

Müssen wir anders verzeichnen? Erschließung zwischen analogen Archivgewohnheiten und digitalen Nutzererwartungen¹

von Mario Glauert

Der folgende Beitrag soll der Frage nachgehen, ob Archive im Zeitalter der digitalen Transformation und der geänderten Nutzererwartungen ihre Unterlagen anders verzeichnen müssen als bisher. Man mag dabei zuerst an Google denken, aber vielleicht auch an neue Erschließungsansätze wie Records in Context, an Linked open Data, Social Tagging, Normdaten, möglicherweise auch an Information Retrieval allgemein, an Text- und Data Mining, Taxonomien oder Ontologien.

Archive, so könnte man polemisieren, folgen immer noch ihren analogen Erschließungsgewohnheiten, während ihre Nutzer längst digitale Erwartungen haben und alles schnell und bequem bei Google finden möchten.²

Andererseits, so könnte man entgegenhalten, haben sich unsere Akten kaum verändert. Und ob wir auf Karteikarten oder über ein Archivinformationssystem für ein Online-Portal erschließen, ändert nichts daran, dass wir auch dort Signatur, Titel und Laufzeit eintragen müssen. Unsere Quellen haben sich nicht verändert und auch die Themen, die die Benutzer darin suchen, sind so vielfältig und unvorhersehbar wie zu Zeiten, in denen Anfragen noch per Post und nicht per E-Mail und WhatsApp eintrafen.³

Was sich zweifellos verändert hat, sind die Suchmöglichkeiten, seit unsere Datenbanken überall und jederzeit online verfügbar sind. In den letzten zehn bis fünfzehn Jahren haben die Archive hier sehr viel erreicht: Mit Unterstützung der Deutschen Forschungsgemeinschaft (DFG) eine groß angelegte Retrokonversion analoger Findhilfsmittel, die nun online zur Verfügung stehen über archivische Datenbanken, regionale Archivportale wie Archive NRW oder auch national über das Archivportal D und die Deutsche Digitale Bibliothek. Und viel früher als erwartet⁴ sind wir jetzt

1 Vortrag auf dem 71. Westfälischen Archivtag am 13.03.2019 in Herford.

2 Vgl. dazu schon vor Google Wolfram Werner, Benutzererwartungen und strukturierte Erschließung, in: Angelika Menne-Haritz (Hrsg.), *Archivische Erschließung. Methodische Aspekte einer Fachkompetenz. Beiträge des 3. Archivwissenschaftlichen Kolloquiums der Archivschule Marburg* (Veröffentlichungen der Archivschule Marburg, Hochschule für Archivwissenschaft 30), Marburg 1999, S. 39–47.

3 Dass die schriftlichen Nutzeranfragen an Archive hingegen „auf einen relativ kleinen Satz ontologischer Entitäten formalisiert werden können“ (S. i), zeigt die Dissertation von Steffen Hennicke, *What is the Real Question? An Empirical-Ontological Approach to the Interpretative Analysis of Archival Reference Questions*, Phil. Diss. Berlin 2016, <https://edoc.hu-berlin.de/handle/18452/19215> [Stand: 29.06.2019, gilt ebenfalls für alle nachfolgenden Hinweise auf Internetseiten].

dabei, mit Hilfe der DFG schon in die Digitalisierung des Archivgutes selbst einzusteigen.

Mit dem rasanten digitalen Wandel aller Lebens- und Arbeitsbereiche wachsen indes die Erwartungen und Ansprüche unserer Nutzer noch viel schneller als der Umfang unserer digitalen Angebote. Müssen Archive in Zukunft deswegen *anders* verzeichnen? Wobei *anders* ja eine wunderbar unverbindliche Formulierung ist und offen lässt, ob wir künftig mehr⁵ oder sogar weniger erschließen müssen. Vielleicht auch mit gleichem Einsatz, eben nur *anders*.

Wie erschließen Archive?

Die Antwort auf die Frage, ob wir künftig anders erschließen müssen, erfordert natürlich zunächst eine Klärung, wie Archive bislang verzeichnen oder ihre Aktentitel bilden.

Für die archivische Verzeichnung gibt es in Deutschland keine Norm. Es bestehen allenfalls Quasi-Standards⁶, die sich in der Praxis etabliert haben und die heute als Grundlage für Verzeichnungsrichtlinien in zahlreichen deutschen Archiven dienen. Dazu gehört die 1961 bzw. in zweiter Auflage 1967 von Johannes Papritz veröffentlichte „Archivische Titelaufnahme bei Sachakten“, die später auch in seine vierbändige Archivwissenschaft eingeflossen ist und im Westen vor allem durch den Ausbildungsunterricht an der Archivschule Marburg nahezu kanonisch wurde.⁷ Für die DDR veröffentlichte die Staatliche Archivverwaltung 1964 die „Ordnungs- und Verzeichnungsgrundsätze“ (OVG), die damit bis heute für die Erschließung und Titelaufnahme in den Archiven der „neuen Länder“ eine breit etablierte Basis bildet.⁸ Mit der Zunahme der Datenbanken kam ab den 1990er-Jahren ISAD (G), der „General International Standard Archival Description“, hinzu.

