, um Archivalien auch auf höheren Aggregationsstufen beschreiben und verwalten zu können. Aus dieser physisch vorgefundenen Ordnung kann man nun eine sachliche ableiten – falls die Archivalien nicht ohnehin schon nach sachlichen Kriterien gelagert waren. Das Bestreben der Archivare, den historischen Zustand der Akten zumindest beim Übergang im Archiv abzubilden (*respect des fonds*), bevorzugt solche Organisation auch in der Beschreibung. Zusätzlich entwickeln Archivare aber auch eigene hierarchische Ordnungen: In der Archivtekonik werden die Archivbestände zum Beispiel nach Zeitschichten, Behördenorganisation oder Geographie hierarchisch angeordnet. In der inneren Ordnung können Akten auf eine vereinheitlichte Registraturordnung oder auf eine von den Archivaren als nützlich erachtete Ordnung verteilt werden, insbesondere wenn die ursprüngliche innere Ordnung kaum mehr nachvollziehbar ist<sup>6</sup>. Im Folgenden sollen Zugangswege zu Archivalien vorgestellt werden, die als radikale Gegenentwürfe zu solchen Erschließungshierarchien gelten könnten: erstens Experimente mit Zugangswegen, die auf dem Muster der Textanordnung, Vertonung der Scans oder dem reinen Zufall beruhen, und zweitens der sogenannte „Google-Suchschlitz“, also ein Suchinterface, das nur eine Volltextsuche auf Grund vom Benutzer frei gewählter Schlagwörter kennt, wobei die Treffer dieser Suche nach im Detail unbekanntem Algorithmen so sortiert werden, dass die „relevantesten“ Treffer an den Anfang gereiht werden.

Ein Beispiel für einen besonders kreativen Zugang bietet die Webseite des John Latham Archive. John Latham (1921–2006) war ein rhodesisch-britischer Künstler, der als Maler und Bildhauer und insbesondere als Konzept- und Performancekünstler gewirkt hat<sup>7</sup>. Die Webseite verwendet die drei Brüder Karamasow Mitya (für Dimitri), Ivan and Alyosha (für Alexej) als Metaphern für verschiedene Zugriffsarten: Mitya lebt in den Tag hinein und kann deshalb mit einer mehr oder wenig zufälligen Zusammenstellung von Bildern der Archivalien zum Stöbern angeregt werden. Ivan ist ein vernünftiger Zeitgenosse, der eine systematische facettierte Suche mit vordefinierten Beschreibungskategorien bevorzugt. Alyosha entwickelt sein Verständnis der Inhalte des Archivs intuitiv an ungewohnten Darstellungsformen wie einer Übersetzung der Daten in Töne<sup>8</sup>. Einen ebenso kreativen Zugang bietet das online digitalisierte Archiv des Verlags Bloodaxe Books<sup>9</sup>. Der Verlag widmet sich zeitgenössischer Poesie. Die Archivwebseite bietet deshalb als zentralen Zugriffsweg die Werke der verlegten Autoren an. Die im Archiv verwahrten Manu-

<sup>5</sup> Zur Komposition als zentralem Prinzip der behördlichen Schriftgutorganisation vgl. Johannes PAPRITZ, *Archivwissenschaft*, 4 Bde. (Marburg <sup>2</sup>1983) I 248–254.

<sup>6</sup> Zur Komplexität der Erschließungspraxis und zur Funktion von Provenienz- und Registraturprinzip darin vgl. Bodo UHL, Die Bedeutung des Provenienzprinzips für Archivwissenschaft und Geschichtsforschung. *Zeitschrift für bayerische Landesgeschichte* 61 (1998) 97–121.

<sup>7</sup> Zu Latham jüngst: A World View: John Latham, hg. von Amira GAD–Joseph CONSTABLE (London 2017).

<sup>8</sup> Athanasios VELIOS, Creative Archiving. A Case Study from the John Latham Archive. *Journal of the Society of Archivists* 32 (2011) 255–271.

<sup>9</sup> The Bloodaxe Archive, <http://bloodaxe.ncl.ac.uk/> [1. 3. 2018].

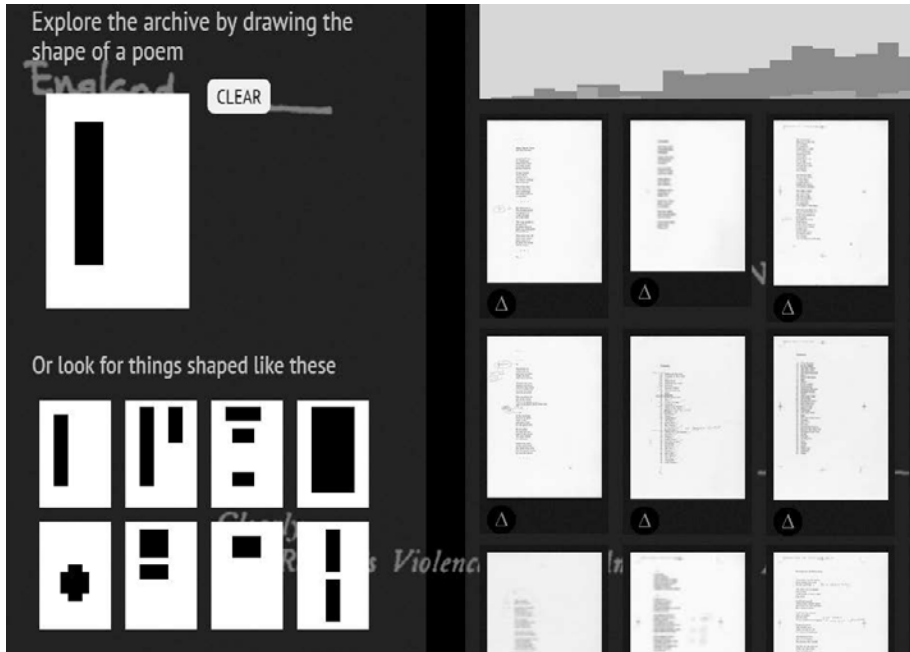


Abb. 1: bloodaxe Zugang über Layoutmuster, <http://bloodaxe.ncl.ac.uk/explore/#/shapes>

skripte sind als Bilder und als automatisch eingeleseene Volltexte zugänglich. Das Archiv kann deshalb die Häufigkeit von Stichwörtern als Suchkriterium verwenden, aber sie auch als Netzwerkgraphik von Kookkurrenzen darstellen. Die Bilder ermöglichen zusätzlich auch den besonders ungewohnten Zugang über die Verteilung des Textes auf der Seite: Texte mit schmalen Kolumnen grenzen sich von die ganze Seite füllenden Texten oder zentrierten Anordnungen ab (Abb. 1)<sup>10</sup>.

