Zeitschrift: Geschichte und Informatik = Histoire et informatique
Herausgeber: Verein Geschichte und Informatik
Band: 11 (2000)

Artikel: Historische Datenbanken : Vorteile und Probleme
Autor: Thaller, Manfred
DOI: https://doi.org/10.5169/seals-8924

Nutzungsbedingungen
Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Mehr erfahren

Conditions d'utilisation
L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. En savoir plus

Terms of use
The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. Find out more

Download PDF: 08.07.2025

ETH-Bibliothek Zürich, E-Periodica, https://www.e-periodica.ch

Historische Datenbanken: Vorteile und Probleme

Manfred Thaller

I. Einleitung

Die Verwendung von Datenbanken in der Geschichtswissenschaft reicht weit zurück. Im Grunde genommen begann Sie vor der Konjunktur, wenn nicht der Erfindung des Begriffs. Eine der ersten internationalen Tagungen zum Thema «EDV-Einsatz in den Historischen Wissenschaften» fand 1963 auf der österreichischen Burg Wartenstein statt – an welcher disziplinären Zuordnung der Titel des Tagungsbandes «The use of Computers in Anthropology»[1] nichts ändert. Der Begriff der «Datenbank» war zu jener Zeit erst im Entstehen, das Thema «database» fehlt dementsprechend – doch 13 der 18 Beiträge beschäftigen sich mit dem Gegenstand «information retrieval» und würden aus heutiger Sicht unter dem Thema unserer Tagung klassifiziert.

Die Beschäftigung mit einem, aus der Sicht einer sich rasch weiterentwickelnden Disziplin wie der Informatik fast steinzeitlich anmutenden, Text von 1963 mag überraschen. In der Tat sind manche der Überlegungen aus heutiger Sicht ein wenig amüsant: So die Zusammenfassung eines Teils der Diskussion durch den Herausgeber im Vorwort, Makroassembler, wenn auch nicht so einfach wie Höhere Programmiersprachen, seien doch wesentlich einfacher als das direkte Programmieren in Maschinensprache, das dem Geisteswissenschaftler eher nicht zu empfehlen sei – ob sich für diesen der höhere Bedienungskomfort der maschinenferneren und daher ineffektiveren Höheren Programmiersprachen gegenüber der nur wenig schwierigeren Makroassemblerprogrammierung aber wirklich auszahle, sei noch unentschieden.

Diese Überholtheit in technischen Details verstellt allerdings manchmal den Blick darauf, dass die grundsätzlichen Überlegungen zum Verhältnis von historischen Wissenschaften und der Datenverarbeitung aus jenen Tagen durchaus noch Gültigkeit haben, wenn sie sich vom zeitgebundenen technischen Detail lösen, wie etwa die ebenfalls in der Zusammenfassung des Herausgebers enthaltene Aussage, für die Nutzung des Rechners in einem Forschungsprojekt «… there are just two jobs required: (1) the

1 Hymes, Dell (Hg.): *The use of Computers in Anthropology*. London etc. 1965.

formulation of the procedure in detail, and (2) the translation of this procedure into machine operations … *And of these two jobs, for most areas of anthropology, the second will be easier than the first.*» (Meine Hervorhebung). Aus dieser Feststellung wird schliesslich eine Aussage abgeleitet, die seither und wohl auch in Zukunft das Leitthema aller Diskussionen unter Historikern über den Einsatz von Datenbanken oder «die Datenverarbeitung» sein sollte. Sollte; denn so explizit wie in diesem frühen Aufsatz geschieht es leider nur selten: Die grundlegende und nicht entschiedene Frage sei, ob der Rechner ein reines Werkzeug zur Administration von Belegen sei oder ob er methodische Relevanz habe.

In dieser viel zu selten explizit geführten Diskussion haben Datenbanken mindestens als *ein*, wenn nicht *das* zentrale Paradigma des Rechnereinsatzes in den historischen Disziplinen stets eine hervorragende Rolle gespielt. Jean-Claude Gardin bezieht sich zwar explizit auf Expertensysteme, wenn er zur radikalsten Formulierung der einen Seite in dieser Debatte kommt: Der wirkliche Wert der Beschäftigung der historischen Disziplinen mit der Rechnertechnologie liege nicht in den Ergebnissen, sondern im Zwang, die eigenen Prämissen endlich so präzise zu formulieren, dass sie rechnergestütztem Kalkül zu Grunde gelegt werden könnten[2]. Nachdem die Expertensysteme als Modethema aber ausgedient haben, würden die meisten Historiker die von ihm beschriebenen konkreten Probleme wohl eher als Bestandteil der Datenbankproblematik betrachten (J.-C. Gardin ist übrigens der einzige Teilnehmer des Treffens von 1963, der seine gesamte akademische Laufbahn bis zur Pensionierung im Umfeld der Rechnernutzung in den historischen Disziplinen verblieben ist.). Und, klarer noch: Wenn Datenbankanwendungen für die Verfechter des Standpunkts, dass die Rechnernutzung als solche methodische Relevanz für die historischen Disziplinen habe(n solle), nahe dem Zentrum der Argumentation stehen – für die Verfechter des Standpunktes, der Rechner als solcher habe als blosses Werkzeug zu dienen, sind diese Anwendungen meist ein Angelpunkt. Was vielleicht am klarsten in der ausbruchsartigen Verdammung der Datenbankanwendungen als einer Verführung weg von der reinen Lehre quantitativer Studien darstellt, mit der Alain Guerreau eine der Konferenzen kommentiert hat, aus denen später die AHC hervorging: «La veritable hysterie de la ‹banque de données› est la simple traduction de

2 Mein Versuch, Gardin, Jean-Claude: Le Calcul et la raison. Essais sur la formalisation du discours savant. Paris 1991. zusammenzufassen.

cette naïve idéologie.»[3] Wobei diese naive Ideologie eben in der Annahme besteht, der Rechneranwendung als solcher könne ein methodischer Wert innewohnen.