Als die Archive ihre Datenbanken dann verknüpfen wollten, stellten sie fest, wie unterschiedlich diese Felder trotz der ähnlichen Verzeichnungsausbildung überall angelegt und gefüllt worden waren. Daher wurden feste Austauschformate und Schnittstellen erforderlich, wie etwa der XML-Standard Encoded Archival Description (EAD, zuerst 1995) zur formalen Beschreibung der archivischen Findhilfsmittel.

Bis heute wirkt diese Entwicklung, die mit dem Aufbau archiv-, sparten- und länderübergreifender Portale noch einmal beschleunigt wurde, in der Praxis indes weniger normierend als nivellierend.

Abseits archivfachlicher Theorien und archivpragmatischer Richtlinien zeigt sich in der Praxis der Erschließung eine große Bandbreite von Verzeichnungsgewohnheiten.⁹ Es gibt zwar meines Wissens keine Studie, die anhand der bestehenden Online-Portale einmal breiter untersucht hat, wie Archive tatsächlich erschließen oder erschlossen haben. Mein Eindruck ist allerdings, dass die Praxis noch sehr viel bunter und kreativer ist als unsere divergierenden Leitlinien.¹⁰

Neue Erschließungskonzepte

Müssen die Archive also heute und in Zukunft „anders“ erschließen, als es uns Papritz, die OVG oder ISAD (G) bisher empfohlen haben oder wie wir es in der Praxis umsetzen?

Die Bibliotheken stellen ihre Katalogisierung von Veröffentlichungen seit einigen Jahren zunehmend von RAK (Regeln für die alphabetische Katalogisierung) auf RDA (Resource Description and Access) um. Nicht nur für den besseren internationalen Austausch bibliographischer Daten. RDA liegt auch der neue Ansatz eines Repräsentationsmodells zu Grunde, der die verschiedenen Ausprägungen, Formen, Formate und Exemplare eines Werkes, egal ob analog oder digital, nicht mehr als völlig voneinander getrennte bibliographische Erfassungseinheiten betrachtet.¹¹

Die deutschen Archive haben RDA, das grundsätzlich ja auch für andere Kultureinrichtungen gedacht war, allerdings kategorisch abgelehnt, da der Werkbegriff, wie er RDA zu Grunde liegt, auf den größten Teil des öffentlichen Archivgutes nicht anwendbar ist.¹²

Der Internationale Archivrat (ICA) hat mit dem Modell „Records in Context“ 2016 einen eigenen neuen Erschließungsansatz vorgelegt.¹³ Ob es gelingt, die komplexen semantischen Verknüpfungen des Referenz-Modells in die ar-

4 Vgl. im Rückblick etwa den Artikel von Peter Müller, Schnell zum Ziel. Erschließungspraxis und Benutzererwartungen im Internetzeitalter, in: Frank M. Bischoff (Hrsg.), Benutzerfreundlich – rationell – standardisiert. Aktuelle Anforderungen an archivische Erschließung und Findmittel. Beiträge zum 11. Archivwissenschaftlichen Kolloquium der Archivschule Marburg (Veröffentlichungen der Archivschule Marburg Institut für Archivwissenschaft 46), Marburg 2007, S. 37–63.

5 Vgl. Franziska Mücke/Susanne Brockfeld, Öffnung der Erschließung, in: Verband deutscher Archivarinnen und Archivare e. V. (Hrsg.), Kompetenz! – Archive in der Wissensgesellschaft. 86. Deutscher Archivtag 2016 in Koblenz, Fulda 2018, S. 209–219, hier: S. 211: „An die erschließenden Archive werden damit zunehmend höhere Anforderungen gestellt: Im Hinblick auf ein primär volltextrecherchierendes Publikum steigen nicht nur die technischen Anforderungen an Archivdatenbanken, sondern steigt auch die Bedeutung einer adäquaten Titelaufnahme und einer guten Aufbereitung der Erschließung durch Datenstrukturen, die dem archivfachlichen Laien unmittelbar verständlich sind.“

6 Vgl. ebd., S. 215, Anm. 26: „eine eher gefühlte als real vorhandene Norm“.

7 Vgl. zum Folgenden den konzisen Artikel von Angelika Menne-Haritz, Erschließung, in: Marcel Lepper/Ulrich Raulff (Hrsg.), Handbuch Archiv. Geschichte, Aufgaben, Perspektiven, Stuttgart 2016, S. 207–217.