Derartige Projekte sind inspiriert vom archivierten Material, verwenden aber nicht die Standardverfahren archivischer Erschließung, um den Benutzer zu seinem Ziel zu führen. Dabei werden archivische Regelwerke und Erschließungsnormen sehr wohl angewendet: Das John Latham Archive verweist in seinem „vernünftigen“ Zugang unter dem Namen „Ivan“ explizit darauf, ISAD(G)-konforme Erschließungsdaten zu verwenden<sup>11</sup>. Derartige Projekte sind nicht nur von der kreativen Leistung der Künstler inspiriert, deren Archivalien verfügbar gemacht werden, sondern auch von einer Forschungsposition der Digitalen Geisteswissenschaften: Geisteswissenschaftliche Forschung ist nicht eine regelgeleitete, reproduzierbare Tätigkeit, sondern eine diskursive Auseinandersetzung mit den zu untersuchenden Objekten. Dabei ist die Benutzerschnittstelle kulturell geprägt wie die Dokumente selbst<sup>12</sup>. Die Gestaltung dieser Schnittstelle ist also auch eine Aussage für sich. Die

<sup>10</sup> Tom SCHOFIELD et al., Archival Liveness. Designing with Collections Before and During Cataloguing and Digitization. *Digital Humanities Quarterly* 9/3 (2015), <http://www.digitalhumanities.org/dhq/vol/9/3/000227/000227.html> [1. 3. 2018].

<sup>11</sup> IA, in: Archive as Event, <http://www.ligatus.org.uk/aae/ivan> [1. 3. 2018].

<sup>12</sup> Johanna DRUCKER, Performative Materiality and Theoretical Approaches to Interface. *Digital Humanities Quarterly* 7/1 (2013), <http://www.digitalhumanities.org/dhq/vol/7/1/000143/000143.html> [1. 3. 2018].

beiden Webportale weisen darauf hin, dass ihr Material sich nicht nur nach den gleichen Kriterien, wie sie schon der Archivbildner angewendet hatte, ordnen lässt.

Diese experimentellen Zugriffsformen sind deshalb auch den Benutzern nicht gerade geläufig. Sehr viel geläufiger ist ein anderer Gegenentwurf zur Gestaltung des Information Retrieval in den Archiven: Der „Google-Suchschlitz“ ist die in der Benutzerkultur wohl am besten etablierte Form der Interaktion mit Informationsressourcen. Er steht primär für eine einfache Volltextsuche, die meist auch in modernen Archivsystemen angeboten wird, neben einer hierarchisch organisierten Archivtektonik oder dem Angebot von kompletten Findbüchern, die zur Lektüre zumindest der Einleitung, aber auch der Kontexte einzelner Einträge einladen. Geoffrey Yeo argumentiert nun, dass die modernen Informationstechnologien sowohl auf der Seite der Benutzer wie der Bestandsbildner die klassischen archivischen Erschließungsmethoden überflüssig machen werden: Forscher organisieren, so beobachtet er, ihre Informationen als flache Datensammlungen ohne hierarchische Ordnung und greifen ebenso primär auf Volltextsuchen zu wie die Bürokrationen, die mit der Einführung von allgemeiner Bürosoftware ihre Akten auch zunehmend als flache Datenbank von Dokumenten verstehen, aus der man mit einer Volltextsuche am leichtesten relevante Informationen extrahieren kann<sup>13</sup>. Man könnte seine Annahmen über das dominante Verhalten zur Informationsorganisation zum Beispiel mit dem Hinweis auf hierarchisch organisierte Dateisysteme oder Systeme zur Abbildung von Geschäftsprozessen in Frage stellen. Ich möchte im Folgenden aber eine andere Perspektive einnehmen und danach fragen, ob der Google-Suchschlitz wirklich ein Konzept der Informationsrückgewinnung ist, das radikal anders als das archivischer Erschließung ist.

Der Google-Suchschlitz ist nämlich erst durch verschiedene technische Maßnahmen im Hintergrund so erfolgreich geworden. Es ist nicht so sehr die Reduktion auf eine nicht-hierarchische Volltextsuche, sondern es sind die Algorithmen, die relevante Treffer zu den Suchwörtern ermitteln. Sie beruhen darauf, dass aus den Suchwörtern auf Suchintentionen geschlossen wird, die sich in statistisch messbaren Eigenschaften des sogenannten „Suchraums“, das heißt der Gesamtheit der durchsuchten Objekte, wiederfinden lassen. Normalerweise werden Suchraum, Suche und zu durchsuchende Dokumente dabei einfach als eine unstrukturierte Ansammlung von Wörtern (*bag of words*) verstanden. Darin kann dann von der Häufigkeit der Wörter in einem Dokument in Relation zur Häufigkeit der Wörter im gesamten Suchraum auf die Relevanz der Wörter für die Beschreibung des Dokuments geschlossen werden. Erkenntnisse der Computerlinguistik haben diesen Ansatz nennenswert weiterentwickelt, indem sie morphologische und lexikalische Normalisierungen sowohl im Suchraum wie in der Suchanfrage vornehmen. Die Relevanz von Dokumenten für eine Suchanfrage ist aber auch Resultat von menschlichen Entscheidungen. In den modernen Suchmaschinen wird die Verweisstruktur des Internet benutzt, um auf diese Entscheidungen automatisch zurückzuschließen: Die Anzahl von Dokumenten, die auf ein Dokument verweisen, ihr Status im Gesamtnetzwerk (viele oder wenige Verweise) oder ihr Host (etwa Universitätsserver) sind Indizien, nach denen Google die Relevanz von Webseiten für eine Suchanfrage zu beurteilen versucht. Mit der Einführung des Google Knowledge Graph ist die Suche intern noch semantisch erweitert worden. Es wird überprüft, ob die Suchtermini in einer Faktendatenbank zu finden sind, aus der sich alternative Bezeichnungen oder Disambiguierungen ableiten lassen. Diese

---

<sup>13</sup> Geoffrey YEO, Contexts, Original Orders, and Item-Level Orientation. Responding Creatively to Users' Needs and Technological Change. *Journal of Archival Organization* 12 (2015) 170–185.

semantische Übersetzung von Suchen ist ein Forschungsbereich im Feld der künstlichen Intelligenz, wird aber experimentell auch schon in Bibliotheksinformationssystemen angewendet<sup>14</sup>. Ähnlich sind auch komplexere statistische Methoden noch nicht im Alltag der Suchmaschinen angekommen. Verfahren des sogenannten „Topic Modeling“ zum Beispiel mit der sogenannten *Latent Dirichlet Allocation*<sup>15</sup> versuchen, mit statistischen Methoden Wörter in einem Corpus an Hand ihrer Nachbarschaften so zu gruppieren, dass sich sinnvolle Zusammenhänge ergeben. Die hinter einer Volltextsuche stehenden Algorithmen sind also komplexer als die einfache Eingabeform es vermuten lässt. Auf welche Kriterien diese Algorithmen achten, ist aber offen. Es könnten also auch für die Archive relevante Kriterien verwendet werden.