Jean-Claude Gardin mag für die meisten Historiker den Inhalten zu abgewandt sein, Alain Guerreau übersehen haben, dass die Informatik eine eigenständige Wissenschaft mit viel grösserer Breite als die Statistik ist: Die grosse Mehrzahl neuerer Texte betrachtet Datenbankanwendungen eindeutig als den Königsweg der Rechneranwendung in den historischen Disziplinen. Sei es, dass sie als eine Weiterführung traditioneller historischer Arbeit mit anderen Mitteln betrachtet werden: «Historians, by necessity, are constantly developing systems for collecting, organizing, and selectively retrieving information. Whether computerized or not, such systems are database management systems or, more simply, databases.»[4] Sei es, dass sie in jenem Stil permanenter Erregung, der die neuere englischsprachige Literatur gerade im Informatikbereich kennzeichnet, geradezu als Heilsweg erscheinen: «The particular appeal of database-centered research is not difficult to understand. Data gathering, organisation, sorting and searching, are tasks routinely carried out by historians. … Of all the subjects that deserve systematic treatment in book form, none is more important than that of database systems. The design and development of databases is central to the transformation of research methods in history which is now underway: to the development of superior systems for the location of primary and secondary sources; … ; to the analysis and presentation of historical information».[5]

Die grosse Mehrzahl neuerer Texte … Auch die neuesten? Auf den ersten Blick könnte man daran irre werden, wenn man feststellt, dass in einem 1998 erschienenen Konferenzband[6] nur 3 von 15 Verfassern Datenbanken erwähnen (wir erinnern uns: 13 von 18, 35 Jahre davor) – doch dies wäre ein Trugschluss. *Alle* 15 Autoren beschäftigen sich mit typischen Datenbankaufgaben, wie «Finden von Informationen», genauso, wie sich Ihre Vorgänger in den sechziger Jahren mit *Information Retrieval* beschäftigt haben. Ihr Medium dafür ist freilich das Internet – und dass dieses Informationen bereitstellt, interessiert mehr als die bei allen grösseren

3 Rezension von «Denley, Peter and Hopkin, Deian: *History and Computing*» durch Guerreau, Alain. In: *Histoire et Mesure,* Bd. 3, 1988, S. 144 - 146.

4 Greenstein, Daniel I: *A Historian's Guide to Computing*. Oxford 1994, S. 61.

5 Harvey, Charles; Press, Jon: *Databases in Historical Research*. Basingstoke 1996, S. xi und 1.

6 Trinkle, Dennis A. (Hg.): Writing, Teaching and Researching History in the Electronic Age. Historians and *Computers*. Armonk and London 1998.

Diensten dahinter stehenden Datenbanken. Wobei der Kenner der Probleme, die sich ergeben, wenn historische Information in grossem Massstab rechnergestützt geordnet und handhabbar gemacht werden soll, freilich nun seinerseits Mühe hat, ein Schmunzeln über die intellektuelle Naivität mancher Beiträge zu unterdrücken, das ihn Guerreau gegenüber nachträglich versöhnlicher stimmt: «To organize all of the bits of data as they came in, I resorted to a modified version of 3x5 cards. Instead of producing physical cards, I typed my information into a personal computer and used it to organize my material into virtual index cards, or what I have come to call «electronic notecards». Over time, I streamlined the note-taking process to allow the computer to perform as much of the sorting and labeling as possible.»[7]

Zusammenfassend:

- Seit Beginn der Beschäftigung mit «dem Rechner» oder «der Informatik» wird die Ordnung von und die Suche nach Informationen als eines der Haupteinsatzgebiete betrachtet.
- Häufig unterschwellig diskutiert, besteht dabei erheblicher Dissens über den *methodischen* Stellenwert der dabei eingesetzten Techniken.
- Während die technischen Werkzeuge und die praktischen Vorgehensweisen zur Realisierung des Ziels «Ordnen und Suchen» rascher Veränderung und vielen Moden unterliegen, wobei erhebliche Fortschritte in der Effektivität und Effizienz erkennbar sind, ist es wesentlich schwieriger, Fortschritte in der intellektuellen Diskussion zu erkennen.

II. Grundbegriffe für eine Typologie von Datenbankanwendungen in der Historischen Forschung

Um bei einer Darstellung von «Vorteilen und Problemen» der Datenbanktechnik in den Geschichtswissenschaften nicht an den zufälligen Äusserlichkeiten eines nur kurzfristig geltenden technischen Entwicklungsstandes kleben zu bleiben, wollen wir daher zunächst versuchen, eine Typologie von Vorgehensweisen zu entwickeln, die sich darauf konzentriert zu beschreiben, was bei verschiedenen Arten der Übertragung von Information, wie sie in historischen Quellen vorkommt, in den Rechner mit dieser Information geschieht, bevor wir uns einer Abschätzung der Vor- und Nachteile auf dieser Basis verglichener konkreter Arbeitsweisen zuwenden.

7 Jones, James A.: «Constructing History with Computers». In *ebdt.,* S. 83 - 88, hier: 83.

Dazu beschreiben wir diese Vorgehensweisen zunächst anhand abstrakter Kategorien.

Um dabei bewusst in der Tradition der in der internationalen AHC zu einem früheren Zeitpunkt stattgefundenen Debatten zu bleiben, wählen wir dazu folgendes (fiktives) Quellenbeispiel, das vom Verfasser bereits früher für ähnliche Zwecke herangezogen wurde[8]:

«At Michaelmas in the fortieth year of Elizabeth, our gracious Queen, Johnny Turner, who holds the smitthy behind the church, has paid the seven guineas he owed to his brother Frederick from the will of his late father, and all his property is now free of any obligations whatsoever.»