8 Vgl. Ilka Hebig [Stahlberg], Zur Entstehungsgeschichte der Ordnungs- und Verzeichnungsgrundsätze der DDR (OVG), in: Angelika Menne-Haritz (Hrsg.), Archivische Erschließung (wie Anm. 2), S. 181–196. Petra Rauschenbach, Einheitliche Ordnungs- und Verzeichnungsgrundsätze in der DDR. Nutzen für Theorie und Praxis, in: Brandenburgische Archive 32 (2015), S. 17–21, http://blha.brandenburg.de/wp-content/uploads/2017/06/Brandenburgische_Archive_32_2015.pdf.

9 Vgl. dazu auch Karsten Uhde, Zwischen Tradition und Online-Mainstream – Archivische Erschließung im 21. Jahrhundert, in: Archivpflege in Westfalen-Lippe 90 (2019), S. 6–10, https://www.lwl-archivamt.de/media/filer_public/ef/e2/efe2639d-074c-42b9-a28a-cf98a6c42080/heft_90_2019.pdf.

10 Vgl. etwa die Anfang 2019 für den 23. Sächsischen Archivtag von Thekla Kluttig unter sächsischen Archiven durchgeführte Umfrage zu den verwendeten Erschließungsstandards und -methoden, <https://saechsischer-archivtag.vda-blog.de/2019/04/17/umfrage-zur-erschliessung-erschliesungsstandards-und-methoden/>.

11 Vgl. Heidrun Wiesenmüller/Silke Horny: Basiswissen RDA. Eine Einführung für deutschsprachige Anwender. Berlin/Boston 2017, S. 3–16.

12 Vgl. Irmgard Christa Becker, Archivische Erschließung und RDA. Stellungnahme der Archivschule Marburg zur Nutzung der RDA (Resource Description and Access) in Archiven, 2013. Text unter: <https://wiki.dnb.de/download/attachments/92447791/rdaKultur2013Becker.pdf>.

13 International Council on Archives Experts Group on Archival Description, Records in Contexts – Conceptual Model for Archival Description, Consultation Draft v0.1, September 2016: <https://www.ica.org/sites/default/files/RIC-CM-0.1.pdf>.

chivische Alltagspraxis zu überführen, erscheint durchaus fraglich. Erste Versuche gibt es allerdings.¹⁴

Wozu und für wen erschließen Archive?

Müssen also auch Archive anders verzeichnen? Vielleicht beginnen wir die Frage mit dem Ziel der Erschließung: Wozu verzeichnen Archive ihre Bestände und Sammlungen? Die Antwort ist einfach: Um das Archivgut findbar, zugänglich und damit nutzbar zu machen.

Diese Feststellung mag banal sein, aber sie erscheint mir dennoch wichtig: Bei der ganzen Diskussion, wie man im Zeitalter der Digitalen Transformation erschließen soll, bleibt es das Wichtigste, überhaupt zu erschließen. Und selbst eine einfache, flache oder archivfachlich unausgereifte Verzeichnung ist besser als gar keine.

Wenn es das Ziel ist, Archivgut findbar zu machen, bliebe die Frage: für wen? Wer soll eigentlich etwas finden? Auch diese Frage erscheint banal, die Benutzer natürlich, aber so einfach und selbstverständlich ist das leider nicht.

Unsere beiden immer noch aktuellen und grundlegenden Regelwerke, Papritz und die OVG, wurden nur bedingt mit Blick auf die Benutzer formuliert. Denn in den 1960er-Jahren hatten die Nutzer noch nicht einmal freien Zugang zu den Findhilfsmitteln, in der DDR noch weniger als im Westen. Und so wie man damals zum Telefonieren in eine Telefonzelle ging, so ging man zum Recherchieren in ein Findbuchzimmer.

Ein Findbuch wurde vorgelegt, im Rahmen der Beratung erläutert und seine Verwendung erklärt. Sie waren nicht offen zugänglich. Und kein Archivar wäre damals auf die absurde Idee gekommen, dass alle Menschen dieser Welt, auch historisch und archivisch gänzlich unerfahrene, völlig frei und ohne Anleitung mit einem Suchschlitz nach diesen Aktentiteln suchen können sollen.

