

# Historische Datenverarbeitung und Landesgeschichte

Micha Röhring

## 1. Vorbemerkung

„Historische Datenverarbeitung“ bezeichnet alle Verfahren der automatisierten, d. h. in der Regel computergestützten Verarbeitung historischer Daten und Informationen. Historische Datenverarbeitung ist keine Hilfswissenschaft der Geschichtsforschung, sie ist ein Werkzeug – keine Methode, sondern eine Technik.<sup>1</sup>

Bewußt wird mit dem Titel an den Aufsatz, den Erwin Riedenauer über die elektronische Datenverarbeitung in der Landes- und Gesellschaftsgeschichte geschrieben hat, angeknüpft.<sup>2</sup> Bei diesem handelt es sich um einen der wenigen Texte über die Anwendung der Historischen Datenverarbeitung in der Landesgeschichte, der für den Interessierten leicht zu erreichen ist. Neben vielen auch heute noch sehr aktuellen Überlegungen referiert Riedenauer darin aber auch einen inzwischen vollständig überholten Stand der Technik. Im Vergleich stellt man dann fest, wie stark sich der Anwendungshorizont der Historischen Datenverarbeitung verändert und erweitert hat. Merkwürdigerweise gibt es daneben kaum Texte, die sich ausdrücklich mit der Anwendung der Datenverarbeitung im Rahmen der Landesgeschichte befassen. Zwar werden Anwendungen beschrieben, die im Bereich der Landesgeschichte anzusiedeln sind, einen Überblick über die Möglichkeiten aber gibt es kaum.

Da hier davon ausgegangen wird, daß die Mehrzahl der Leser mit den Techniken der Datenverarbeitung in der Geschichtswissenschaft kaum vertraut ist und die Fachliteratur zur Historischen Datenverarbeitung oftmals ein großes Eigeninteresse des Lesers voraussetzt, sollen hier zunächst

- die Techniken der Historischen Datenverarbeitung dargestellt und daran anschließend (einer bewährten Methode folgend)
- anhand von konkreten Projekten ihr Einsatz illustriert werden. Darüber hinaus führen dann
- Überlegungen zum Einsatz der Historischen Datenverarbeitung in der Landesgeschichte und
- Forderungen für zukünftige Entwicklungen.

In der folgenden Darstellung überwiegen Konzepte und Modelle gegenüber konkreten technischen Lösungen. Aussagen über Möglichkeiten und Grenzen, technische Details oder auch Unzulänglichkeiten existierender Systeme sind für Rezensionen oder Aufsätze in EDV-Zeitschriften geeignet. Vor diesem Hintergrund erscheint es wenig sinnvoll, kurzlebige technische Konzepte vorzustellen. Der Schwerpunkt liegt auf längerfristig angelegten theoretischen Konzepten, die mit den Bedürfnissen der Geschichtswissenschaft zusammenhängen.

## 2. Motivation

Wozu eigentlich den Computer verwenden? Grundsätzlich bleibt festzustellen, daß seine Verwendung kein Ausweis besonderer Wissenschaftlichkeit ist. Die Geschichtswissenschaft hat in ausreichendem Maß Fragen, die sich sehr gut auf traditionelle Weise lösen lassen. Computer arbeiten sicherlich sehr schnell und können einem viel Arbeit abnehmen. Im Gegensatz zu mathematischen Problemen, wo drei oder vier Zahlenpaare und vielleicht hundert Zeilen Programmtext einen Rechner stunden- oder tagelang beschäftigen können, benötigt man zur Lösung historischer Fragestellungen in der Regel sehr viele Daten, die man – auch das ist der Regelfall – der Maschine erst durch viel Handarbeit präsentieren muß.

Computereinsatz beginnt sich dann zu lohnen, wenn eine größere Menge von Daten mehrfach benötigt wird. Sei es, daß man nach unterschiedlichen

Kriterien sortierte Listen erzeugen will oder einerseits eine statistische Auswertung, andererseits eine Aufbereitung für den Druck plant. Ebenso kann man – und das ist ein wichtiges Ziel – eine Quelle durch verschiedene Personen zu unterschiedlichen Zwecken auswerten lassen.

Frühneuzeitliche Stadtrechnungen etwa kann man geschlossen betrachten, um eine Frage zum städtischen Haushalt zu klären, man kann gezielt die Angaben zu Zünften, Wehrwesen, Baumaßnahmen u. a. herausfiltern, man kann Listen der erwähnten Personen erstellen und die Quelle zu einer gedruckten Edition aufbereiten. Diese technischen Möglichkeiten müssen dabei natürlich von einer konkreten Fragestellung geleitet sein, dürfen nicht Selbstzweck werden.

Gerade in der Tatsache, daß man eine maschinenlesbare Quelle von ganz unterschiedlicher Seite betrachten kann, ohne ihre eigentliche Substanz zu verändern (die Erzeugung einer Druckvorlage etwa verändert nicht die Originaldaten, sondern ergibt eine Kopie mit den nötigen Zusätzen für den Druck – Fußnoten, Seitenzahlen, Überschriften usw.), scheint ein wesentlicher Vorteil zu sein.

Aus dieser zunächst sehr einfach und naheliegend scheinenden Idee, den Computer zum Suchen, Sortieren, Drucken, Korrigieren, Rechnen und anderem einzusetzen, erwachsen bei der Anwendung auf historische Phänomene einige Komplikationen. Diese sind in der Art der Daten begründet. Wahrscheinlichkeiten und Unwägbarkeiten, Vermutungen über mögliche Zusammenhänge (A *könnte* das Kind von B sein – oder auch nicht) fallen unter das Stichwort „Unschärfe“, ein in der Datenverarbeitung noch nicht sehr weit entwickelter Terminus. Scheinbar triviale Fragen wie etwa: *Gib mir alle Personen zwischen 18 und 25 Jahren, die aus Preußen stammen*, angewandt auf Datenbestände vergangener Jahrhunderte, implizieren Schwierigkeiten, die bestenfalls mit den ausgefeiltesten Methoden der Künstlichen Intelligenz „auf Knopfdruck“ zu lösen sind.<sup>3</sup> Zur Lösung dieser und anderer Schwierigkeiten hat sich das Arbeitsgebiet der Historischen Datenverarbeitung herausgebildet, nicht als Selbstzweck, sondern in Anlehnung an konkrete historische Projekte und Untersuchungen.

### 3. Konzepte der Historischen Datenverarbeitung

In der Historischen Datenverarbeitung wurden sehr unterschiedliche Konzepte entwickelt, die an die jeweiligen Fragestellungen angepaßt sind. Zu einem erheblichen Teil spiegeln sie aber auch die technische Entwicklung und insbesondere auch die Erweiterung der Theorie der Datenverarbeitung von rein numerischen Problemen hin zu text-, bild- und objektorientierten Verfahren wider. Im Folgenden soll eine Übersicht über die bislang entwickelten Techniken und Anwendungen gegeben werden.<sup>4</sup>

Etwas systematisierend – und damit vereinfachend – werden folgende Punkte angeschnitten:

1. quantitative Methoden und Statistik,
2. datenbankorientierte Verfahren,
3. Objekte als Quellen historischer Forschung.