Wir wollen uns hier mit einer Typologie der Ansätze beschäftigen, mit denen Historiker im Laufe der Zeit datenbankartige Dienstleistungen realisiert haben, die wir salopp als «Ordnen und Suchen» charakteristiert haben. Dazu sind natürlich nicht notwendigerweise Datenbankprogramme im engeren Sinne nötig. Wir beginnen daher mit statistischer Software. Für diese würde der Informationsgehalt unseres fiktiven Beispiels etwa folgendermassen aufbereitet:

```
09071598_John_________________Turner_______________456_0147_110_7_0
```

d.h., durch die Position eines Codes oder Textfragments an einer bestimmten Stelle wird festgelegt, welche Art von Information durch die einzelnen Fragmente und Codes ausgedrückt wird.

Was üblicherweise durch ein «Codebook» etwa folgender Art dokumentiert wird:

```
Month                  01 02
Day                    03 04
Year                   05 08
Christian Name         10 29
Surname                30 49
Occupation             50 52
Amount in sh.          54 57
Payment to             59 61
   (1xx == Relative
    11x == Brother)
Type of obligation     63 63
Obligation remaining?  65 65
```

8 Thaller, Manfred: «The Need for Standards: Data Modelling and Exchange». In: Greenstein, Daniel: (Hg.): *Modelling Historical Data. Towards a Standard for Encoding and Exchanging Machine-Readable Texts.* St. Katharinen 1991 (= *Halbgraue Reihe zur Historischen Fachinformatik* A 11), S. 1 - 18.

Wenn wir versuchen, zur besseren Vergleichbarkeit verschiedener Ansätze, diese Vorgehensweise auf der Basis einiger abstrakter Kategorien zusammenzufassen, können wir sagen:

1. Visuelle Charakteristika der Quelle werden ignoriert.
2. Die Syntax der Quelle («welche Zeichenkette bezieht sich auf welche Kategorie») wird ignoriert, da diese Zuordnung durch Umpositionierungen ausgedrückt werden.
3. Die Semantik der Quelle wird ignoriert, da die Begriffe durch Codes vereinheitlicht werden.

Würden wir dieselbe Quelle zur Bearbeitung in einer relationalen Datenbank speichern, erhielten wir etwa folgendes Ergebnis:

Surname	*1st name*	*Occupation*	*Payment in sh.*	*Day*	*Month*	*Year*
Turner	John	Blacksmith	147	7	Sep	1598

Wenden wir auf diese Darstellung die gleichen Kategorien an wie eben, erhalten wir:

1. Visuelle Charakteristika der Quelle werden ignoriert.
2. Die Syntax der Quelle («welche Zeichenkette bezieht sich auf welche Kategorie») wird ignoriert, da diese Zuordnung durch Umpositionierungen ausgedrückt werden.
3. Die Semantik der Quelle wird ignoriert, da die Begriffe durch Codes vereinheitlicht werden.

Mit anderen Worten: Diese Darstellung einer Quelle auf dem Rechner ist der Quelle selbst exakt gleich ähnlich oder unähnlich als die frühere. «Blacksmith» mag dem menschlichen Leser als unmittelbarer verständlich erscheinen als die Codegruppe «456», hat mit dem eigentlichen Inhalt der Quelle («who holds the smitthy behind the church») aber genauso viel oder wenig zu tun: In *beiden* Fällen findet eine Umsetzung oder Codierung statt. (Wenn Codierungen in Codesysteme, die eine statistische Auswertung optimieren sollen, auch meist, aber keineswegs immer, etwas anders angelegt sind als solche, die die Verarbeitung mit Datenbanksoftware vorbereiten sollen.)

Geeignete Software vorausgesetzt[9], sind jedoch durchaus auch Arten der Umsetzung denkbar, die wesentlich näher an der Quelle bleiben – wie etwa die folgende:

```
<payment>
<date>Michaelmas 40 Eliz I</date><sum>7g</sum>
  <furtherobligation>none</furtherobligation><reason>will</reason>
</payment>
<holder>
<surname>Turner</surname><cname>Johnny</cname>
  <occupation original="holds the smitthy behind the
church">blacksmith</occupation>
</holder>
<addressee>
<cname>Frederick</cname><relationship>brother</relationship>
</addressee>
```

Freilich sei betont, dass diese Vorgehensweise ganz bestimmte Softwarefunktionalitäten und eine gezielt eingesetzte Logik bei der Handhabung des Auszeichnungsystems voraussetzt, die diesem nicht zwingend innewohnen muss. Was sich ergibt, wenn wir versuchen, diesen Ansatz anhand der von uns erarbeiteten Kategorien zu charakterisieren:

1. Visuelle Charakteristika der Quelle werden ignoriert.
2. Die Syntax der Quelle («welche Zeichenkette bezieht sich auf welche Kategorie») wird ignoriert; diese Zuordnung wird durch die Auszeichnung ausgedrückt.
3. Die Semantik der Quelle wird beibehalten. Sie wird entweder durch Software aufgelöst (Datierungen, Geldmengen) oder durch eine Aufteilung auf unterschiedliche Auszeichnungsebenen verwaltet (Beruf).

Unter Fortschreibung dieses einmal begonnenen Ansatzes können wir uns der Quelle noch wesentlich stärker nähern; etwa durch folgende Form der Aufbereitung für den Rechner:

---

9 Dem Leser des Jahres 2000 werden die folgenden Beispiele automatisch als Anwendungsfälle für die derzeit diskutierten Markupsprachen wie SGML und XML erscheinen. Diese Sprachen *eignen* sich unstreitig für derartige Anwendungen (auch wenn ihre Verwendung erst dann und nur dann Sinn macht, wenn klar ist, welche Software die mit ihrer Hilfe ausgezeichneten Texte unterstützt), sie sind jedoch keine *Voraussetzung* für die beschriebenen Ansätze. In der Tat kennen die historischen Disziplinen Beispiele für die Verwendung der im folgenden beschriebenen Vorgehensweisen, die lange vor die Erfindung von SGML zurückreichen – z.B. die für das Earls Colne Projekt an der Universität Cambridge entwickelte Softwareumgebung und Auszeichnungssprache: King, Tim J.: «The Use of Computers for Storing Records in Historical Research». In: Historical Methods, vol. 14, 1981, S. 59 - 64.

```
<payment><date> At <feast> Michaelmas </feast> in the <year>
fortieth </year> year of <ruler> Elizabeth </ruler></date>, our
gracious Queen, <person><cname> Johnny </cname> <surname> Turner
</surname>, who <status> holds the smitthy behind the church
</status> </person>, has paid the <amount> seven guineas </amount>
he owed to his <addressee> <relationship> brother
</relationship><cname> Frederick </cname></addressee> from the
<reason> will of his late father </reason>, and all his property
is now <further> free of any obligations </further>
whatsoever.</payment>
```