Die vordergründige Annahme einer auf nicht-archivische Suchformen fokussierten Benutzergruppe ist auch an einer anderen Stelle zu hinterfragen: Passen sich die modernen Historikerinnen entweder ganz der archivischen Erschließungsweise an oder hegen sie nur naive Erwartungen aus der nicht-archivischen Welt der Suchmaschinen? Wohl kaum. Peter Haber geht zum Beispiel in seiner umfassenden Analyse des Verhältnisses von Geschichtswissenschaften zu den digitalen Informationstechnologien genau davon aus, dass „Jede Suche [...] geprägt [wird] vom Vorwissen, das der Suchende mitbringt, von der Kenntnis der entsprechenden Zusammenhänge und Fachbegriffe, von der Erfahrung und der Routine im Umgang mit den Hilfsmitteln“<sup>16</sup>. Amerikanische Archivwissenschaftlerinnen haben schon zu Beginn des Jahrtausends diese Position empirisch untermauern können<sup>17</sup>. Yakel und Torres bescheinigen den Historikerinnen, dass sie in der Auseinandersetzung mit der Überlieferung eine „archival intelligence“ entwickeln<sup>18</sup>. Hea Lim Rhee entwirft 2012 ein Modell archivalischer Suche, in dem nach der Suche (*searching stage*) die *accessing stage* (Zugriff) und die *processing stage* (Verarbeitung) neue Suchanfragen anstoßen können, also in eine erneute *searching stage* münden<sup>19</sup>. Es entsteht dabei etwas, das nennenswerte Ähnlichkeiten zum hermeneutischen Zirkel hat. Ich möchte diese iterative oder explorative Methode als „lernende Suche“ bezeichnen und anregen, dass sie in der Diskussion über die Gestaltung von Informationssystemen einen festen Platz bekommt, denn sie ist ein Kern geschichtswissenschaftlicher Forschungspraxis, in der Wissen über das Forschungsobjekt schon während der Suche des Quellenmaterials erzeugt wird.

Eine zweite Perspektive auf die Historiker im Archiv ergibt sich aus geschichtstheoretischen Überlegungen der jüngeren Vergangenheit. In der Debatte um die Rolle der Geschichtsdarstellung für die Geschichtswissenschaft hat sich herausgestellt, dass die Erzäh-

<sup>14</sup> Etwa in der Zusammenarbeit zwischen der Sächsischen Landes- und Universitätsbibliothek Dresden mit der Firma AvantgardeLabs: vgl. Jens MITTELBACH–Robert GLASS, A Library Data Management Platform Based on Linked Open Data. *ResearchGate* (Dezember 2014), <http://doi.org/10.13140/RG.2.1.1178.0965> [1. 3. 2018].

<sup>15</sup> David M. BLEI–Andrew Y. NG–Michael I. JORDAN, Latent Dirichlet Allocation. *Journal of Machine Learning Research* 3 (2002/03) 993–1022.

<sup>16</sup> Peter HABER, Digital Past. Geschichtswissenschaft im digitalen Zeitalter (München 2011) 163.

<sup>17</sup> Wendy M. DUFF–Catherine A. JOHNSON, Accidentally Found on Purpose: Information-Seeking Behavior of Historians in Archives. *The Library Quarterly. Information, Community, Policy* 72 (2002) 472–496.

<sup>18</sup> Elizabeth YAKEL–Deborah A. TORRES, AI: Archival Intelligence and User Expertise. *The American Archivist* 66 (2003) 51–78.

<sup>19</sup> Hea Lim RHEE, Modelling Historians' Information-seeking Behaviour with an Interdisciplinary and Comparative Approach. *Information Research* 17/4 (2012), <http://InformationR.net/ir/17-4/paper544.html> [1. 3. 2018].



lung eine elementare Form der Aneignung von Vergangenheit ist, und zwar unabhängig von der Gesamtstruktur der historischen Darstellungen durch Wissenschaftler<sup>20</sup>. Für die Archive bedeutet das, dass auch bei der Konsultation von Archivmaterial narrative Muster angewendet werden: Der Verlauf einer Verhandlung in der Korrespondenz, der Bericht über eine behördliche Maßnahme in internem Schriftgut, die vor Gericht verhandelten Einzelschicksale, die Reaktion auf einen politischen Vorschlag – die alle auch als Teile von analytischen Studien verwendet werden können. Neben einer analytischen, kategorisierenden Erschließung könnte eine moderne Archiverschließung also auch Narrationen als Suchwege unterstützen.

Die Kernfrage archivischer Erschließung aus der Perspektive der digitalen Geisteswissenschaften ist also nicht die Frage, ob in anderen Lebensbereichen entwickelte Technologien und Formen der Benutzung von Informationsressourcen auch auf die Archivwelt übertragen werden sollten, sondern ob und wie diese spezifische Art der Auseinandersetzung mit dem in den Archiven bewahrten Kulturerbe mit digitalen Mitteln unterstützt werden kann. Archivisches Information Retrieval muss sich also nicht an angebliche Standards der Suchmaschinenteknologie anpassen, sondern Archivinformationssysteme müssen die Technologien auswählen oder neu entwickeln, die am besten zu den Methoden der Geschichtswissenschaft und des Archivwesens passen.

Wie oben skizziert, leitet die etablierte archivische Ordnungslehre und Erschließungspraxis aus dem *respect des fonds*, dem „Provenienzprinzip“, dem „Registaturprinzip“ und der „Archivteknik“ hierarchische Modelle der Erschließung ab. Damit sind fünf theoretische Ansprüche verbunden, die begründen, warum diese Organisation der Informationen die bestmögliche ist:

- Der Archivbildner hat die bestmögliche Informationsstruktur gefunden.
- Die Organisation nach Provenienz (und nach Registatur) bildet die Entstehungszusammenhänge ab.
- Einzelstücke sind ohne ihren Entstehungskontext nicht verständlich.
- Die Hierarchie erleichtert die Suche.
- Eine hierarchische Ordnung erleichtert die Erschließungsarbeit.