Für die gewählte Grenzziehung – die allerdings zu einem gewissen Grad willkürlich bleibt – sprechen sowohl methodisch historische Gründe als auch

technische Unterschiede der jeweils primär eingesetzten Datenverarbeitungstechniken.

### 3.1. Quantitative Methoden/Statistik

Entsprechend seinem Namen war es nur natürlich, den *Rechner* zunächst für vorwiegend numerische Probleme einzusetzen, zumal die Theorie der Datenverarbeitung und die technischen Möglichkeiten der Verspeicherung von Daten bis vor gar nicht so langer Zeit kaum etwas anderes zuließen.

Während Riedenaier noch der Meinung war, die Verwendung statistischer Methoden würde sich auf wenige, grundlegende Maßzahlen der deskriptiven Statistik (Mittelwert, Median, ...) beschränken und er die Hauptanwendung in der Registererstellung sah<sup>5</sup>, bahnte sich von Frankreich und den angelsächsischen Ländern ausgehend eine völlig andere Entwicklung an. Mit dem Vordringen der Historischen Sozialforschung, die auf einen ausgefeilten Methodenapparat setzt und eine enge Theoriebindung anstrebt<sup>6</sup>, wurden die Fragestellungen und die statistischen Modelle zunehmend komplexer. Von der Preis- und Bevölkerungsentwicklung über die inzwischen weit ausgebauten Modelle zur historischen Demographie, zu kollektiven Biographien und historischer Wahlanalyse führte der Weg.<sup>7</sup> Die Entwicklung der statistischen Methoden und ihrer Anwendung auf historische Problemstellungen lief schneller ab, als die Ausbildung und Umsetzung von Konzepten zur Registererstellung. In jedem Fall hatte der Einsatz der Datentechnik dort, wo es um die Bewältigung von Zahlenmaterial ging, sehr tiefgreifenden Einfluß auf die historische Forschung.

Mit Hilfe quantitativer Methoden sollen historische Prozesse und Strukturen *auf der Basis von Wahrscheinlichkeitsannahmen auf ihre Regelmäßigkeit hin überprüft*<sup>8</sup> werden. Die Statistik wird dabei eingesetzt, um *numerische Daten zu verdichten, zu ordnen, zu gruppieren und zu veranschaulichen*<sup>9</sup>. Nicht notwendigerweise benötigt man hierzu Techniken der Datenverarbeitung. In vielen historischen Arbeiten werden Methoden der deskriptiven Statistik angewandt, ohne daß der Forscher sich als „Quantifizierer“ bezeichnen würde oder mehr als einen Taschenrechner eingesetzt hätte. Da sich aber quantitative Methoden im Umfeld der Historischen Sozialforschung, die nicht das einzelne Individuum, sondern eine Gesamtheit von Individuen betrachtet, konzentrieren, liegt es nahe, den Computer zur Bewältigung von Massendaten einzusetzen.

Insbesondere wird dies hilfreich, wenn man mehr als nur Mittelwerte und ähnlich einfache Kennzahlen ermitteln will, die oftmals die Wirklichkeit mehr verschleiern als Erkenntnis vermitteln können. Im Bereich des „Testens“ von Hypothesen ist es erforderlich, den Zusammenhang zwischen unterschiedlichen Variablen zu untersuchen (z. B. Industrialisierungsgrad einer- und Konfessionen, Einkommensniveaus oder Wahlergebnisse andererseits).

Zur statistischen Analyse werden meist sogenannte Statistikpakete eingesetzt. Dabei handelt es sich um Sammlungen von Programmen, die auf die Lösung statistischer Aufgaben zugeschnitten sind. In der Historischen Sozialforschung haben sich vor allem die beiden Programmpakete SPSS (Statistical Package for the Social Sciences) und SAS (Statistical Analysis System) etabliert.<sup>10</sup> Sie erlauben die Errechnung nahezu aller in Frage kommender statistischer Maßzahlen ohne nennenswerten Programmieraufwand. Ihre volle Ausnutzung setzt allerdings fundierte Kenntnis der erforderlichen statistischen

Verfahren und insbesondere auch ihrer Aussagekraft voraus, da ansonsten das numerische Ergebnis nicht interpretiert werden kann.

Man darf nicht aus den Augen verlieren, daß Quantifizierung nicht gleichbedeutend mit Computereinsatz ist. Genausowenig ist der Schluß erlaubt, Historische Datenverarbeitung sei gleichbedeutend mit der Anwendung quantifizierender Methoden. Es sei hier einmal die Prognose gewagt, daß diese in der Vergangenheit sicher häufiger zutreffende Vermutung in Zukunft immer weniger Berechtigung haben wird.

Öfters wird kritisiert, daß man mit Statistik „alles beweisen“ könne. Diese Kritik kann in der Regel zwei Ursachen haben: entweder eine – oft mehr emotionale – Ablehnung des Verfahrens als solches, meist begleitet von Unkenntnis der quantitativen Methodik oder – und das muß man als gravierender betrachten – ihre unqualifizierte und unreflektierte Anwendung. Beim Einsatz von quantitativen Methoden ist eine fundierte Kenntnis ihrer Reichweite notwendig. Die Kritik an Ergebnissen quantitativer Forschung muß bei korrekter statistischer Auswertung bei der vorher notwendigen Begriffs- und Theoriebildung einsetzen. Wer glaubt, die Statistik liefere stets naturwissenschaftlich exakte Ergebnisse, sitzt einem fatalen Irrtum auf. Ihrer Einbindung in den Kontext der historischen Forschung, insbesondere einer geschichtswissenschaftlichen Begriffsbestimmung, ist größte Aufmerksamkeit zu widmen.

### **3.2. Datenbankorientierte Verfahren**

Der rasante Aufschwung, den die Datenbanktechnik seit Beginn der 80er Jahre genommen hat, hat auch die Historische Datenverarbeitung grundlegend beeinflußt. Es ist abzusehen, daß Datenbanksysteme (welcher Art auch immer) zunehmend zum Kernstück der Datenverarbeitung in den Geschichtswissenschaften werden.