Offensichtlich wird der Text hier als fortlaufend lesbare Darstellung erhalten und seine Struktur wird für den Rechner durch die Auszeichnung nur unterstrichen, nicht in eine völlig andere Darstellungsweise übertragen. Oder, systematischer:

1. Visuelle Charakteristika der Quelle werden ignoriert.
2. Die Syntax der Quelle wird durch die Auszeichnung hervorgehoben, sodass sie als Kriterium durch die interpretierende Software verstanden werden kann.
3. Die Semantik der Quelle wird beibehalten. Sie wird entweder durch Software aufgelöst (Datierungen, Geldmengen) oder durch eine Aufteilung auf unterschiedliche Auszeichnungsebenen verwaltet (Beruf).

Und eine sehr kleine weitere Änderung bringt uns schliesslich zu einer Verbindung mit einer photographischen Abbildung der Quelle:

```
<payment original="http://meinserver/archiv/dok2345.jpg"><date> At
<feast> Michaelmas </feast> in the <year> fortieth </year> year of
<ruler> Elizabeth </ruler></date>, our gracious Queen,
<person><cname> Johnny </cname> <surname> Turner </surname>, who
<status> holds the smitthy behind the church </status> </person>,
has paid the <amount> seven guineas </amount> he owed to his
<addressee> <relationship> brother </relationship><cname>
Frederick </cname></addressee> from the <reason> will of his late
father </reason>, and all his property is now <further> free of
any obligations </further> whatsoever.</payment>
```

Durch diese Verbindung wird schliesslich auch die letzte unserer Ausgangsannahmen abgelöst; wir erhalten damit:

1. Visuelle Charakteristika der Quelle werden bewahrt.
2. Die Syntax der Quelle wird durch die Auszeichnung hervorgehoben, sodass sie als Kriterium durch die interpretierende Software verstanden werden kann.

3. Die Semantik der Quelle wird beibehalten. Sie wird entweder durch Software aufgelöst (Datierungen, Geldmengen) oder durch eine Aufteilung auf unterschiedliche Auszeichnungsebenen verwaltet (Beruf).

Dabei sei nachdrücklich betont, dass es uns um keine chiliastische Heilslehre geht. Viele dieser Charakteristika sind unabhängig voneinander, so ist auch folgendes Beispiel denkbar:

Surname	*1st name*	*Occupation*	*Payment in sh.*	*Day*	*Month*	*Year*	*Original*
Turner	John	Blacksmith	147	7	Sep	1598	http://meinserver/archiv/dok2345.jpg

Was uns zu folgender Beschreibung an Hand der von uns gewählten Dimensionen bringen würde:

1. Visuelle Charakteristika der Quelle werden bewahrt.
2. Die Syntax der Quelle («welche Zeichenkette bezieht sich auf welche Kategorie») wird ignoriert, da diese Zuordnung durch Umpositionierungen ausgedrückt werden.
3. Die Semantik der Quelle wird ignoriert, da die Begriffe durch Codes vereinheitlicht werden.

III. Eine Typologie von Datenbankanwendungen in der Historischen Forschung

Wenn wir akzeptieren, dass die drei genannten Dimensionen für die Abbildung von Informationen in historischen Quellen, jeden Versuch zur Speicherung derartiger Informationen zum «Ordnen und Suchen» innerhalb eines einheitlichen Kategoriensystems beschreibbar machen, können wir eine Reihe von Ansätzen unterscheiden, die zunächst völlig unterschiedlich aussehen, aber eben doch nur unterschiedliche Ausprägungen des Phänomens «Datenbanken in der historischen Forschung» sind. Unterschiedliche Ausprägungen eines einheitlichen Phänomens – für die die gestellte Frage nach den Vorteilen und Problemen von Datenbanken in der historischen Forschung deshalb unterschiedlich beantwortet werden muss.

Hierfür schlagen wir folgende Typologie von Datenbankanwendungen in den historischen Disziplinen vor:

1. Exzerptsammlungen und Bibliographien

Ziel dieser Art von Datenbankanwendung in der Geschichte ist die Vereinfachung handwerklicher Vorgänge mit dem geringstmöglichen Aufwand. Derartige «persönliche Datenbanken» werden als Hilfsmittel der eigenen

Arbeit angelegt. Der Prozess, mit dessen Hilfe das Datenmaterial exzerpiert wird, bleibt «im Hinterkopf» und ist daher Schwankungen ausgesetzt.

Zur Umsetzung dieses Konzepts dient ganz offensichtlich die mittlerweile nahezu «automatisch» verfügbare Software, die bei neu beschafften Rechnern meist mit der Grundausstattung mit installiert wird. (In der Regel also Microsoft Access oder Filemaker, eventuell auch Microsoft Excel, wobei offen bleiben soll, ob letzteres – eigentlich eine Tabellenkalkulation – wirklich als «Datenbanksoftware» bezeichnet werden kann.)

Der Nutzen dieses Ansatzes ist offensichtlich: Der Aufwand ist minimal, es besteht keinerlei Notwendigkeit, den eigenen methodischen Ansatz überdenken zu müssen. In der Tat hat dieser Ansatz aber sogar für den überzeugten Vertreter der grundsätzlichen Wünschbarkeit der Beschäftigung mit rechnergestützten Methoden in der Geschichte etwas Bestechendes: Eben durch den minimalen Aufwand bei der Verwendung «automatisch» vorhandener Software findet für viele Historiker durch die Verwaltung der eigenen Notizen und/oder Bibliographien oft ein unbemerkter Einstieg in die Informationstechnologien statt, der den potentiellen Interessentenkreis für anspruchsvollere Verfahren laufend erhöht.