Unter diesen fünf Gründen sind zwei, die einer näheren Betrachtung mit den Erfahrungen der Informationserschließung großer digitaler Ressourcen nicht gut standhalten. Clay Shirky hat 2005 den Sieg der Suchmaschinen gegenüber hierarchischen Webverzeichnissen nachvollzogen und verallgemeinert: Er argumentiert sowohl mit der konkreten Entwicklung in der Erschließung von Webressourcen wie grundsätzlich. Er erklärt den Siegeszug der Suchmaschinen, die über Stichwörter und Verweise zwischen den Ressourcen die Informationsbedürfnisse der Benutzer befriedigen, gegenüber hierarchischen Webverzeichnissen wie dem Yahoo/DMOZ-Katalog mit der Mehrdeutigkeit der Objekte: Texte sind nicht physische Objekte, die genau einen Platz im Regal haben müssen, sondern können auf verschiedene Themen verweisen. Unter den Bedingungen von Hypertext ist dann die Hierarchie nur noch ein Verweis unter vielen. Klassifikationen funktionieren nur, wenn es um ein kleines Corpus mit klaren abgrenzbaren Kategorien geht. Über diese

---

<sup>20</sup> Eine Zusammenfassung bietet Wolfgang HARDTWIG, Formen der Geschichtsschreibung. Varianten historischen Erzählens, in: Geschichte. Ein Grundkurs, hg. von Hans-Jürgen GOERTZ (Reinbek 1998) 169–188, hier 174f.; vgl. Achim SAUPE–Felix WIEDEMANN, Narration und Narratologie. Erzähltheorien in der Geschichtswissenschaft, Version: 1.0, in: Docupedia-Zeitgeschichte (28. 1. 2015), [http://docupedia.de/zg/saupe\\_wiedemann\\_narration\\_v1\\_de\\_2015](http://docupedia.de/zg/saupe_wiedemann_narration_v1_de_2015) [1. 3. 2018].

Kategorien müssen diejenigen, die erschließen, ebenso Bescheid wissen wie die Benutzer der Erschließung<sup>21</sup>.

Aus informationspraktischer Sicht ist die hierarchische Organisation also ein Hilfsmittel, das zur Organisation von Papierakten in Behörden mit professionellen Registaturen nützlich war und so auch helfen kann, die Akten im Archiv wiederzufinden. Die hierarchische Ordnung der Akten in Aktenplan und Archivverzeichnung war nützlich – vielleicht sogar essentiell –, um Einzelstücke in der Menge der Papierdokumente zu finden. Für die Benutzer war die Ordnung schon immer fremd und musste erst erlernt werden. Unter den Bedingungen elektronischer Datenverwaltung ist diese Ordnung nicht mehr die gegebene. Geoffrey Yeo hat 2015 darauf hingewiesen, dass der mediale Übergang auch andere – und eventuell bessere – Erschließungswege zur Verfügung stellt. Diese kommen zunehmend zum Einsatz, sowohl auf Seite der Aktenproduzenten in der digitalen Aktenführung als auch auf Seite der wissenschaftlichen Benutzer in ihren Forschungsdatenbanken. Beide verwenden kategoriale Beschreibungen von Einzeldokumenten, als inhaltliche Beschreibung meist in Form von Schlagwörtern. Hinzu kommt, dass die Erschließungskategorien der Archivbildner sich nur selten mit den Erschließungskategorien der Forscher decken, die Einzeldokumente also von den Benutzern in neue Kontexte eingeordnet werden. Er schließt daraus, dass zwei der oben angeführten Gründe für eine hierarchische Erschließung wegfallen: Der Archivbildner hat weder notwendig die bestmögliche Informationsstruktur gefunden, noch erleichtert eine hierarchische Klassifikation notwendig die Suche<sup>22</sup>. Damit reduzieren sich die Gründe für den Erhalt einer hierarchischen und an der Organisation der Dokumente wie beim Archivbildner orientierten Erschließung auf die Abbildung von Entstehungszusammenhängen, auf die Kontextualisierung der Akten und auf die Arbeitspragmatik. Diese Gründe sind gut genug, die spezifischen Eigenschaften einer solchen Erschließung in der digitalen Umgebung fruchtbar zu machen.

Die Verfahren der Suchmaschinen zur Gewichtung der Treffer und zur Suchexpansion sind zunächst einmal nicht auf Archive übertragbar. Das liegt einerseits daran, dass Archive gewöhnlich keine umfangreichen Texte, sondern nur kurze beschreibende Metadaten den Maschinen zur statistischen Bewertung bereitstellen können. Auch eine Gewichtung auf Grund von Verweisen erscheint durch die geringe Anzahl von expliziten Verweisen in den Beschreibungen vernachlässigbar. Bei einer näheren Betrachtung wird jedoch deutlich, dass zumindest eine morphologische Anreicherung und eine Gewichtung auch für Suchmaschinen in Archivbeschreibungen sinnvoll sein können. Thomas Grotum hat so 2015 in seiner Stellungnahme zur Veröffentlichung des Archivportal-D die Zusammenarbeit zwischen Archivwissenschaft und Digital Humanities eingefordert: „Bei der Entwicklung entsprechender Werkzeuge im Rahmen der Digital Humanities könnten Archive und Forschung kooperieren“<sup>23</sup>. Aus Sicht der Digital Humanities scheint der Einsatz von Lemmatisierungswerkzeugen in den Beschreibungsdaten eine

<sup>21</sup> Clay SHIRKY, *Ontology is Overrated: Categories, Links, and Tags*, in: Clay Shirky's Writings About the Internet (2005), [http://shirky.com/writings/ontology\\_overnated.html](http://shirky.com/writings/ontology_overnated.html) [1. 3. 2018].

<sup>22</sup> YEO, *Contexts* (wie Anm. 13).

<sup>23</sup> Thomas GROTUM, *Schritt für Schritt ins Paradies? Anforderungen an Archivportale aus der Sicht eines Historikers*, in: *Netz werken. Das Archivportal-D und andere Portale als Chance für Archive und Nutzung. Beiträge zum 19. Archivwissenschaftlichen Kolloquium der Archivschule Marburg*, hg. von Irmgard Christa BECKER–Gerald MAIER–Karsten UHDE–Christina WOLF (Veröffentlichungen der Archivschule Marburg, Hochschule für Archivwissenschaft 61, Marburg 2015) 251–268, hier 268.