Ein Datenbanksystem (DBS) bietet die Möglichkeit, unterschiedliche Datenbestände für unterschiedliche Anforderungen zu bearbeiten. Es stellt ein Beschreibungsmodell für die Daten, das sogenannte Datenmodell bereit, innerhalb dessen die Selektion und Weiterverarbeitung der Daten erfolgt. Ein DBS ist meist ein sehr komplexes und anwendungsmächtiges Programmpaket. Dabei gestaltet sich seine Anwendung in der Praxis wesentlich einfacher und ökonomischer, als man vielleicht zunächst erwarten würde. Zwischen dem Anwender und den tatsächlich vom Computer verspeicherten Daten liegen mehrere Abstraktionsebenen, die bewirken sollen, daß auf der einen Seite eine möglichst effiziente Verspeicherung und auf der anderen Seite eine möglichst menschliche Darstellung der Daten gewährleistet ist. Die Summe der Elemente, die die dafür erforderlichen Transformationen durchführen, ist das DBS.<sup>11</sup>

Eine konkrete, im Kern auf ein Datenbanksystem gestützte, Anwendung dieser vielfältigen Möglichkeiten stellt die „Historical Workstation“ dar. Unter diesem Titel<sup>12</sup> soll ein Konzept verstanden werden, das einen speziell an den Bedürfnissen eines Historikers ausgerichteten Computerarbeitsplatz definiert. Die Historical Workstation setzt sich zusammen aus folgenden Komponenten:

- Einem Datenbanksystem, das den speziellen Erfordernissen historischer Daten Rechnung trägt;

- Datenbanken mit Hintergrundwissen aus den historischen Hilfswissenschaften (Chronologie, Maßsysteme, Währungen ...);
- Datenbanken mit historischem Quellenmaterial;
- einem System der Künstlichen Intelligenz (KI) zur transparenten Verwaltung dieser Teilkomponenten;
- einer Schnittstelle zu einem Satzsystem (für Druckvorlagen);
- einer Schnittstelle zu Statistiksystemen;
- einer Verbindung zwischen transkribierten (d. h. byte-kodierten) Texten und digitalisierten Abbildungen der originalen Quellen.<sup>13</sup>

Zusätzlich zu den von Thaller angeführten Komponenten sollte man in die Liste noch das Satzsystem und die statistische Software sowie zusätzliche Komponenten, etwa für linguistische oder philologische Analyse, aufnehmen.<sup>14</sup>

Man könnte diese Zusammenstellung auf den ersten Blick für Science Fiction halten. Sieht man aber einmal von dem KI-Subsystem ab (an dessen Realisierung gearbeitet wird), so sind alle genannten Komponenten verfügbar. Das Datenbanksystem  $\kappa\lambda\epsilon\iota\omega$  (Kleio) verfügt über eine Schnittstelle zu den Statistikpaketen SAS, SPSS und Cens-Sys.  $\kappa\lambda\epsilon\iota\omega$  bietet zudem sowohl eine Komponente für kartographische Daten wie auch für die Verwaltung, Selektion und Analyse digitalisierter Bilder. Inzwischen ist es auch möglich geworden, maschinenlesbare Editionen im  $\kappa\lambda\epsilon\iota\omega$ -Eingabeformat zu beziehen – wenn auch natürlich erst wenige Quellen vorliegen; schließlich muß man berücksichtigen, daß auch die gedruckten Editionen nicht vom Himmel gefallen sind. Wissenschaftliche Satzsysteme, die einfachen Textverarbeitungen vorzuziehen sind<sup>15</sup>, gibt es ebenfalls auf dem Markt.

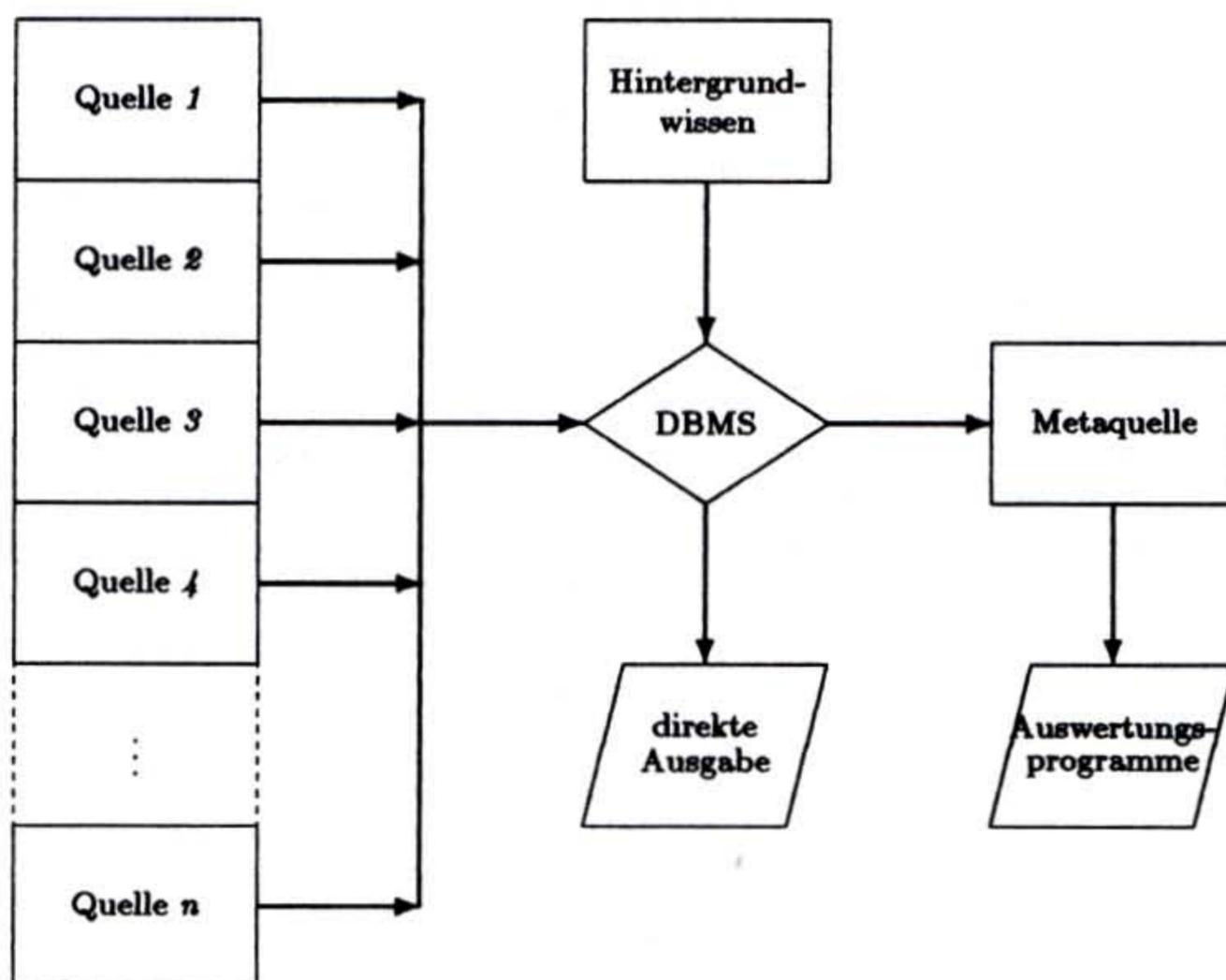


Abb. 1: Schema der Historical Workstation

Die Funktion, die sich für das Datenbanksystem als Vermittler zwischen den unterschiedlichen Komponenten der Historical Workstation ergibt, ist aus Abbildung 1 zu ersehen.