Die Nachteile dieses Ansatzes sind je nach Blickwinkel gering oder fundamental: Die gesammelten Daten eignen sich genau so wenig zur Weitergabe oder zur langfristigen Verwendung wie hastig angelegte Notizzettel, da zahllose undokumentierte Annahmen in sie eingehen, die, sobald sie nicht mehr täglich verwendet werden, auch dem Datenproduzenten meist nicht mehr transparent sind. Das ist ein fundamentales und ernsthaftes Problem, wenn die Illusion besteht, dass derart rasch und ohne längeres Nachdenken angelegte Sammlungen eine Zukunftsinvestition darstellen. Werden Datenbanken dieser Art als das verstanden was sie sind, nämlich «Schmierzettel» auf technologisch anspruchsvollerem Niveau, entfällt dieses Problem.

2. Statistische Datensammlungen

Statistische Datensammlungen, auch wenn sie ohne analytischen Anspruch, schlicht zur besseren Darstellung von unübersichtlichen Details mit den Mitteln der deskriptiven Statistik, begonnen werden, sollten das Ergebnis klarer methodischer Reflexion sein. Quellengruppen, bei denen Einzelaussagen an sich wenig Aussagekraft haben, wie einzelne Preis- oder Altersangaben, setzen nun einmal ein bewusst anderes Vorgehen voraus als Quellengruppen, bei denen der einzelne Beleg einen in sich diskutierbaren Sachverhalt darstellt. Die Praxis zeigt freilich, dass gerade die Leichtigkeit, mit der moderne Statistik- und Grafiksoftware die Erstellung von auf

quantitativen Merkmalen aufbauenden Grafiken möglich macht, hier oft zu Verschleierungseffekten führt. Wer es schwierig findet, in tabellarischer Form wiedergegebene quantitative Sachverhalte zu interpretieren, sollte ernsthaft prüfen, ob die plausibel erscheinenden daraus abgeleiteten Grafiken ihn – oder sie – nicht in die Irre führen.

Das Medium für die Realisierung dieses Ansatzes ist die leicht zugängliche Statistiksoftware – nach wie vor wohl vor allem SPSS.

Vorteil und Nachteil dieses Ansatzes sind implizit bereits in der einleitenden Diskussion beschrieben worden. Wenn die statistischen Kunstlehren[10] ernst genommen werden, entsteht echter methodischer Mehrwert. Und selbst abgesehen davon: Manche Arten historischer Quellen sind nur quantitativ zu interpretieren. Werden sie dies nicht, entstehen Artefakte. Artefakte entstehen freilich auch, wenn der Enthusiasmus für das neu entdeckte statistische Rüstzeug dazu verleitet zu glauben, man könne dadurch auf die Quellenkritik verzichten. Noch in Volkszählungslisten der USA im 19. Jahrhundert, also in Quellen, die durch eine «zahlenbewusste», moderne Bürokratie angelegt wurden, schlagen manche Forscher mittlerweile vor, alle auf «Null» oder «Fünf» endenden Altersangaben als potentielle Datenfehler zu betrachten.[11] Was recht plausibel ist, wenn man bedenkt, dass eine der bestdokumentierten Untersuchungen ergeben hat, dass im benachbarten Kanada die Altersangaben von 23 % aller eindeutig in zwei durch zehn Jahre getrennten Volkszählungen in diesen beiden Volkszählungen um mehr als drei Jahre differieren.[12] Dies ebenfalls im 19. Jahrhundert: Wie sinnvoll unter diesen Umständen die Interpretation von nach-Komma-Differenzen in Indikatoren der historischen Demographie für frühneuzeitliche Gemeinschaften ist, möge jeder selbst entscheiden.

3. Protostatistische Datenbanken

Während die beiden vorigen, durch uns eingeführten Kategorien auf unmittelbare Wiedererkennung hoffen durften, ist dieser Begriff wahrscheinlich erklärungsbedürftig. Wir verstehen darunter alle jene Vorgehensweisen, bei denen Daten in eine Datenbank eingegeben werden, um Ziele zu erreichen, die an sich denen quantitativen Vorgehens gleichen – also z.B. um in sich nicht aussagekräftige Einzelbelege so ordnen zu

10 Wohl immer noch empfehlenswert: Thome, Helmut: *Grundkurs Statistik für Historiker*. Bd. I und II, Köln 1989/1990 (= Historische Sozialforschung Beiheft 1 und 2).

11 Guest, A.M.: «Notes from the National Panel Study: Linkage and Migration in the Late Nineteenth Century». In: *Historical Methods*, vol. 20, 1987, S. 63 - 77, hier: S. 67.

12 Pueyez, C.; Roy, R.; Martin, F.: «The Linkage of Census Name Data: Problems and Procedures». In: *Journal of Interdisciplinary History*, vol. 14, 1983, S. 129 - 152, hier: S. 147.

können, dass sie interpretierbar werden oder um eine sehr grosse Anzahl von Belegen komfortabel verwalten zu können. Für viele Leser mag in diesem Zusammenhang das Epitheton «protostatistisch» als überflüssig erscheinen, da diese beiden Ziele – Ordnen und der raschen Suche zugänglich machen – schliesslich als Motivenbericht für den Einsatz von Datenbanksystemen in der Geschichtsforschung überhaupt stehen können und von uns, leicht abgewandelt, weiter oben auch eben dafür verwendet wurden.