Zudem ist die Frage, für welche Benutzer wir erschließen oder welche unserer Zielgruppen wir bei der Verzeichnung im Blick haben, streng archivfachlich natürlich falsch gestellt. Denn anders als unsere (anderen) Angebote der Öffentlichkeitsarbeit hat archivische Erschließung eben keine Zielgruppe im Fokus, sondern erfolgt einzig mit Blick auf die Herkunft und Entstehung der Unterlagen, auf ihre Provenienz, die Zuständigkeiten und Kompetenzen der Provenienzstellen, Aktenbildner, Verwaltungen.

Unser archivischer Januskopf schaut hier nur in eine Richtung und sein Blick richtet sich bei der Erschließung eben nicht auf die Benutzer¹⁵ – im Gegenteil: Völlig unabhängig von möglichen Fragestellungen oder Forschungsinteressen soll sich die Erschließung einzig auf den Entstehungszweck der Unterlagen richten. Es geht bei der klassischen archivischen Titelbildung eben nicht darum, Antworten zu antizipieren. Ein archivischer Aktentitel soll gerade nicht alle denkbaren Suchstichworte enthalten oder aufzählen, sondern den Entstehungszweck der Akte aufzeigen. Ein fachlich gut formulierter Aktentitel kann daher extrem kurz sein – und eben dadurch alle in der Akte enthaltenen Themen umfassen.

Angelika Menne-Haritz hat dies einmal den „Kern des archivischen Paradoxons“ genannt: Die Erschließung „führt präzise und eindeutig zu den Aussagen der Unterlagen, ohne diese Aussagen selbst zu benennen oder sich auch nur um sie zu kümmern.“¹⁶ Es geht bei der Titelbildung nicht darum, Antworten zu antizipieren, sondern „Aussagepotentiale“ darzustellen.¹⁷

Erschließung für Suchmaschinen

Eigentlich sollten wir die Benutzung und deren mögliche Fragen bei unserer Verzeichnung von Archivgut also gar nicht im Hinterkopf haben. Zudem erschließen wir heute faktisch auch nicht mehr für Benutzer, die vor Ort in den Archiven in Karteien und Findbüchern blättern, sondern für Datenbankabfragen und mathematische Recherchealgorithmen. Wir erschließen für Benutzer, die mit Hilfe von Suchfunktionen oder Suchmaschinen recherchieren. Wenn wir also eine hohe Nutzung haben wollen, müssten wir unsere Erschließungsrichtlinien dann nicht auf diese Suchalgorithmen orientieren? „Suchmaschinenoptimiert“ nennt man das, und jede kommerzielle Webseite lebt heute davon, bei Google möglichst schnell und prominent gefunden zu werden.

Auch für Algorithmen und Suchmaschinen haben Papritz und die OVG vor über fünfzig Jahren ihre Richtlinien natürlich nicht formuliert. Und auch in den jüngeren Erschließungsrichtlinien und -standards gibt es erstaunlicherweise kaum einmal eine einleitende Überlegung, für wen oder was die Verzeichnung eigentlich ein „Finden“ ermöglichen soll.

Müssen wir also anders erschließen, wenn unsere Aktentitel heute für Suchmaschinen optimiert sein sollen und nicht mehr für das analoge Recherchieren unterwiesener Benutzer im Findbuchzimmer?

Ich denke, es lohnt sich, das Problem einmal informationswissenschaftlich zu betrachten. Worin liegt eigentlich die Schwierigkeit?

Kleine Theorie der Klassifikation

Unsere „klassische“ archivische Erschließung besteht bekanntermaßen immer aus zwei miteinander verschränkten Tätigkeiten: aus Ordnung und Verzeichnung. Die hierarchisch angelegte Ordnung orientiert sich an den Kompetenzen und Zuständigkeiten der Provenienzstelle und bildet die Struktur des Bestandes ab. Die Klassifikation führt über mehrere Stufen hinab zum Aktentitel. Titel und Klassifikationen zählen dabei keine Inhalte auf (auch Enthält-

14 Vgl. Gerhard Müller/Silke Jagodzinski, Die Erschließung des Kontexts. Neue Perspektiven auf ein bewährtes Prinzip, in: Archivpflege in Westfalen-Lippe 90 (2019), S. 10–17, https://www.lwl-archivamt.de/media/filer_public/ef/e2/efe2639d-074c-42b9-a28a-cf98a6c42080/heft_90_2019.pdf.

15 Vgl. Zur Debatte um diese „retrospektive“ Erschließung die nachdenkenswerteste Übersicht von Geoffrey Yeo, Continuing Debates about Description, in: Heather MacNeil/Terry Eastwood (Hrsg.), Currents of Archival Thinking, Santa Barbara 2016, S. 163–192.