Im Zentrum der bisherigen Überlegungen steht das Datenbanksystem *κλειω*, das speziell für die historischen Wissenschaften unter der Federführung des Max-Planck-Instituts für Geschichte, Göttingen, entwickelt wird.<sup>16</sup> Zwar ist *κλειω* keine zwingende Voraussetzung, tatsächlich gibt es aber kaum eine Alternative, zumal kommerzielle Programme in der Regel auf historische Daten nur unzureichend anwendbar sind. Die von *κλειω* unterstützte Datenstruktur<sup>17</sup> und etliche seiner spezifischen Funktionen (für verschiedene Kalenderstile, historische Währungen, Ausgleich von Namensschreibweisen usw.) führen dazu, daß es unter den in der historischen Forschung eingesetzten Systemen eine Klasse für sich bildet.

Auf die Nützlichkeit der Künstlichen Intelligenz, insbesondere der sogenannten Expertensysteme kann hier nicht eingegangen werden. Angemerkt sei nur, daß sich hier Möglichkeiten für Mustererkennung in Bildern und automatische Erkennung gedruckter Texte ergeben. Daneben scheint sich ein weites Feld im Bereich der Integration von Wissensbasen über historische Währungen, Maßsysteme und ähnliches zu eröffnen. Der entscheidende Nachteil ist bis auf weiteres die Tatsache, daß Expertensysteme zur Zeit nur über ein sehr eingeschränktes Wissen verfügen und insbesondere nur mit – im Maßstab von Historikern – sehr kleinen Datenmengen arbeiten.

### 3.3. Objekte als Quellen

Folgt man der Definition des Begriffs „historische Quellen“, die nach P. Kirn *alle Texte, Gegenstände oder Tatsachen, aus denen Kenntnis der Vergangenheit gewonnen werden kann*<sup>18</sup>, umfassen, so scheint es nur natürlich, daß der Historiker sich auch mit Objekten befaßt. Tatsächlich überwiegt aber die Betrachtung der schriftlichen Überlieferung. Allerdings hat die Arbeit mit gegenständlichen Quellen in den letzten Jahren an Bedeutung gewonnen, zum einen im Bereich der Bildanalyse, wo, im Gegensatz zur Kunstgeschichte, das Abgebildete in den Vordergrund tritt; zum anderen im Umfeld der historischen Museen, deren Darstellung sich an anderen Motiven als denen einer volkskundlichen Präsentation orientiert.

In der Praxis erweist es sich als nicht einfach, die sehr heterogenen musealen Informationen zu verwalten und vollständig zu nutzen. Versuche mit musealen Informationssystemen gibt es bereits seit den 60er Jahren.<sup>19</sup> Die dort beschrittenen Wege sind aber von Historikern eigentlich nie weiter verfolgt worden, nicht zuletzt wegen der sehr aufwendigen Formalisierung.<sup>20</sup> Dennoch haben gerade die Arbeiten von Gundlach und Lückcrath eine andauernde Nachwirkung in anderen Disziplinen.

Kunstgeschichte und Volkskunde arbeiten sehr gerne mit sogenannten „Thesauren“, das sind Klassifizierungssysteme, die eine einheitliche Beschreibung ähnlicher Merkmale durch unterschiedliche Bearbeiter sicherstellen sollen. Hier ist in diesen Bereichen auch beachtliches geleistet worden.<sup>21</sup> Nachteilig ist aber, daß sich die Erstellung eines Thesaurus an einer vorab definierten Terminologie orientieren muß. Ob hier in allen Bereichen wirklich konsensfähige Lösungen gefunden werden können, darf mit Recht angezweifelt werden. So würde ein sinnvolles, alle Sammlungen eines Museums abdeckendes Inventarsystem eine Reihe vernetzter Thesauren erfordern. Damit wird das Problem ihrer Erstellung potenziert, zumal bei unterschiedlichen Fachinteressen der beteiligten Wissenschaftler.<sup>23</sup>

Die Interessen von Kunstgeschichte und Geschichtswissenschaft gehen hier auseinander. Während erstere primär an der Erstellung von Katalogen mit mehr oder weniger standardisierten Informationen zum kunsthistorischen Hintergrund des Objekts arbeiten, gehen die Erwartungen der Historiker deutlich über diesen Ansatz hinaus. Die inhaltliche Analyse im Sinne der Geschichtswissenschaft setzt eine detaillierte Beschreibung der Einzelobjekte, eine Verbindungsmöglichkeit zu Schriftquellen und umfangreiche Such- und Auswertungsstrategien voraus. Dieser Ansatz *is very different from such done by curators of museums or by art historians* und setzt als Forschungsinstrument eine *much more accurate and detailed*<sup>24</sup> Beschreibung des ikonographischen Inhalts voraus. Wenn die hiermit verbundenen methodischen Untersuchungen zur Handhabung von Bildinhalten als historische Quellen erfolgreich abgeschlossen sind<sup>25</sup>, dann könnten Bilder *sources for the historian, opening a wide field of interpretation, of new aspects and of additional information to written sources*<sup>26</sup> werden.

Der entscheidende Unterschied im Anspruch der Disziplinen manifestiert sich im Vergleich des kunsthistorischen Datenbanksystems MIDAS mit κλειω. Während die zentrale Aufgabe für das kunsthistorische System das *Retrieval*, d. i. das Wiederfinden von Information, ist<sup>27</sup>, steht bei κλειω die Aufbereitung der Quelle für weitergehende maschinengestützte Analyse im Vordergrund. Demzufolge verfügt MIDAS über einen hochqualifizierten, sehr umfangreichen aber dennoch beschränkten Klassifikationsrahmen<sup>28</sup>, während bei κλειω das Schwergewicht auf der Erstellung einer quellenangepaßten Umgebung beruht. Gerade im Vergleich der beiden Systeme wird der grundlegend unterschiedliche Ansatz deutlich. Nicht, daß die Geschichtswissenschaft Thesauren ablehnen würde, vorgefertigte Klassifikationsschemata wären aber für sie sinnlos, da der Thesaurus fragestellungsbezogen erstellt werden muß. Das gilt auch dann, wenn sich Historiker z. B. mit Bildern und Objekten befassen, denn ihr Zugang ist ein anderer als der der Kunstgeschichte oder der Volkskunde.

Neben diesen eher universellen Verfahren und Methoden gibt es noch eine ganze Reihe von Möglichkeiten, Datenverarbeitung für die Geschichtswissenschaft nutzbar zu machen. Nicht angesprochen wurde etwa die Möglichkeit der Registererstellung (bis hin zu KWIC [Key Word In Context]-Indices), der Visualisierung von Ergebnissen in Karten oder Graphiken, linguistischer Analysen von fortlaufenden Texten und anderes mehr.