Wenn wir auf diesem Epitheton beharren, so deshalb, weil wir hier eine erhebliche potentielle Gefahr aufzeigen wollen. Der Übergang zwischen einer privaten Exzerptsammlung, mit der wir unsere Typologie begonnen haben, und einer Datenbank, die gezielt für die Nutzung durch andere erstellt wird, ist fliessend – in der Tat liegt der einzige wirkliche Unterschied in dem Aufwand, der getrieben wird, um die Entscheidungen, die bei der Bildung der Datenbankkategorien oder Felder getroffen werden, transparent zu machen und permanent zu dokumentieren. (Demgegenüber ist der technische Unterschied, dass bei Datenbanken, die sich an ein weiteres Publikum wenden, manchmal mächtigere Instrumente als Microsoft Access oder Filemaker, z.B. Oracle, zum Einsatz kommen, gering: zumal mit der steigenden Rechnerleistung die Notwendigkeit dazu potentiell sinkt.)

Wie wir bei der Entwicklung unserer Beschreibungskategorien – visuelle Eigenschaften, Semantik und Syntax einer Quelle – zu zeigen versuchten, sind die tatsächlichen konzeptuellen Unterschiede zwischen statistischen Datensätzen und «Faktendatenbanken», wie sie traditionell mit Microsoft Access oder Filemaker verwaltet werden, sehr gering. Sie *erscheinen* jedoch vielen als fundamental und so manche Datenbank wurde der Erstellung eines statistischen Datensatzes vorgezogen, weil man glaubte, sich dadurch die saubere Kategorienbildung ersparen zu können – worauf die Unübersichtlichkeit gezählter Listen dann nachträglich zu mehr oder minder willkürlichen Zusammenfassungen zwang. Die grosse Gefahr derartiger Illusionen über die Vermeidbarkeit sauberen konzeptuellen Arbeitens bei der Datenbankkonstruktion liegt also in einer Tendenz zu methodisch unreflektierter Pseudostatistik.

4. Quellenorientierte Datenverarbeitung[13]

Diese Vorgehensweise beim Einsatz von Datenbanken wird manchmal eng mit dem Verfasser und einem von ihm geschriebenen Softwarepaket[14] identifiziert. Es ist richtig, dass dieses Softwarepaket explizit zur Unterstützung quellenorientierten Vorgehens geschrieben wurde; richtig ist auch, dass es als Mittel verwendet wurde, um bestimmte Konzepte innerhalb eines quellenorientierten Paradigmas direkt zu implementieren und zu testen[15]. So ehrend es ist, ist es aber trotzdem falsch, ihn als den Erfinder der dadurch zusammengefassten Konzepte zu sehen. Diese entstammen vielmehr einem Unbehagen über Probleme bei der Umsetzung historischer Quellen in statistische Datensätze in den siebziger Jahren einer-[16] und dem frühen Bestreben, den gesamten Informationsgehalt historischer Quellen verarbeitbar zu machen, ohne bei der Eingabe Vorentscheidungen darüber zu treffen, welchem Analysezweck die Dateneingabe dienen solle, andererseits[17].

Letztere Charakterisierung gibt auch jetzt noch den besten Einstieg in das Ziel der Technik: Angestrebt wird eine *bewusste* Abbildung des Informationsgehaltes einer Quelle als solcher in ein rechnergestütztes Medium. Die Semantik der Quelle wird *bewusst* transformiert und meist selbst zum

---

13 Vgl. in diesem Band: Grassi, Giulio: «Prinzipien der quellennahen Datenverarbeitung exemplifiziert an einem Forschungsprojekt des Schweizerischen Nationalfonds». S. 39 - 58. – Bosshard, Werner; Malamud, Sibylle; Sutter, Pascale: «Komplexe sozialhistorische Abfragen in modularen Datenbanken – ein praxisnaher Werkstattbericht. Das Nationalfondsprojekt ‹Soziale Beziehungen im Alltag einer spätmittelalterlichen Stadt, Zürich im 15. Jh. › unter der Leitung von Prof. Dr. Hans-Jörg Gilomen». S. 59 - 78. – Gigandet, Cyrille: «Dans la toile des Lumières. Modèle d'analyse des correspondants suisses de Jean Henri Samuel Formey». S. 123 - 154. – Dietrich, Urs: «Die Datenbank EuroClimHist im Wandel». S. 155 - 166.

14 Thaller, Manfred: Kleio. A Database System. St. Katharinen 1993 (= Halbgraue Reihe zur Historischen Fachinformatik B 11).

15 Vgl. z.B. Thaller, Manfred: «Databases and Expert Systems as Complementary Tools for Historical research». In: *Tijdschrift voor Geschiedenis*. Bd. 103, 1990; mit einer darauf fussenden Untersuchung zur Umsetzung durch Levermann, Wolfgang: *Kontextsensitive Datenverwaltung*. St. Katharinen 1991 (= *Halbgraue Reihe zur Historischen Fachinformatik* Bd. B8).

16 Z.B.: Overton, Mark: «Computer Analysis of an Inconsistent Data Source: the Case of Probate Inventories». In: *Journal of Historical Geography*. Bd. 3, 1977, S. 317 - 326 oder Ginter, Donald E. et al.: «A Review of Optimal Input Methods: Fixed Field, Free Field, and the Edited Text». In: *Historical Methods Newsletter*. Bd.10, 1977, S. 166 - 171.

17 Z.B.: King, Tim J.: «The Use of Computers for Storing Records in Historical Research». In: *Historical Methods*. Vol. 14, 1981, S. 59 - 64 oder Baum, H.P.; Sprandel, Rolf: «Die Erforschung von Lehensregistern in Verbindung mit der EDV. Ein sozialgeschichtliches Projekt an der Universität Würzburg». In: *Jahrbuch der historischen Forschung in der Bundesrepublik Deutschland, Berichtsjahr 1980*. Stuttgart 1981, S. 49 - 55.