Erweiterung der archivischen Erschließungssysteme, die dem gewöhnlich in lexikalischen Grundformen suchenden Benutzer die Suche nennenswert erleichtert. Dabei kann auf etablierte Forschungsergebnisse im Bereich der Computerlinguistik zurückgegriffen werden. Weniger etabliert sind Methoden, den Suchraum zusätzlich noch um Referenzen auf Lexika zu erweitern, die etwa Über- und Unterbegriffe verzeichnen können, wie zum Beispiel in den WordNets<sup>24</sup>, oder Wissen aus der historischen Semantik einfließen zu lassen<sup>25</sup>.

Um das Prinzip der Treffergewichtung sinnvoll anzuwenden, müssen andere Kriterien als die von Google verwendeten Links angewendet werden. Mit der Archivteknik bietet die klassische Archiverschließung dafür einen guten Kandidaten. Die verschiedenen Hierarchieebenen, insbesondere aber der Provenienzzusammenhang, sind nämlich Formen vereinfachter Informationserfassung: Die Erschließungsangaben eines Bestandes sind für die meisten darin enthaltenen archivalischen Einheiten gültig. Ein Treffer in einer Bestandsbeschreibung repräsentiert also eine Vielzahl an Treffern in Aktenverzeichnungen – insbesondere dann, wenn diese Akten selbst noch gar nicht verzeichnet sind. Auch für die Sucherweiterung kann eine hierarchische Ordnung von Archivgut nutzbar gemacht werden, wenn nämlich auch die Beschreibungen der Teile eines Bestandes als Deskriptoren des Gesamtbestandes verstanden werden. Wenn diese Einsichten über die Funktion der Archivteknik in die Gewichtung von Suchtreffern integriert werden, dann sollte eine Suche nach „Enteignung Jüdischer Besitz“ im Archivportal-D den Bestand B 323 im Bundesarchiv („Treuhandverwaltung von Kulturgut bei der Oberfinanzdirektion München“) vor dem Schriftverkehr mit dem Reichsleiter Martin Bormann (B 323/102) aus dem Bestand sortieren. Eine solche archivisch geprägte Methode des Information Retrieval wendet also archivfachliches Wissen auf die vorhandenen technischen Konzepte an: aus dem „Google-Suchschlitz“ wird ein „Archivsuchschlitz“.

Das Archivportal-D hat zum Beispiel in seinem Suchinterface einen weiteren Schluss aus den spezifisch archivischen Bedingungen der Erschließung gezogen. Die Benutzer suchen zwar per Stichwortsuche in der Gesamtheit der Erschließungstexte und bekommen einzelne Verzeichnungseinheiten, egal welcher Erschließungsebene, als Treffer geliefert. Sie können die Trefferliste aber nach weiteren archivischen Erschließungskategorien einschränken. In dieser facettierten Suche sind die Provenienz, die Verzeichnungsstufe und der Archivalientyp im zweiten Suchschritt ebenso zusätzlich anwendbare Kategorien wie Schlagwörter, Personen, Orte oder Zeiträume. Diese Lösung nutzt die Funktion des Provenienzprinzips und den Erhalt der originalen hierarchischen Gliederung als Abbild der Entstehungsumstände und als Erläuterung des Kontextes von Einzelstücken.

Die Erschließungstiefe dem Benutzer als Suchkriterium anzubieten und sie als Kriterium für die Gewichtung der Treffer zu verwenden, kann in der Erschließungspraxis dazu motivieren, ausführliche Bestandsbeschreibungen langwierigen Einzelaktverzeichnungen vorzuziehen. Die Erleichterung der Erschließungsarbeit durch eine hierarchische Organisation wird also durch die auf Mathematik aufbauende Möglichkeit verstärkt, summarische Beschreibungen dem Benutzer mit dem gleichen Gewicht anzubieten wie

---

<sup>24</sup> Das älteste dieser hierarchischen Lexika ist das englische Wordnet, erstellt und gepflegt an der Princeton University: Wordnet. A Lexical Database for English, <https://wordnet.princeton.edu/> [1. 3. 2018]. Für die deutsche Sprache wird eine entsprechende Ressource an der Universität Tübingen erstellt: GermaNet. A German WordNet, <http://www.sfs.uni-tuebingen.de/GermaNet/> [1. 3. 2018].

<sup>25</sup> Zum Problem der Formalisierung von Wissen alleine auf Grundlage von modernem Sprachgebrauch vgl. Kim H. VELTMAN, Towards a Semantic Web for Culture. *Journal of Digital Information* 4/4 (2004), <http://journals.tdl.org/jodi/article/view/113> [1. 3. 2018].

eine Vielzahl von Einzelverzeichnissen in einem Bestand. Es ergibt sich daraus jedoch die Forschungsaufgabe, realistisch einzuschätzen, welches Gewicht eine Bestandsbeschreibung erhalten soll. Dazu wären die Deskriptorenräume ausführlicher Bestandsbeschreibungen mit der Gesamtheit der Deskriptoren für die enthaltenen kleineren Verzeichnungseinheiten zu vergleichen und gegebenenfalls Schlüsse für die Praxis der Bestandsbeschreibung zu ziehen.

Provenienz- und Registraturprinzip sind auch in der Archivfachwelt nicht unumstritten, weil sie nur den Zustand der Aktenorganisation bei der Übernahme im Archiv abbilden. Die Entwicklung der Aktenorganisation als Teil der Behördengeschichte bleibt dabei unsichtbar<sup>26</sup>. Es ist aber akzeptiert, dass Provenienz- und Registraturprinzip ebenso wie andere Erschließungsmethoden einen zum Verständnis der Akten notwendigen Kontext liefern<sup>27</sup>. Aus der oben angeführten erkenntnistheoretischen Position heraus geht es also darum, diesen Kontext – wie alle von den Archivaren ermittelten Kontexte – zum Teil der Narrationen werden zu lassen, in denen die geschichtswissenschaftlichen Benutzer die Archivalien rezipieren. Damit meine ich nun nicht die eher theoretische Position, dass auch die Verfahren der Archivare prinzipiell auch Meistererzählungen realisieren<sup>28</sup>. Es geht vielmehr um die konkrete Rezeption der Archivalien als Zeugnisse von menschlichen Handlungen, die in eine sinnvolle Abfolge gebracht werden, um sie zu verstehen. Die Benutzer sollten sich aus der archivischen Erschließung Narrative konstruieren können, die ihnen helfen, zu den Einzelakten zu gelangen, und die ihnen helfen, sich eigene Narrationen aufzubauen, um zu historischen Erkenntnissen zu kommen. Können die – offensichtlich narrativen – Behörden- und Bestandsgeschichten gemeinsam mit kategorialen Beschreibungen wie Aktentitel, Laufzeit oder Indizes mit Hilfe von Informationstechnologien so umorganisiert werden, dass sie diese Prozesse unterstützen?