#### 4. Projekte

Zur Veranschaulichung sollen hier drei Projekte, die mit Mitteln der Historischen Datenverarbeitung arbeiten, dargestellt werden. Die Beispiele wurden aus dem Bereich der Landes- bzw. Stadtgeschichte gewählt. Sie können als typische Vertreter verschiedener Forschungsrichtungen stehen und zeigen die Anwendungsbreite der Historischen Datenverarbeitung. Die Wahl fiel auf:

1. HETRINA/HESAUS, die beiden Datensammlungen zu den Hessischen Truppen im amerikanischen Unabhängigkeitskrieg und zur Hessischen Auswanderung im 19. Jahrhundert,
2. LEHNBU, die Datenbank zur spätmittelalterlichen Wirtschaftsgeschichte des Hochstifts Würzburg,
3. das Regensburger Bürger- und Häuserbuch, als Kombination unterschiedlicher, komplexer Quellenbanken.

Die Auswahl der Projekte wurde durch zwei Kriterien wesentlich beeinflusst. Sie sollten hinreichend ausführlich dokumentiert und für eine bestimmte Vorgehensweise stellvertretend sein. Ein weiterer wichtiger Punkt ist es – und es wurde versucht, ihn im gesamten Aufsatz durchzuhalten –, daß die entsprechende Literatur leicht und für jedermann zugänglich ist.<sup>29</sup> Zudem ist es notwendig, daß von in Frage kommenden Projekten die technische Seite dokumentiert ist. Aus den the-

matischen historischen Arbeiten kann man in der Regel kaum Schlüsse auf die angewandte Technik ziehen. So wünschenswert es gewesen wäre, alle vorgestellten Forschungen aus der hessischen Landesgeschichte zu wählen, so wenig schien dies unter diesen Prämissen realisierbar.<sup>30</sup>

#### 4.1. HETRINA/HESAUS

Aus dem Bereich der hessischen Landesgeschichte sind insbesondere die beiden Projekte HETRINA und HESAUS hervorzuheben.<sup>31</sup> Diese beiden Unternehmen basieren allerdings auf eher konventionellen und inzwischen wenig innovativen Methoden der Datenverarbeitung.

Das Projekt HETRINA (d. i. *Hessische Truppen in Amerika*) ist stark geprägt von den bis in die 70er Jahre vorherrschenden Perspektiven des Computereinsatzes in den Geschichtswissenschaften. Ganz dem Prinzip des Zählens, Sortierens und Registererstellens folgend, durch gewisse Umstände der archivalischen Tagesarbeit wahrscheinlich gefördert, stellt HETRINA im wesentlichen eine Liste der im amerikanischen Unabhängigkeitskrieg in den Jahren 1776 bis 1784 eingesetzten Soldaten in hessen-kasselischen Regimentern dar.<sup>32</sup> Bei einer geschätzten Gesamtzahl von 25 000 Soldaten rechnete man aufgrund von Doppelnennungen beim HETRINA-Projekt mit mindestens 60 000 Einträgen.<sup>33</sup> Folgt man den Ausführungen von Eckhart G. Franz<sup>34</sup>, so war damals an eine über die Listenform hinausgehende Auswertung nicht gedacht. Neben wissenschaftlicher Recherche nach einzelnen Offizieren und Soldaten sollten die zunehmenden genealogischen Anfragen schnell, ohne langwieriges Aktenstudium, beantwortet werden können.<sup>35</sup>

Ähnlich ist die Situation beim HESAUS-Projekt, in dem die registrierte Auswanderung aus Kurhessen im 19. Jahrhundert erfaßt wird. Die gesamte HESAUS-Datei enthält etwa 90 000 Einträge von rund 50 000 Einzelpersonen.<sup>36</sup> Auch Inge Auerbach wendet sich in ihrer Erläuterung zu den Auswandererlisten ausdrücklich an den Genealogen.<sup>37</sup> Es wurden dafür ausschließlich Archivalien des Staatsarchivs eingesetzt, mit dem primären Ziel einer erleichterten Auskunftserteilung an Genealogen und Lokalhistoriker.<sup>38</sup>

Es wird der Vorschlag gemacht, anhand der gedruckten Indices Auswertungen hinsichtlich Herkunft, Todesursache oder Desertion vorzunehmen<sup>39</sup>, eine Idee, die leider Zeitverschwendung und viele Fehlermöglichkeiten nach sich zieht und als wenig tauglich abzulehnen ist. Es entsteht die groteske Situation, daß ein großer Datenbestand (zumindest am Anfang der Projekte) primär für die schriftliche Edition erfaßt wurde.<sup>40</sup> Eine quantitative Bearbeitung des Materials ist nicht bekannt. Zwar gibt ein Aufsatz von Inge Auerbach<sup>41</sup> einige statistische Übersichten wieder. Von einer statistischen Analyse kann aber gar keine Rede sein. Zudem hält das wiedergegebene Zahlenmaterial den weiter oben genannten Kriterien für eine fundierte quantitative Analyse (Begriffsbildung, Angabe fehlender und unbekannter Werte, Fehlerabschätzung usw.) nicht stand.

Die Daten wurden nach einem relativ engen Schema erhoben. Dies ist zweifellos eine Folge der in der Startphase verwendeten Lochkartentechnik, die nur sehr wenig Spielraum für quellennahes Vorgehen ließ. Zum Teil scheinen aber auch arbeitsökonomische Erwägungen eine nicht unerhebliche Rolle bei der Wahl von Kürzeln für Berufe und ähnliches gespielt zu haben. Ihre Ziele werden die Listen mit Sicherheit erfüllen. Das gewählte Eingabeformat stellt zudem für die Listenerstellung die effizienteste Erfassungsmethode dar.

In Anbetracht der langen Laufzeit der Projekte wäre eine nachträgliche veränderte Eingabekonvention vermutlich sogar schädlich gewesen. Heute sollte man aber nur noch im Notfall ein so starres Erfassungsschema wählen. Der Trend geht – wie die beiden nächsten Projekte zeigen – zur quellenorientierten Datenverarbeitung, die solche Vereinheitlichungen nicht zuläßt und damit aber weniger eingeschränkte Verwendungen der Daten ermöglicht.

#### 4.2. LEHNBU

Bei der Datenbank, die von Rolf Sprandel am Institut für Geschichte in Würzburg unter dem Titel LEHNBU (Lehnbuch) aufgebaut wurde, handelt es sich um eines der größten, komplexesten und interessantesten Projekte der historischen Datenverarbeitung in Deutschland.<sup>42</sup> Obwohl inzwischen als eines der „älteren“ Projekte anzusprechen, sind die hier realisierten Konzepte und Standards immer noch richtungsweisend. Insbesondere ist hervorzuheben, daß es sich um eine der leider noch immer wenigen Datenbasen handelt, die tatsächlich für sehr unterschiedliche Fragestellungen nicht nur benutzt werden *könnten*, sondern auch tatsächlich benutzt *wurden* und *werden*.