Forschungsgegenstand. Die Auswertungsziele bleiben zu Projektbeginn *bewusst* offen. Im Prinzip könnte man die quellenorientierte Datenverarbeitung also einfach als eine Form des Einsatzes der Datenbanktechnologie betrachten, die versucht, die Quellenkritik besonders ernst zu nehmen. Dementsprechend kann sie prinzipiell mit gängiger Software implementiert werden, setzt wegen der Eigenart historischer Datenstrukturen und/oder der Semantik des Quelleninhalts jedoch meist ad hoc programmierte Komponenten oder dedizierte Systeme voraus.

Diesen Prämissen entsprechend bietet sie offensichtliche Vorteile: Die Ergebnisse sind präzise und inhaltlich «sauber». Das auf den Rechner gebrachte Material bleibt langfristig verwertbar. Der Prozess der Formalisierung wird voll dokumentiert. Auch sehr hohe quellenkritische Ansprüche werden befriedigt.

All dies bringt freilich Kosten mit sich: Der Arbeitsaufwand ist hoch, ein langer Vorlauf steht vor den ersten präsentablen Ergebnissen. Da die visuellen Charakteristika der Quelle meist verloren gehen, und die Datenbanken nicht geeignet sind, das vertraute Bild einer edierten Quelle wiederzugeben, wird die Aufbereitungsarbeit im Fach nicht als genuin wissenschaftliche Leistung anerkannt.

Erschwerend kommt hinzu, dass detailbezogene Quellenkritik und zeitgemässe Arbeitstechniken im postmodernen Zeitalter Öffentlichkeiten ansprechen, die miteinander meist ohnehin im Konflikt liegen.

5. Digitale Editionen[18]

Das Ziel derartiger Projekte besteht in der Präsentation einer Quelle auf einem digitalen Medium in einer Form, die den Ansprüchen der Editionswissenschaften gerecht wird, wobei der Informationsgehalt (in unterschiedlichem Ausmass) den Kriterien quellenorientierter Ansätze entspricht. Eine Einbindung der visuellen Wiedergabe der Originalquelle wird meist angestrebt. (Wir verstehen hier unter einer «digitalen Edition» eine, die gezielt versucht, die Informationstechnologien einzusetzen, um Leistungen zu erbringen, die durch gedruckte Editionen *nicht* erbracht werden können. Gedruckte Editionen, die den Rechner bloss zur Realisierung des Druckbildes verwenden, fallen u. E. aus dem hier diskutierten Bereich völlig heraus.)

18 Vgl. in diesem Band: Kropač, Ingo H.; Boshof, Heidrun: «Digitale Edition eines umfangreichen Quellenkorpus: Vorgehensweise und Probleme bei der Aufbereitung, Strukturierung und Kategorisierung des Quellenmaterials». S. 93 - 112.

Realisiert werden derartige Projekte heute meist auf der Basis einer SGML- oder XML-orientierten Auszeichnung, während frühere Projekte meist eigene Auszeichnungssysteme entwickelt haben. Der Unterschied ist freilich geringer als man annehmen möchte: Für die eigentliche Datenbankfunktionalität, jenseits der Präsentation in einem bekannten, also standardisierten Browser, sind nach wie vor meist Eigenentwicklungen nötig.

Der Vorteil dieses Ansatzes besteht vor allem in der hervorragenden Einpassung in die editorische Tradition des Faches und der hohen Akzeptanz, die er daher oft findet. Auf Grund der höheren Effektivität ergibt sich eine unmittelbare Beschleunigung[19] der primären editorischen Aufgaben, mit Ergebnissen, die viele der Vorteile quellenorientierter Arbeit übernimmt.

Eine Konsequenz dieses Ansatzes kann als abschreckendes Problem oder als zusätzlicher Vorzug angesehen werden: Projekte dieser Art sind fast immer so arbeitsaufwändig, dass sie von der inhaltlichen Forschung letztlich abgelöst sind, sich also auch in einem Punkt in die Tradition editorisch-hilfswissenschaftlicher Arbeit einfügen, der von manchem als abschreckend betrachtet wird.

6. Digitale Archive/Digitale Bibliotheken[20]

Schliesslich haben die letzten Jahre die Entstehung sehr grosser Datenbanken gesehen, die für die historischen Wissenschaften von grosser Relevanz sind, nämlich Datenbanken, die gezielt versuchen, ganze Sammlungen – geschlossene Bestände von Archiven[21] und/oder Bibliotheken[22] – in

---

19 Ein äusserst eindrucksvolles Beispiel bietet hier *Wittgenstein's Nachlass: The Bergen Electronic Edition.* Oxford etc. 1998 - 2000. Als erste substantielle vollelektronische Edition eines grösseren Korpus, von 20'000 teils hand-, teils maschinschriftlichen Seiten, d.h., *ohne* eine Druckfassung, aber als vollständige Transkription mit den elektronischen Äquivalenten kritischer Apparate, begleitet von einer kompletten photographischen Wiedergabe, wurde das Werk in zehn Jahren (im Einklang mit dem Arbeitsplan !) abgeschlossen.

20 Vgl. dazu Froidevaux, Yves; Bär, Peter; Ritter, Gerold (Hgg.): *Informatik im Archiv – L'Informatique des archives.* Geschichte und Informatik – Histoire et Informatique, Vol. 10, 1999.

21 Vgl. z.B. http://www.archive.geschichte.mpg.de/duderstadt/dud.htm und dazu: Ebeling, Hans-Heinrich; Thaller, Manfred: *Digitale Archive. Die Erschließung und Digitalisierung des Stadtarchivs Duderstadt.* Göttingen, 1999. Vgl. auch http://memory.loc.gov/ammem/gwhtml/.