Die Digital Humanities haben dazu kein Patentrezept anzubieten. Sie engagieren sich zwar im Bereich der Narratologie, sind aber derzeit noch vorwiegend damit beschäftigt, den Befund aufzunehmen: Digitale Techniken werden vorwiegend dazu eingesetzt, Er-

<sup>26</sup> Insbesondere Peter Scott und Mario Fenyo argumentierten dafür, auf der den Fonds untergeordneten Ebene der Serien zu erschließen: Peter SCOTT, The Record Group Concept: A Case for Abandonment. *The American Archivist* 29 (1966) 493–504; Mario D. FENYO, The Record Group Concept: A Critique. *The American Archivist* 29 (1966) 229–239. Aus den systematischen Überlegungen hinter dieser Kritik leiten sich etwa Positionen zur funktionalen Beschreibung von Archivgut ab: Peter HORSMAN, Wrapping Records in Narratives. Representing Context through Archival Description, in: ITS. International Tracking Service – News – 12.10.2011 Internationale Archivtagung beim ITS, [https://www.its-arolsen.org/fileadmin/user\\_upload/Metanavigation/Aktuelles/2011\\_PDF/04\\_Horsman\\_text.pdf](https://www.its-arolsen.org/fileadmin/user_upload/Metanavigation/Aktuelles/2011_PDF/04_Horsman_text.pdf) [1. 3. 2018]. Zur Geschichte der Archivtheorie: Terence Gordon COOK, What is Past is Prologue: A History of Archival Ideas since 1898, and the Future Paradigm Shift. *Archivaria. The Journal of the Association of Canadian Archivists* 43 (1997) 17–63. Weiters vgl. Jefferson BAILEY, Disrespect des Fonds: Rethinking Arrangement and Description in Born-Digital Archives. *Archive Journal* 3 (2013), <http://www.archivejournal.net/essays/disrespect-des-fonds-rethinking-arrangement-and-description-in-born-digital-archives/> [1. 3. 2018].

<sup>27</sup> Heather MACNEIL, The Context is All: Describing a Fonds and its Parts in Accordance with the Rules for Archival Description, in: *The Archival Fonds: From Theory to Practice*, hg. von Terence M. EASTWOOD (Ottawa 1992) 195–225.

<sup>28</sup> Wendy M. DUFF–Verne HARRIS, Stories and Names: Archival Description as Narrating Records and Constructing Meanings. *Archival Science. International Journal on Recorded Information* 2 (2002) 263–285; oder weniger allgemein: Heather MCNEIL, Picking Our Text. Archival Description, Authenticity, and the Archivist as Editor. *The American Archivist* 68 (2005) 264–278. Archivare haben aber auch schon praktische Konsequenzen aus der Erkenntnis zu ziehen versucht: Melanie FEINBERG, Designing Collections for Storytelling: Purpose, Pathos, and Poetry. *Information Research* 15/3 (2010), <http://www.informationr.net/ir/15-3/colis7/colis701.html> [1. 3. 2018].

zählstrukturen in literarischen Texten aufzufinden, zu dokumentieren und zu annotieren. Forschungen zur aktiven Erzeugung von Narrativen (*story telling*) finden mit Studien zur künstlichen Intelligenz statt. Sie sind aber noch nicht sehr weit fortgeschritten<sup>29</sup>.

Dennoch können etablierte archivische Erschließungsmethoden einen narrativen Zugang zu den Archivalien unterstützen. Erstens müssen sie ernst nehmen, dass die Organisation der Archivalien im Bestand Kontext für ihre Bestandteile liefert. Der Informationsgewinn aus der Organisation von Einzelakten in Bestände liegt für die Archivbenutzer darin, dass sie sich damit über Tätigkeiten der historischen Akteure informieren können: Die Behörden agieren in ihrem Schriftgut, sie kommunizieren, entscheiden, ordnen an, verbieten. Wo sie nicht direkt durch das Schriftgut agieren, dokumentieren sie doch ihr Handeln, das Handeln der in ihnen beschäftigten Personen oder das Handeln derjenigen, die mit ihnen kommunizieren. Die klassische hierarchische Erschließung auf Bestandsebene nach ISAD(G) berücksichtigt das mit den Angaben zur Geschichte des Bestandsbildners und zum Inhalt des Bestandes – historische Erzählungen, die bei Lektüre beim Archivbenutzer ein Verständnis von den im Bestand zu erwartenden Informationen erzeugen. Das ICA schlägt für diese Bedürfnisse zwei weitere Standards vor: ISAAR(CPF) für die Beschreibung der Bestandsbildner und ISDF für die Aufgaben und Tätigkeiten der Akten erzeugenden Instanzen<sup>30</sup>. ISAAR(CPF) hat im XML-Schema EAC-CPF konkretisiert einigermaßen Verbreitung in der Archivwelt gefunden. ISDF ist bislang nur von akademischem Interesse, vermutlich auch weil es an konkreten technischen Implementationen fehlt, mit denen diese Informationen in die digitalen Abläufe alltäglicher Archivarbeit integriert werden können.

Die Datenmodellierung, das heißt die Entwicklung von modellhaften Konzepten und ihre Abbildung in konkrete technische Lösungen, ist nun ein zentrales Forschungsfeld der Digital Humanities<sup>31</sup>. Die archivische Erschließung hat mit den Kategorien von ISAD(G) ein relativ flaches Modell entwickelt. Die verbale Beschreibung der Bestände – beziehungsweise der Einzelakten – enthält viele Informationen, die anders organisiert werden könnten, um sie für weitere Suchanfragen zugänglich zu machen. Sie werden für einen „erzählenden“ Zugang zum Archivmaterial besonders attraktiv, wenn sie Verbindungen zu Beschreibungen in anderen Erschließungseinheiten identifizieren. Dafür steht mit Schlagwortindizes ein etabliertes Mittel zur Verfügung: Die klassischen Angaben zu Personen und Orten helfen, ihr Vorkommen auch bestandsübergreifend leicht zusammenzuordnen. Technische Potentiale ergeben sich daraus, mit modernen Mitteln der Wissensorganisation zu verschlagworten.

Die Erschließungsarbeit, die zurzeit bei den Staatlichen Archiven Bayerns an den

---

<sup>29</sup> Pablo GERVÁS, *Story Generator Algorithms*, in: *Handbook of Narratology*, hg. von Peter HÜHN–Jan Christoph MEISTER–John PIER–Wolf SCHMID, 2 Bde. (Berlin–Boston 2014) 2 825–835; David HERMAN, *Cognitive Narratology*, in: ebd. 1 46–64.