Die Datenbasis, deren Aufbau kontinuierlich seit 1976 durchgeführt wurde, enthält z. Z. Lehenbücher des Hochstifts Würzburg, sogen. *libri diversarum formarum*<sup>43</sup>, ein Lehenbuch der Würzburger Dompropstei, ein Lehenbuch der Grafschaft Castell, zwei Würzburger Salbücher, drei Zins- und ein Regelbuch. Mit diesen Quellen wird ein Zeitraum von 1303 bis 1529 abgedeckt.<sup>44</sup> Die Quellen erlauben in erster Linie wirtschafts- und sozialhistorische Analysen. Es lassen sich die Strukturen eines regionalen Herrschaftssystems, wirtschaftliche und soziale Verflechtungen in großer Zahl aufdecken. Studien zu Besitzakkumulation, Wirtschaftsweise, Währungs- und Maßwesen oder auch Preisentwicklung<sup>45</sup> lassen sich erarbeiten. Hier deutet sich schon der immense Vorteil des 1976 gewählten Verfahrens einer relativ quellennahen und quellenbezogenen Datenerfassung an. Im Gegensatz zu vielen anderen Projekten, wo z. T. nur eine einzige oder wenige ähnlich gelagerte Arbeiten verfaßt werden können, liegt die Leistungsfähigkeit des LEHNBU-Projekts in seiner sehr großen Bandbreite. Trotz einer bereits jetzt ungewöhnlichen Anzahl von vorliegenden Arbeiten<sup>46</sup> sind die Möglichkeiten noch nicht erschöpft. Zwar wird man die unterschiedlichen Grade der – sehr gemäßigten – Formalisierung nicht mehr beseitigen können<sup>47</sup>, dennoch erscheint das Material aufgrund der einfachen Eingabekonventionen trotz des langen Bearbeitungszeitraums als sehr homogen. Man kann die Weitsicht der Entwickler, die sich die Erkenntnis von Jean-Paul Genet, daß es *kein Kodieren nach der zu kodierenden Wirklichkeit, sondern nur ein Kodieren von einer bestimmten Fragestellung aus*<sup>48</sup> gibt, nicht stark genug hervorheben.

Die hier wiedergegebenen zwei Einträge (Abb. 3) zeigen beispielhaft die wesentlichen Elemente. In den sogenannten „Geschäftsordnungszeilen“, die mit den Symbolen ‚/‘, ‚&‘, ‚;‘ eingeleitet werden, sind Quellenangaben und -sprache, Datierung und kodierte Informationen zu der Art des Quelleneintrags angegeben. Es folgt der Quellentext, in dem bestimmte Teile, etwa Personen- und Ortsnamen, Immobilien, Geldsummen oder Waren mit bestimmten Symbolen gekennzeichnet sind. Häufig vorkommende Begriffe werden durch festgelegte Abkürzungen wiedergegeben (etwa *recepit in feodum* = ‚rf, *Bischof und Hochstift Würzburg* = ‚bhw, *Zehnte* = ‚zz). Ortsangaben wird – soweit sie

sich auflösen lassen – das Autokennzeichen des alten Landkreises beigeordnet.

Die Vorteile dieser quellenkonservativen Methode zeigen sich schnell. Durch die nur an der Quelle und nicht an einem bestimmten technischen System orientierte Aufbereitung der Texte ist es im Prinzip gleichgültig, wie man die Information weiterverarbeiten will. Insbesondere ist der gesamte textliche Inhalt der Quelle – Eingabefehler einmal ausgeschlossen – erhalten. Man kann für die Auswertung verschiedene eigene Programme entwerfen, die z. B. nur die Aufgabe haben, bestimmte Merkmale zu zählen.<sup>49</sup>

Eine sehr anschauliche Methode der Auswertung ist die kartographische Darstellung bestimmter Phänomene. So läßt sich das Material nach einem bestimmten Grundkriterium (Besitz einer Familie, verliehene Mühlen oder – wie im Beispiel Abb. 2 – den im Lehenbuch aufgeführten Weinäckern) selektieren und in zwei zeitliche Querschnitte aufteilen. Man erhält dadurch ein Bild der geographisch-zeitlichen Entwicklung von Wirtschaftsstrukturen oder Besitzverhältnissen.

Es ist sehr zu bedauern, daß dieses Projekt, dessen Wert für die Landesgeschichte – übrigens in kleineren Teilen auch für Hessen – kaum anzuzweifeln ist, eigentlich bis heute keine Nachahmung gefunden hat. Zwar dürfte es schwer sein, eine Region mit ähnlichen Herrschaftsverhältnissen wie sie in Franken typischerweise vorkommen, zu finden. Unabhängig davon aber liegt ein Modell vor, daß sich auf völlig andere Quellen und Verhältnisse übertragen ließe.<sup>50</sup>

### 4.3. Regensburger Bürger- und Häuserbuch

Stellvertretend für die jüngsten Projekte, die zum Teil eng an die Entwicklung der Historical Workstation angebunden sind, sei hier das Regensburger Bürger- und Häuserbuch genannt. Gerade hier wird auf eine massive Verwendung der Datenverarbeitung gesetzt. Erste schriftlich fixierte Ergebnisse lassen einen erfolgreichen Verlauf der einzelnen Teilprojekte erhoffen.<sup>51</sup>

Das 1984 gestartete Projekt steht von seinem Umfang und seiner Konzeption z. Zt. an der Spitze der Historischen Datenverarbeitung. Ziel soll es sein, Datenbanken zu erstellen, die die drei Bereiche Personen, Häuser, Realien umfassen und sich miteinander verknüpfen lassen.<sup>52</sup> Dem Projekt kommt die umfangreiche, wenn auch sehr komplexe Quellenüberlieferung der Reichsstadt zugute. Da von Beginn an mit einem Anfall von ca. 0,5 Millionen Datensätzen, vergleichbar 2 Millionen Karteikarten (incl. Querverweisen), gerechnet wurde, hat man am federführenden Stadtarchiv Regensburg von Beginn an auf den großzügigen Einsatz datentechnischer Mittel gesetzt.<sup>53</sup>

Innerhalb dieses Projektes werden sehr unterschiedliche Quellengattungen bearbeitet, die in Zukunft in einer übergeordneten Datenbank zusammengefaßt werden können. Zur Zeit werden Archivrepertorien, ein Inventar des Regensburger Stadtmuseums und ein Bildarchiv sowie verschiedene historische Datenbanken bearbeitet<sup>54</sup>.