16 Angelika Menne-Haritz, Erschließung (wie Anm. 7), S. 207.
17 Ebd.

Vermerke sind dafür eigentlich nicht vorgesehen), sondern beschreiben Zuständigkeiten und Entstehungszwecke. Ziel ist dabei die Vermeidung von Redundanzen, ein Grundsatz, der übrigens nicht erst seit ISAD (G) besteht.

Durch diese Redundanzen vermeidende Verzeichnung im Rahmen einer hierarchischen Ordnung sind archivierte Aktentitel extrem kontextabhängig. Diese Kontextabhängigkeit der Titelbildung (nicht der Akten) ist auch der Grund, warum es für die Archive ein spartenspezifisches Archivportal-D innerhalb der Deutschen Digitalen Bibliothek gibt.

Unsere archivistische Erschließung besteht also aus hierarchischer Klassifikation *und* Titel. Sie sieht dadurch zwar strukturiert aus, ist aber keine Klassifikation im informationswissenschaftlichen Sinne¹⁸, wie z. B. eine bibliothekarische Systematik, deren Struktur auf einem kontrollierten Vokabular basiert und die in allen Bibliotheken identisch verwendet werden kann. Unsere Erschließung ist behördenspezifisch, also für jeden Bestand in jedem Archiv anders strukturiert, und folgt auch keinen festen, abgestimmten Begriffshierarchien. Dies zwingt Archivar und Nutzer immer von Neuem zum mühsamen Navigieren entlang einer unbekanntenen, hoch proprietären Klassifikation – oder eben zum Suchen.

Vom Nutzen und Nachteil einer Klassifikation

Grundsätzlich gibt es zwei Auswege aus diesem Dilemma: Wir ändern unsere Erschließung oder wir ändern die Suchfunktion.

Die häufige Kritik, die maschinelle Volltextsuche sei für Archivgut nicht adäquat und Benutzern müsse die navigierende Recherche vermittelt werden, geht zunehmend an der Realität vorbei und übersieht die massiven informationswissenschaftlichen Schwächen unserer Erschließung.¹⁹

Der Erfolg von Google war am Anfang vor allem ein Sieg der Suchmaschine gegenüber hierarchischen Webverzeichnissen wie Yahoo.²⁰ Denn funktionierende Klassifikationen setzen einen begrenzten Corpus mit klaren und abgrenzbaren Kategorien voraus. Diese Kategorien müssten nicht nur die erschließenden Archivare, sondern auch die Benutzer kennen und beherrschen.²¹ In einer Behörde mit professionell geführter Registratur und gepflegtem Aktenplan vermochte die hierarchische Organisation des Schriftgutes ein ebenso effektives Ablegen wie präzises Auffinden der Akten zu gewährleisten. „Für die Benutzer von Archiven“, so fasste Georg Vogeler jüngst zusammen, „war die Ordnung schon immer fremd und musste erst erlernt werden.“²² Hinzu kommt, dass die Erschließungskategorien der Registraturbildner sich nur selten mit den Erschließungskategorien der Benutzer decken, die Akten also von den Benutzern in immer neue sachthematische Kontexte eingeordnet werden.²³

Im Ergebnis muss man festhalten, dass die archivistische Klassifikation informationswissenschaftlich betrachtet weder eine gute, d. h. strukturierte und kontrollierte Informationsstruktur ist, noch den Anwendern die Recherche er-

leichtert. Für die Nutzer von heute ist Google der Maßstab für erfolgreiche Recherchen. Etwas Anderes einzufordern erschiene wohl zwecklos und zunehmend weltfremd.

Volltextsuche vs. Google

Das Hauptproblem bei der Diskussion um die Nachteile einer Volltextsuche für die Archivrecherche liegt darin, dass viele die Volltextsuche z. B. von Archive NRW mit Google verwechseln. Beides heißt zwar „Volltextsuche“ und hat einen Suchschlitz. Aber es wäre vermessen anzunehmen, unsere archivischen Datenbankabfragen könnten nur ansatzweise mit der hochkomplexen Recherchefunktion konkurrieren, die wir nutzen können, um Milliarden von unstrukturierten Webseiten, Texten, Bildern und anderen Informationen zu durchsuchen. Mit Hilfe von Statistik, Linguistik, Semantik und vielem mehr kann Google relevante Seiten und Informationen finden, auch ohne dass dort der gesuchte Begriff überhaupt vorhanden ist. Google nutzt für seine Suchräume zudem eine Masse zusätzlicher Informationsressourcen, neben Wikipedia und Wikidata etwa komplexe semantische Wörterbücher, die morphologische und lexikalische Normalisierungen ebenso erlauben, wie das Finden von Synonymen oder semantische Verknüpfungen (Knowledge Graph). Google weiß, ob die Pizzeria um die Ecke noch offen ist, auch wenn der Inhaber seine Öffnungszeiten nicht in einem bestimmten XML-Format oder in einer Datenbank abgelegt hat.