<sup>30</sup> ISDF: *International Standard for Describing Functions*, in: ICA. *International Council on Archives – International Standards*, <http://www.ica.org/en/isdf-international-standard-describing-functions> [1. 3. 2018]; ISAAR(CPF): *International Standard Archival Authority Record for Corporate Bodies, Persons and Families*, 2nd Edition, in: ICA. *International Council on Archives – International Standards*, <https://www.ica.org/en/isaar-cpf-international-standard-archival-authority-record-corporate-bodies-persons-and-families-2nd> [1. 3. 2018].

<sup>31</sup> Willard McCARTY, *Humanities Computing* (Basingstoke–New York 2005) 29–72; DERS., *Modeling: A Study in Words and Meanings*, in: *A Companion to Digital Humanities*, hg. von Susan SCHREIBMAN–Ray SIEMENS–John UNSWORTH (Blackwell Companions to Literature and Culture 26, Malden–Oxford–Carlton 2004) 254–270.

Wiedergutmachungsakten der Nachkriegszeit geleistet wird, ist ein gutes Beispiel dafür, worauf es hierbei zu achten gälte<sup>32</sup>. Die für die Registratur der Akten angelegten Karteikarten sind eine gute Basis für die Erschließung. Sie enthalten Informationen wie die Namen der Verfolgten und der Anspruchsberechtigten, die jetzigen Eigentümer und ihre Vorgänger. Diese Informationen sind zentral für die Klarstellung von Restitutionsansprüchen an enteignete Kunst. In der summarischen Beschreibung eines Aktenitels und einfacher Personenschlagwörter gehen diese strukturellen Informationen aber nur in einen nicht weiter untergliederten Aktenbetreff ein. Die bayerischen Archivarinnen haben in den Erfassungsrichtlinien für die Rollen der an den Verwaltungsverfahren beteiligten Person ein kontrolliertes Vokabular („Antragsteller“, „Antragsgegner“, „Verfolgter“, „Berechtigter“) entwickelt, das der externe Dienstleister bei der Digitalisierung der Karteikarten zu verwenden hat. Diese pragmatische Lösung tilgt also nicht etwa textuelle Information, sondern ordnet sie nur anders an, reduziert die Struktur der Vorlage. Das scheint die Folge eines technischen Systems zu sein, nämlich der verwendeten Eingabemaske für die Erschließungsdaten. Die benutzte Software nutzt aber nicht die an sich vorhandenen technischen Methoden, denn die digitalen Hilfsmittel zur Modellierung erlauben es, über die einfache Kategorisierung als „Person“, „Ort“ oder „Sachschlagwort“ hinauszugehen<sup>33</sup>. Mit den Methoden der Klassenbeziehungen können nämlich Verallgemeinerungen in Wissensrepräsentationen so ausgedrückt werden, dass die Erschließungsinformationen sowohl in der allgemeinen Kategorie „Person“ wie in einer speziellen Kategorie als „Verfolgter“ auffindbar sind. In der Museumswelt ist zum Beispiel mit dem Standard des CIDOC-CRM ein Modell entwickelt worden, das solche Verallgemeinerungen bis auf die abstrakte Unterscheidung von temporalen und nicht-temporalen Konzepten mitdenkt<sup>34</sup>.

Das CIDOC-CRM setzt auf einem Basismodell auf, das sich in den digitalen Geisteswissenschaften zunehmender Beliebtheit erfreut: der Repräsentation von strukturierten Daten als Graphen, das heißt als Beziehungen zwischen zwei Knoten. Dieses Modell lässt sich auf eine sehr einfache intellektuelle Struktur abbilden, nämlich einfache Aussagesätze aus „Subjekt“ (Knoten), „Prädikat“ (Beziehung) und „Objekt“ (Knoten). Dieses Modell ist zum Beispiel Grundlage der Aktivitäten des W3C im Bereich des sogenannten „Semantic Web“, in dem strukturierte Daten über Web verknüpfbar und austauschbar gemacht werden<sup>35</sup>. In der Archivwelt ist dieser Ansatz mit der Arbeit an „Records in Contexts“ (RiC) in Angriff genommen worden. RiC ist ein konzeptionelles Modell, das die Dokumentation und ihre Organisation sowie die Akteure, ihre Rollen, Aufgaben und Aktivitäten zueinander in Beziehung setzt. Es kann damit sowohl die Aktivitäten bei der Entstehung der Akten, ihre Organisation in der Registratur und im Archiv abbilden als

<sup>32</sup> Ich danke Herrn Dr. Julian Holzapfl für den Einblick in die Erschließungspraxis und die fruchtbare Diskussion über die dahinter stehenden Konzepte.

<sup>33</sup> EAD 2002 sieht im Bereich für die Beschreibung mit kontrollierten Vokabularien neben den Kategorien *persname*, *geogname* und *subject* auch noch Organisations- und Familiennamen (*corpname*, *famname*), Berufe und Funktion (*occupation*, *function*) sowie archivalische Genres (*genreform*) vor: Encoded Archival Description Tag Library, Version 2002, in: The Library of Congress – Standards – EAD, <http://loc.gov/ead/tglib/elements/controlaccess.html> [1. 3. 2018].

<sup>34</sup> CIDOC-CRM Conceptual Reference Model, <http://cidoc-crm.org/> [1. 3. 2018]. Das offizielle Dokument zur Definition des Standards ist: Definition of the CIDOC Conceptual Reference Model, Version 6.2.2, September 2017, hg. von Patrick LE BOEUF et al., in: CIDOC-CRM Conceptual Reference Model – Resources, [http://cidoc-crm.org/sites/default/files/2017-09-30%23CIDOC%20CRM\\_v6.2.2\\_esIP.pdf](http://cidoc-crm.org/sites/default/files/2017-09-30%23CIDOC%20CRM_v6.2.2_esIP.pdf) [1. 3. 2018].

<sup>35</sup> Semantic Web, in: W3C. World Wide Web Consortium – Standards, <https://www.w3.org/standards/semanticweb> [1. 3. 2018].

auch die Dokumentation zu bestimmten Personen, Behörden und Aktivitäten durchsuchbar machen. Noch sind die vorgeschlagenen fast 800 Beziehungen zwischen den Objekten wenig entwickelt. Sie zeigen aber deutlich, welches Potential sie für eine gänzlich anders konzipierte Suche bieten: RiC-R204 definiert zum Beispiel „Agent authored Record“ oder RiC-R511 „Activity resulted in Record“<sup>36</sup>.