Es zeigt sich, daß durch die inzwischen erreichte Komplexität der Projektname „Regensburger Bürger- und Häuserbuch“ *mittlerweile nur noch ein Arbeitstitel*<sup>55</sup> ist. Ziel sind Personenbiographien, Hausbiographien, realienkundliche Forschungen und die Entwicklung eines Modells zur Verknüpfung stadtgeschichtlicher Einzelquellen.<sup>56</sup> Die größte Schwierigkeit bei einem derarti-

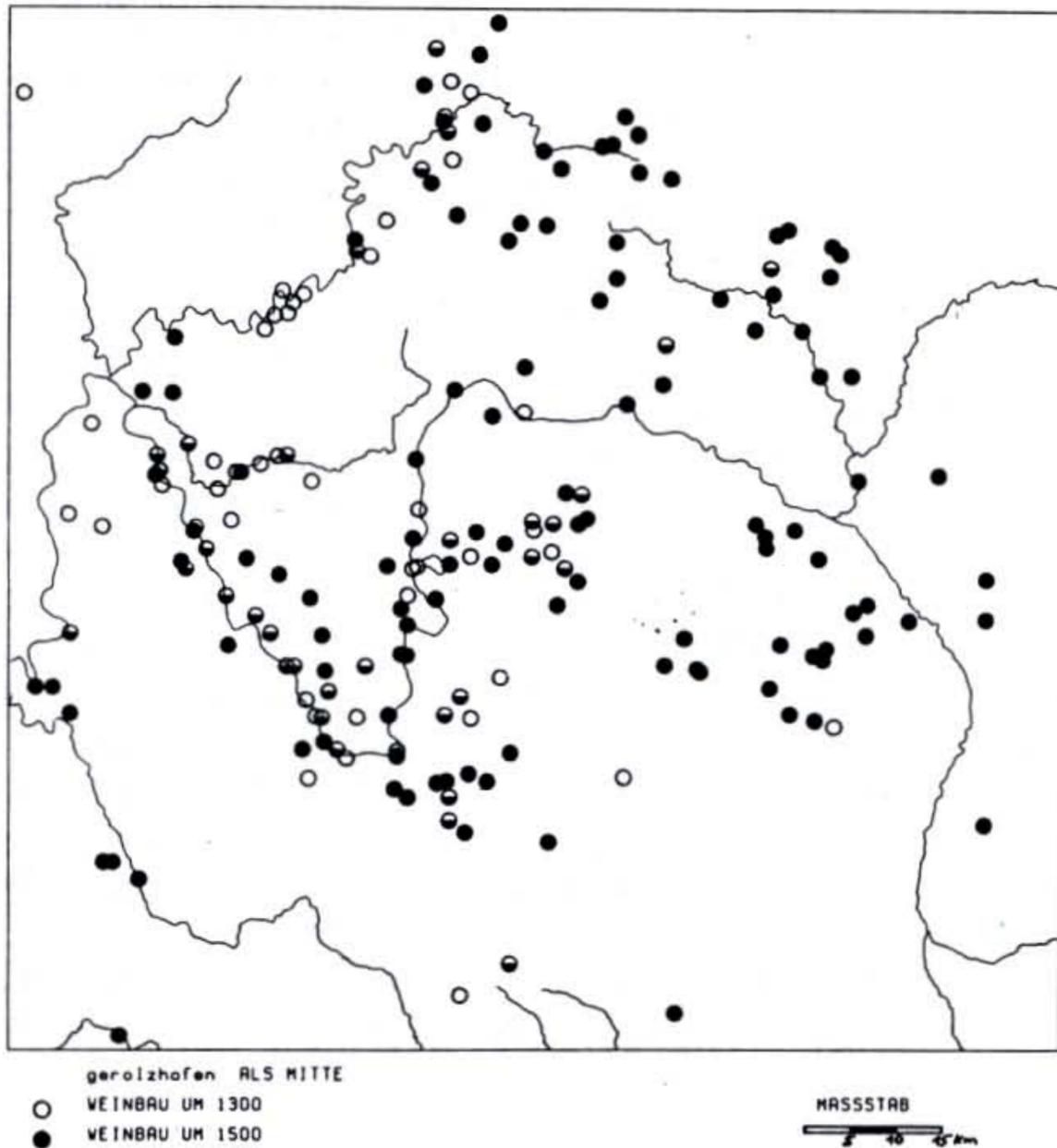


Abb. 2: Die Entwicklung des Weinbaus im spätmittelalterlichen Franken, beschrieben auf der Grundlage von Lehenbüchern.

Die Vorlage für die Landkarte wurde freundlicherweise von Dr. Dieter Rödel zur Verfügung gestellt.

010360

```

/lf 11, fol 23r, dt
&+/xe 1401 okt 03/+
;n;r;l,rl,uu
(hans & zollner, ritter) 'rfra
den halben 'zz zu (=birkenfeld {lk hoh}),
seinen teil von (=sulzbach {*lk hoh}),
das gut zu (=*/bynhawsen/* zu (=hofheim {lk hoh})),
den 'zz mit hoeften und wiesen zu (=bundorf {lk hoh}),
1 hube zu (=eyershausen {lk koen}),
:/$4 lb/: geld zu (=*/nidernlawringe/* (=stadtlauringen {lk hoh}))
seinen teil des 'zz zu (=friesenhausen {lk hoh}),
seinen teil des 'zz zu (=altenmuenster {lk hoh}),
8 gueter zu (=neuses {*lk hoh}) 'aiz und
die seen zu (=neuses {lk hoh}) und (=bundorf {lk hoh}),
*/die er umb halbe hat sein lebtage /* vom 'bhw.

```

071001

```

/wsb 1, fol 92r, dt
&+/xe 1470 000 00/+
;n;b;de
:/$20 d/: an +/martini {nov 11}/+, :/$20 d/: an +/walpurgis {mai 01}/+,
:/$4 metzen weizen/: und :/$1 fastnachtshuhn/:
von einer halben hube (zu (=dettelbach {lk kt})),
besitzer (hans & dreissigacker) und (klaus & reinhart).

```

Abb. 3: Auszug aus der Datenbank „LEHNBU“

Der Datenbankauszug wurde freundlicherweise von Dr. Dieter Rödel zur Verfügung gestellt.

gen Vorhaben dürfte in der Ausarbeitung eines verbindenden Modells für die sehr unterschiedlichen Quellen liegen. Solange ein solches fehlt, erhält man nur viele nebeneinander stehende Datenbasen, die einer gemeinsamen Auswertung mehr oder weniger große Hindernisse entgegenstellen. Darum ist es nicht verwunderlich, daß man in den Teilprojekten versucht, die für die übergeordneten Fragestellungen relevanten Informationen nach einem einheitlichen Schema zu erfassen und gleichzeitig – auch dies ein bemerkenswertes und für Arbeitsgruppen, die mit κλειω arbeiten, fast schon typisches Charakteristikum – den gesamten Quellentext ebenfalls aufzunehmen. Somit ist sowohl die Vernetzung disparater Quellen nach einheitlichen Kriterien als auch die auf den individuellen Teilquellen beruhende Auswertung möglich. Einleuchtend ist, daß in der Gestaltung des verbindenden Modells der kritische Punkt liegt. Hier gibt es bislang nur relativ wenig praktische Erfahrung.

Im Kontrast zum HETRINA/HESAUS-Projekt wird die ungeheure Weiterentwicklung der Historischen Datenverarbeitung sichtbar. Mit den drei vorgestellten landes- und stadtgeschichtlichen Projekten werden auch drei Generationen praktischer Arbeit mit dem Computer erkennbar, beginnend mit dem stark formalisierenden, rein seriellen Ansatz von HETRINA mit sehr engen technischen Grenzen, über das bereits sehr viel breiter angelegte Würzburger Unternehmen bis hin zu den sehr komplexen Regensburger Arbeiten. Während sich aber letzteres erst noch im ganzen bewähren muß und der technische und methodische Ansatz von HETRINA überholt ist, kann die Würzburger Datenbank durchaus Vorbildcharakter beanspruchen.