Man darf sich von dem einfachen Suchschlitz nicht täuschen lassen: Im Vergleich zu Google bleiben die Volltextsuchen unsere archivischen Suchportale weit zurück. Denn dort findet zumeist tatsächlich nur eine einfache Suche nach Stichworten über die Einleitungstexte, Bestandsnamen, Klassifikationen, Titel und Enthält-Vermerke statt bzw., sofern vorhanden, über die ergänzten Index-Begriffe.

Natürlich mit fatalem Ergebnis: Relevante Akten werden nicht angezeigt, weil sie die gesuchten Begriffe nicht enthalten oder in anderer Schreibweise oder nur Synony-

18 Zur Einführung Ulrich Reimer, B 3 Wissensorganisation, in: Rainer Kuhlen (Hrsg.) Grundlagen der praktischen Information und Dokumentation. Handbuch zur Einführung in die Informationswissenschaft und -praxis, Berlin u. a. 62014, S. 172–181.

19 Vgl. zum Folgenden den anregenden Beitrag von Georg Vogeler, Zum Verhältnis von klassischen Formen der Archiverschließung und den Zugängen der Digital Humanities zum Information Retrieval, in: Elisabeth Schögl-Ernst/Thomas Stockinger/Jakob Wührer (Hrsg.), Die Zukunft der Vergangenheit in der Gegenwart. Archive als Leuchtturm im Informationszeitalter, Wien 2019, S. 199–212, dessen vorab vom Autor bereitgestellten Druckfahnen dieser Text wichtige Inspirationen verdankt.

20 Vgl. Clay Shirky, Ontology is Overrated, in: Clay Shirky's Writings About the Internet, 2005, http://shirky.com/writings/herecomeseverybody/ontology_overrated.html.

21 Clay Shirky, Ontology is Overrated (wie Anm. 20), fasst als Bedingungen für das Funktionieren solcher ontologischen Klassifikationen zusammen: „Small corpus, formal categories, stable entities, restricted entities, clear edges“ bzw. „expert catalogers, authoritative source of judgment, coordinated users, expert users“.

22 Georg Vogeler, Zum Verhältnis (wie Anm. 19), S. 206.

23 Zum „Kontext“ vgl. die Denkanstöße von Christian Keitel, Zwölf Wege ins Archiv. Umriss einer praktischen Archivwissenschaft, Stuttgart 2018, S. 180–187.

me. Die Benutzer finden nur die Begriffe, die sie eingegeben haben, aber nicht die Informationen, die sie suchen.

Das liegt aber nicht nur an unserer Erschließung, sondern vor allem an der Unzulänglichkeit der dafür eingesetzten Volltextsuche. Die Informationswissenschaftlerin Jutta Lindenthal hat es vor Jahren in einem Gutachten zur Verbesserung der Deutschen Digitalen Bibliothek (DDB) auf den Punkt gebracht: „Für facettenbasierte Retrievalfunktionen und semantische Suchfunktionen, wie die Deutsche Digitale Bibliothek sie anbietet oder anstrebt, sind standardisierte, kontrollierte und konsistente Daten jedoch eine unabdingbare Voraussetzung.“²⁴ Solche Strukturen und Angaben liegen den archivischen Erschließungen aber nicht zu Grunde.

Vor allem liefern unsere Suchfunktionen bislang noch kein überzeugendes Ranking der Suchergebnisse. Wie wichtig, vielmehr: wie entscheidend dies ist, zeigt uns Google aber jeden Tag: Das Ranking in der Trefferliste ist das eigentliche Erfolgskriterium. Und mit wachsenden Datenmengen wird diese Relevanzbewertung im Zusammenspiel mit Vollständigkeit (recall) und Genauigkeit (precision) immer wichtiger.

Diese Relevanzbewertung von Treffern für eine Suchanfrage ist bei den Archivrecherchesystemen kaum entwickelt. Dass Benutzer also bei einfachen Suchanfragen zu viele Treffer ohne Relevanzsortierung erhalten, ist kein Problem der Volltextsuche, sondern der fehlenden Ranking-Algorithmen unserer Archivprogramme.

Vom Titel zum Tag

Wir müssten also vor allem unsere Suchfunktionen verbessern. Aber müssen wir auch unsere Verzeichnung verbessern, um sie für Suchfunktionen besser findbar zu machen?