Aus einer solchen Konzeption der archivischen Beschreibung ergeben sich Zugangsformen, welche das Wissen über die Archivalien, über ihren Inhalt und über ihren Entstehungskontext den assoziativen und narrativen Modellen der Forscher näher bringen. In verschiedenen Archiven sind Experimente mit einschlägigen Darstellungsmethoden ihrer Erschließung gemacht worden. Das Staatsarchiv Thurgau hat zum Beispiel seine Beständeübersicht als Netzplan visualisiert<sup>37</sup>. Weiter gehen die Archivabteilung der New York Public Library<sup>38</sup> und ein Portal zu Bologneser Archivalien des 19. und 20. Jahrhunderts unter dem Namen „Città degli archivi“<sup>39</sup>. Beide übersetzen die Deskriptoren (Personen, Organisationen, Orte, Inhalte und Ereignisse) der Archivalien in eine graphische Visualisierung aus Knoten und Beziehungen. Die Benutzer können sich damit zu jedem beliebigen Knoten die einschlägigen Archivalien, aber auch die Beziehungen zu anderen Deskriptoren anzeigen lassen, und sich so durch die Assoziationsketten der Erschließung hangeln. Während das Bologneser Projekt die Beziehungen auch, wie in RiC vorgeschlagen, konkret benennt, sind die Beziehungen zwischen den New Yorker Archivalien nicht klassifiziert. Dafür dienen dort auch die Verzeichnungseinheiten als Knoten, die zu weiteren Deskriptoren führen. Dieser Unterschied ist relevant für die Frage danach, welche Konsequenzen digitale Instrumente für die archivische Erschließung haben: Die amerikanische Darstellung benutzt nämlich auch die Organisation der in den Dokumenten enthaltenen Information durch die Archivbildner und die Archive. Sie zeigt, dass sich die Informationen, die in der Erschließungshierarchie und in den Standards archivischer Erschließung enthalten sind, den Vorgehensweisen des lernenden Suchens annähern können, wenn sie in andere Präsentationsformen gebracht werden.

Am Ende dieses Beitrags kann ich deshalb zu den eher experimentellen Formen der Präsentation von archivischer Erschließung zurückkehren und ein Beispiel anführen, bei dem sich ein Designer Gedanken über die digitale Präsentation von Archiverschließung gemacht hat. Mitchell Whitelaw hat Visualisierungen der National Archives of Australia entworfen, die als Zugangsweg zu den einzelnen Archivalien benutzt werden können. In einer ersten Form visualisiert er die Größe der Bestände je Bestandsbildner. Whitelaw verwendet dafür sowohl die Zahl der Verzeichnungseinheiten als auch die Zahl der Regalmeter und bildet damit eine innere Hierarchie der Erschließung ab, die auf abstrakterer Ebene bleiben kann und damit wenige Verzeichniseinheiten für umfangreiches Archivmaterial erzeugt, oder mehr Details aufnimmt, die sich auf eine geringe Menge von Papier verteilen. Dass eine solche Visualisierung natürlich den unterschiedlichen Lagerungsbedingungen von zum Beispiel Urkunden und Papier Rechnung trägt, ist ein Hinweis auf die Notwendigkeit einer archivfachlich informierten Form der Darstellung. Eine zweite

<sup>36</sup> Records in Contexts – Conceptual Model, in: ICA. International Congress on Archives – International Standards, <http://www.ica.org/en/egad-ric-conceptual-model> [1. 3. 2018].

<sup>37</sup> Urban STÄHEL, Archivnetzplan. Ein neues Findmittel für die Online-Suche in Archivbeständen. *Informationswissenschaft. Theorie, Methode und Praxis* 4/1 (2016) 266–278.

<sup>38</sup> Terms Explorer, in: The New York Public Library. Archives and Manuscripts, <http://archives.nypl.org/terms> [1. 3. 2018].

<sup>39</sup> archIVI. Città degli archivi, <http://www.cittadegliarchivi.it> [1. 3. 2018].

Visualisierung listet die in den Titeln vorkommenden Wörter so auf, dass sie zum einen auf ihre Häufigkeit über die Zeit und zum anderen auf mit ihnen gemeinsam vorkommende Wörter verweisen<sup>40</sup>. Gerade diese zweite Visualisierung nähert sich wieder den narrativen Elementen historischer Forschung an, indem die Texte der Archivverzeichnung aufgebrochen und in neue Zusammenhänge gestellt werden. Die Benutzer bekommen also gewissermaßen alle möglichen Geschichten erzählt, die sich aus den Aktentiteln bilden lassen.

Aus Sicht der Digital Humanities gestaltet sich das Verhältnis von klassischen Formen der Archiverschließung und den Zugängen der Digital Humanities zum Information Retrieval also als Anspruch, nicht die etablierten hierarchischen Formen durch flache Erschließung zu ersetzen, sondern die Gründe und das Potential der etablierten Formen von Erschließung im Information Retrieval der Archivinformationssysteme nutzbar zu machen. Die Anforderung an ein solches muss also nicht nur heißen, einen hierarchischen Zugang etwa über die Archivtekonik zu ermöglichen, sondern diese auch in der Suchumgebung zum Beispiel über Indizes oder die Stichwortsuche zu berücksichtigen. Wenn eine Stichwortsuche eine Gewichtung der Treffer nötig macht, dann ist dabei die Bedeutung der übergeordneten Erschließungsebenen entsprechend zu berücksichtigen. Umgekehrt kann mit einer solchen Berücksichtigung auch die Archivpraxis guten Gewissens die Erschließung von Beständen, Bestandsgruppen, Aktengruppen und anderer Aggregationen von Einzelarchivalien bevorzugen, da Treffer in diesen Beschreibungen hervorgehoben werden. Es ist sogar zu überlegen, ob nicht die Aktentitel innerhalb einer Aggregation im Suchraum durch die Beschreibung übergeordneter Einheiten angereichert werden sollten. Experimente mit der Visualisierung von hierarchischer Erschließung geben weitere Beispiele, wie sie informativ in die Präsentation der Archivinformationen integriert werden kann. Darüber hinaus unterstützt Archiverschließung auch narrative Zugänge, wenn die Bezüge, die sich aus Schlagwörtern, Erschließungshierarchien und verbaler Kontextualisierung zum Beispiel in Form von Behörden- und Bestandsgeschichte ergeben, den Benutzern als Lesepfade durch das Archivgut angeboten werden können. Hier ist noch Forschungsbedarf, der einer Zusammenarbeit zwischen Digital Humanities und Archivwissenschaft eine rosige Zukunft verspricht.

---

<sup>40</sup> Series Browser, in: NAA. National Archives of Australia – About us – Research Grants, <http://www.naa.gov.au/about-us/grants/ian-maclean/series-browser.aspx> [1. 3. 2018].