22 Vgl. z.B. http://www.mpier.uni-frankfurt.de/dlib/, http://gallica.bnf.fr/ oder http://moa.umdl.umich.edu/index.html. Leider ist der ganze Bereich der digitalen Ressourcen derzeit von einer aus finanzierungstrategischen Gründen völlig überzogenen Debatte um technische Standards und Empfehlungen gekennzeichnet, deren Zahl die der tatsächlich realisierten signifikanten digitalen Ressourcen mittlerweile bei weitem übersteigt. Relativ nahe an der Realität bleiben: Carpenter, Leona; Shaw, Simon; Prescott, Andrew (Hgg.): *Towards the Digital Library.* London 1998.

digitaler Form bereitzustellen, meist mit der Absicht, das Material im Internet greifbar zu machen. Das Ziel besteht explizit nicht in der Bereitstellung kritischer Editionen, sondern in der möglichst grosser Mengen an Rohinformationen. «Zehnmal mehr, als durch eine Edition bearbeitbar, zehnmal besser erschlossen, als durch herkömmliche Findbücher oder Kataloge.»

Als Medium dafür dienen im Wesentlichen die gleichen Instrumente, die im vorigen Punkt beschrieben wurden: also SGML- und XML-kodierte Datensammlungen, hinter denen spezialisierte Datenbanken, sogenannte *Document Management Systems* stehen.

Über die Attraktivität der grundsätzlichen Vision, das Archivmaterial für eine historische Arbeit in qualitativ hochstehenden, digital-photographischen Reproduktionen ohne Zeitverzug vom eigenen Schreibtisch aus einsehen zu können, muss man wahrscheinlich nur wenige Worte verlieren. Trotzdem sei darauf hingewiesen, dass dieser Entwicklung, abgesehen von der unmittelbaren instrumentellen Attraktivität, auch methodisch einiger Stellenwert zukommen könnte: Wenn *alle* Dokumente und Bücher gleichermassen zugänglich sind, entfällt die praktische Notwendigkeit jener Konzentration auf eine relativ beschränkte Zahl gut edierter und/oder zugänglicher Texte, die in vielen historischen Disziplinen auch heute noch zu beobachten ist. Eine derartige «Dekanonisierung der Geisteswissenschaften» kann weitreichende Folgen haben.

Dass die Anlage von Dokumentationssystemen mit mehreren hunderttausend Seiten Inhalt dedizierte Projekte voraussetzt, dürfte offensichtlich sein; dass sie nicht als Bestandteil primär persönlichem Forschungsinteresse entspringender Arbeit verwirklicht werden können, kann daher nicht als Gefahr betrachtet werden. Eben deshalb, weil es sich hier um unmissverständliche Dienstleistungen handelt, die mit inhaltlichen Ergebnissen nicht verwechselt werden können, scheint dem Verfasser auch der Vorwurf positivistischer Faktenhuberei diesen Systemen nicht angemessen. Mit gleichem Recht könnte man vorschlagen, Archive sollten ihre Findbücher durch theoriebasierte Essays ersetzen, mit dem man digitalen Archiven vorwirft, ihr Material nicht kritisch zu analysieren. Eine vage, wenn im Moment auch durchaus lebhafte Diskussion um eine Entfremdung des Historikers vom Original und um den Verlust der haptischen Dimension der Quellenarbeit sei anerkannt. Dem Verfasser scheint jedoch, dass hier übersehen wird, dass es ihm nicht um das Verbot des Zugangs zu den zentralen Originalen einer Arbeit geht, sondern um den Zugang zu jenem Material, den er derzeit aus praktisch-organisatorischen Gründen schlechterdings nicht hat.

Eine Gefahr ist jedoch unübersehbar: In vielen europäischen Ländern stehen die Digitalisierungsprogramme völlig unter Kontrolle der bewahrenden Einrichtungen – Archive, Bibliotheken und Museen. Bleibt dies so, verringert sich der Einfluss der Disziplin darauf, welches Material wie zugänglich gemacht wird, den sie in den derzeitigen editorischen Gremien noch hat.

IV. Zusammenfassung

Die Frage nach den Vorteilen, die historische Datenbanken der Forschung bringen und den Problemen, die sie aufwerfen, schien uns nicht direkt beantwortbar, weil das Phänomen historischer Datenbanken zu uneinheitlich ist, um generelle Antworten zu erlauben. Nachdem wir versucht haben, die unterschiedlichen Erscheinungsformen zu beschreiben, möchten wir dies jedoch in Thesenform nachholen.

Vier Vorteile scheinen uns offensichtlich:

1. Die mühsame Kärrnerarbeit wird durch den Einsatz von Datenbanken auf vielen Ebenen einfacher.
2. Dies geschieht im Bereich moderner Anwendungssoftware mittlerweile so selbstverständlich, dass Datenbanken oft gar nicht mehr als solche erkannt werden.
3. Gleichzeitig gibt es genügend hinreichend weit gediehene Projekte, um zu zeigen, dass methodische Fortschritte bei Durchdringung und Integration des Materials möglich sind: Anspruch der Intersubjektivität.
4. Selbst wenn davon ausgegangen wird, dass die Geisteswissenschaften keine Wissenschaften, sondern postmodernistische Darstellungskünste sind, bleibt eine radikale Verbesserung des Zuganges zur Information.

Dem stehen aus unserer Sicht vier Gefahren gegenüber:

1. Die Einfachheit des Einstiegs in moderne Software verführt dazu, den intellektuellen Aufwand zu unterschätzen, der bei der Übernahme von Informationen in Datenbanken zu leisten ist. Dadurch entsteht die Gefahr von Scheinpräzision.
2. Die Erstellung einer Datenbank nach einem der methodisch anspruchsvolleren Paradigmata stellt die gleichen methodisch-intellektuellen Ansprüche wie die Anlage einer klassischen Edition – ohne jedoch in gleicher Weise Anerkennung zu finden.
3. Die erste dieser beiden Gefahren verstärkt die zweite: Einfache Lösungen machen wegen ihrer Einfachheit den Aufwand für komplexere

unglaubwürdig, und der von ihnen ausgehende Ruch schlampiger Quellenarbeit diskreditiert die Leistungen, die in die anspruchvolleren Ergebnisse investiert werden müssen.

4. Werden Geisteswissenschaftler zu passiven Konsumenten von durch andere erstellten fachwissenschaftlichen Informationssystemen?