In der Praxis geschieht das wohl schon vielfach. Beim Stöbern durch die deutschen Archivportale zeigt sich, dass oftmals die Aktentitel, vor allem aber die Enthält-Vermerke dazu benutzt, Papritz würde wohl sagen: missbraucht werden, um dort möglichst viele Begriffe unterzubringen und aufzuzählen, verbunden mit der Hoffnung, dass ein Benutzer diese Stichwörter über die Suchfunktion finden kann.

Im Prinzip ist das natürlich ein gangbarer Weg, angesichts der unzulänglichen Suchfunktion den Erfolg des Information Retrieval zu verbessern. Allerdings ein sehr, sehr aufwendiger. Und da auch hierfür keine kontrollierten Vokabularien verwendet werden, ist ein Erfolg natürlich weiterhin von dem Zufall abhängig, welche Suchwörter der Nutzer später tatsächlich eingibt.

Informationswissenschaftlich betrachtet geht es bei diesem Weg darum, der archivischen Verzeichnung kategoriale Beschreibungen hinzuzufügen: Ich vergebe für Dokumente oder Akten zusätzliche Schlagwörter, Index-Begriffe, modern: Tags, die der Benutzer wiederfinden kann, wenn er diese Wörter beispielsweise bei einer Volltextsuche eingibt.

Das ist zwar recht „unarchivisch“ gedacht. Aber wenn man in die Dokumentenmanagementsysteme heutiger Verwaltungen und Unternehmen schaut, findet man ge-

naues dieses Tagging, das die Mitarbeiter verwenden, um ihre Dokumente und Texte zu kennzeichnen und wiederzufinden. Es ist neben der Volltextsuche, die oft zu viele irrelevante Treffer bringt, die gängigste Art der Wissensorganisation. Aber auch hier ergibt sich das Problem, dass diese Schlagwörter zumeist nicht zentral koordiniert bzw. – informationswissenschaftlich gesprochen – kontrolliert sind, sondern individuell vergeben werden. Dahinter liegen keine Systematik und kein abgestimmter Thesaurus. Ob also ein anderer Verwaltungsmitarbeiter die Dokumente damit findet, ist völlig offen. Vom Archivar oder einem (viel) späteren Benutzer ganz zu schweigen.

Den gleichen Effekt und die gleichen Probleme erzielt man, wenn man Archivgut ungesteuert von Benutzern verschlagworten oder – in Web. 2.0-Sprache – taggen lässt. Die Bibliotheken haben dieses Social Tagging zumeist schon vor einigen Jahren wieder aufgegeben. Es funktioniert als Crowdsourcing wunderbar bei der ersten Identifizierung von Personen, Orten und Bildern. Aber für die Indexierung von Sachakten, deren Aussage- und Auswertungspotential unbegrenzt ist, bleiben alle diese „Tags“ oder „Folksonomies“ immer nur sehr individuelle und nutzungsspezifische Inhaltsangaben, – wenn Sie nicht ohnehin nur Aussagen enthalten wie „tolle Akte“, „historisch wichtig“ oder einen Smiley und einen „Daumen hoch“.

Die Vergabe von Indexbegriffen, die traditionelle Form Archivgut mit Schlagworten anzureichern, ist in den veröffentlichten Verzeichnungsrichtlinien der deutschen Archive zumeist ebenfalls nur für Personen- und Ortsnamen vorgesehen. Für Sachbetreffende fehlen dazu sowohl innerhalb der Archive wie auch deutschlandweit oder gar spartenübergreifend abgestimmte Thesauri.²⁵

Normdaten

Alle diese Anreicherungen mit Metadaten, auch so könnte man diese Form der Schlagwortergänzung bezeichnen, sind natürlich ungemein aufwendig und erfordern zusätzliche Ressourcen. Angesichts der zunehmenden Rückstände, besonders bei der Erschließung, ist diese Form der erweiterten Erschließung daher kaum zu empfehlen.

Dabei stellen sich insbesondere für Orts- und Personennamen noch höhere Ansprüche an die Archive. Sie sollen künftig nicht nur einheitlich in die Erschließungen aufgenommen werden, sondern zugleich mit sogenannten

24 Jutta Lindenthal, Datenqualität und Retrieval. Vorschläge zur Verbesserung der Suche in der Deutschen Digitalen Bibliothek, 4. Februar 2016, Version 1.0. http://jl.balilabs.de/DDB/DQ/DDB_Datenqualität%20C3%A4t_Retrieval_1.0.pdf.

25 Vgl. dazu zuletzt Frank M. Bischoff, Überlegungen zur Zukunftsfähigkeit archivischer Webangebote, in: Archivpflege in Westfalen-Lippe 90 (2019), S. 36–47, hier: S. 42. – Einen Versuch unternimmt jetzt das vom Bundesarchiv und dem Landesarchiv Baden-Württemberg initiierte Projekt zum Aufbau einer Infrastruktur zur Implementierung sachthematischer Zugänge im Archivportal-D am Beispiel des Themenkomplexes „Weimarer Republik“, das mittels Ergänzung einer kontrollierten Sachsystematik im Archivportal D ein sachthematisches Inventar für die Weimarer Republik zu erstellen versucht, allerdings ohne erneute Autopsie der Quellen, sondern nur anhand der bereits vorhandenen Verzeichnungsangaben. <https://www.landesarchiv-bw.de/web/63525>.

Normdaten verknüpft werden, also etwa mit der Gemeinsamen Normdatei der Deutschen Nationalbibliothek (DNB).

Der Grund ist einleuchtend: Wenn man Otto Müller oder den Ortsnamen Neustadt eingibt, dürfte es ziemlich schwierig werden, den richtigen Ort oder den richtigen Müller zu finden. Wenn Benutzer künftig in spartenübergreifenden Portalen wie der DDB oder sprachenübergreifend in der Europeana suchen wollen oder anstelle einfacher Treffer künftig auch semantische Verknüpfungen möglich sein sollen – immer dann werden wir um Normdaten nicht herumkommen.²⁶

Doch auch hier stellt sich die Frage, ob diese Verknüpfungen und Annotationen von Erschließungsangaben in allen Archiven manuell angelegt werden müssen (z.B. eine Georeferenzierung von Orten) oder ob wir nicht unseren Ehrgeiz drauf richten sollten, dies über intelligente, maschinelle Verfahren zu erreichen.

Zusammenfassung

Bei der Diskussion um die Frage, wie man im Zeitalter der Digitalen Transformation erschließen soll, bleibt es das Wichtigste, dass Archive ihre Bestände überhaupt erschließen. Statt mit großem Aufwand umfassende Enthält-Vermerke zu formulieren, sollten wir die knappen Personal-Ressourcen besser in die Erschließung der Rückstände investieren. Und selbst eine einfache, flache oder archivfachlich unausgereifte Verzeichnung ist besser als gar keine.

Die archivische Klassifikation ist informationswissenschaftlich betrachtet keine gute, d.h. strukturierte und kontrollierte Informationsstruktur, und sie erleichtert den Benutzern auch nicht die Recherche. Auch unsere stark kontextabhängigen Aktentitel sind nicht für eine einfache maschinelle Volltextsuche gemacht. Beides erlaubt aber eine wirtschaftliche und für alle Auswertungszwecke gleichermaßen offene Erschließung unserer Bestände.

Statt unsere Erschließung zu ändern oder die „Datenqualität“ durch die Ergänzung von Schlagworten und Tags

für Suchmaschinen zu optimieren, sollten wir viel mehr Mühe darauf verwenden, die Leistungsfähigkeit der Suchmaschinen vor unseren Portalen deutlich zu verbessern. Dies erscheint einfacher und praktikabler, als erneut alle Akten in die Hand zu nehmen und Millionen von Erschließungsdaten anzupassen oder anzureichern. Die Bibliotheken gehen seit einigen Jahren diesen Weg und entwickeln anstelle der klassischen OPACs zunehmend „Recovery Systeme“, die den Nutzern komfortable, intuitive Suchfunktionen und „Entdeckungen“ über alle (analogen wie digitalen) Medienbestände und Informationsressourcen der Bibliothek bieten sollen.

Google bleibt der Maßstab, an dem sich auch Archive werden orientieren müssen. Das bedeutet etwa, bei der Volltextsuche linguistische, semantische und statistische Methoden einzubinden sowie den Suchraum durch eigene und fremde Kontexte zu erweitern – und zwar sowohl für die Ermittlung von Treffern als auch für deren Ranking.

Und vielleicht können unsere vergleichsweise kleinen Archivdatenbanken nebenher auch von den erstaunlichen Fortschritten profitieren, die sich derzeit bei der Entwicklung intelligenter, neuronaler Mining- und Recherchewerkzeuge für die Analyse riesiger Forschungsdatenbestände (Big Data) zeigen. ■



Prof. Dr. Mario Glauert
Brandenburgisches Landeshauptarchiv, Potsdam
mario.glauert@blha.brandenburg.de

²⁶ Vgl. Gerhard Müller, Die Normdatei als ein Mittel in der Erschließung von Archivbeständen, in: Brandenburgische Archive 32 (2015), S. 8–16, http://blha.brandenburg.de/wp-content/uploads/2017/06/Brandenburgische_Archive_32_2015.pdf.