### **5. Historische Datenverarbeitung in der Landesgeschichte : Mögliche Projekte und Anwendungen**

Zweifellos ist ein Einsatz der Historischen Datenverarbeitung in der Landesgeschichte<sup>57</sup> eher an praktischen Überlegungen als an methodischen Prinzipien orientiert. Unverkennbar ist aber, daß ein erheblicher Teil der bisherigen Forschung, gerade aus dem Umfeld der historischen Demographie, starke landeshistorische Bezüge hat. Das ist zunächst ein rein arbeitstechnischer Effekt. Die Datenaufnahme verursacht in Projekten der quantitativen Forschung den größten Zeitaufwand. Will man einen für demographische Fragestellungen sinnvollen Zeitraum von mehreren Generationen abdecken, so steht man selbst bei relativ kleinräumigen Studien sehr schnell vor einem gewaltigen Datenberg. Selbst mit der Beschränkung auf Stichproben<sup>58</sup> wird man nur ein vergleichsweise kleines geographisches Gebiet abdecken können. Historische Demographie besitzt in Frankreich, England und den skandinavischen Ländern eine grundlegende Funktion für die Regionalgeschichte als spezifische Form der Landesgeschichte.<sup>59</sup>

Gerade das Mengenproblem fällt vielleicht in der Landes- und Regionalgeschichte nicht so stark ins Gewicht, wie bei anderen Untersuchungen. Oft herrscht bei der Betrachtung kleinräumiger Phänomene eine Konzentration auf einen relativ kleinen, zumindest aber vergleichsweise geschlossenen Quellenkomplex vor. Man muß sich vorstellen, daß es durchaus im Bereich des Machbaren liegt, die gesamte Monumenta Germaniae Historica maschinenlesbar zu erfassen. Würde das aber einen großen Vorteil bedeuten? Eher nicht, da für den einzelnen Forscher aus diesem riesigen Komplex<sup>60</sup> vielleicht zwei oder drei Urkunden von Interesse sind. Es ist zur Zeit auch noch relativ

schwierig, eine Bibliothek maschinenlesbarer Quellen so zu verwalten, daß komplexe Querverbindungen zwischen den einzelnen Teilen zu erschließen wären – ganz abgesehen von der fehlenden breitgefächerten Quellenbasis. Man wird sich noch einige Zeit in der Landesgeschichte seine Datenbestände leichter schaffen können als in anderen Teilbereichen der Geschichtswissenschaft.

Die „klassischen“ Quellen, die bislang mit datentechnischen Mitteln bearbeitet wurden, sind auch landesgeschichtliche Quellen: Kirchenbücher, Steuerregister, Personenregister unterschiedlichster Art, Güterverzeichnisse, Lehenbücher, Gerichtsprotokolle und andere mehr.

Eine für die Landesgeschichte interessante Möglichkeit bietet sich in der maschinellen Edition von Quellentexten. Neben der Option, einmal maschinell erfaßte Texte sehr einfach, schnell und vor allem vollständig nach bestimmten Kriterien durchsuchen zu können, ergeben sich zumindest bei größeren Editionsprojekten, möglicherweise aber auch schon bei relativ kleinen Textmengen, erhebliche Vorteile während der editorischen Arbeit. Bei einem entsprechend ausgearbeiteten methodischen Instrumentarium spricht sogar einiges dafür, daß hier effektiv finanzielle Mittel eingespart werden können.

Im Prinzip geht es darum, den vollständigen Text historischer Quellen maschinenlesbar zu erfassen. Dabei werden alle für die Edition wichtigen Informationen über die Quelle, insbesondere der Varianten- und Sachapparat, alle für die Registererstellung nötigen Markierungen sowie die typographischen Auszeichnungen eingegeben. Während der Erfassung der Texte lassen sich begleitende Auswertungen der bereits erfaßten Daten durchführen. Auf diese Weise kann man Erkenntnisse über Echtheit oder Überlieferung von Urkunden ermitteln, Diktatvergleiche anstellen oder auch Lücken im Text durch Vergleich von formelhaften Wendungen in ähnlichen Urkunden füllen.<sup>61</sup> Bei diesem Vorgehen müssen drei Bereiche miteinander verknüpft werden: die Datenbank für die Texte, das System für die Textanalyse und die typographische Endgestaltung. In den Projekten des Forschungsinstituts für Historische Grundwissenschaften der Universität Graz, auf die hier Bezug genommen wurde, verwendet man für diese Zwecke κλειω als Datenbanksystem, SPSS als Statistiksystem zur Textanalyse und TEX zur Satzerstellung.<sup>62</sup> Diese so erfaßten, standardisierten und dokumentierten Datensätze könnten dann auch an andere Projekte zur Auswertung weitergegeben werden.<sup>63</sup> Dabei gilt die Vorgabe, daß ein derartiger Datensatz *zumindest die gleichen Anforderungen erfüllen muß wie eine moderne Edition im herkömmlichen Druck*<sup>64</sup>.

Die Landesgeschichte wäre in jedem Fall gefragt, wenn es um die Erstellung von Wissensbasen geht. Das Hintergrundwissen, das erforderlich ist, um große Quellenbanken rationell verwenden zu können, wird zu einem nicht geringen Teil in einer regionalbezogenen Forschung gewonnen werden können. So werden in Zukunft geographische Datenbanken ebenso erforderlich sein wie zuverlässige und umfassende Informationssammlungen über historische Währungen, Preise, Maß- und Kalendersysteme, aber auch biographisches Material und genealogische Daten. Die Wissensbasen könnten neben ihrem allgemeinen Nutzen auch gerade für die Regional- und Heimatforschung von großem Interesse sein. Mit entsprechend leistungsfähigen Expertensystemen in den zentralen wissenschaftlichen Bibliotheken etwa könnten sich auch Nicht-Experten (auf dem Feld der Genealogie, der Wappenkunde, der Siegel-

kunde usw.) Fachwissen auf dem Gebiet der verschiedenen historischen Hilfswissenschaften erschließen.

Für die hessische Landesgeschichte wäre es wünschenswert, wenn die Datenbestände von HETRINA und HESAUS maschinenlesbar ediert und für die Forschung frei verfügbar wären. Hier stünde eine Informationsquelle zur Verfügung, die für die Landesgeschichte großen Wert haben könnte. Ich sage *könnte*, denn eine Verfügbarkeit auf einem Großrechner mit sehr engen Zugangsmöglichkeiten oder als gedruckte Liste kann in diesem Fall nur als zweitbeste Lösung angesehen werden. HETRINA und HESAUS wären auch interessante Bestände für eine Wissensbasis mit genealogischen und prosopographischen Daten.

## 6. Forderungen

Wer datentechnische Methoden in der Geschichtswissenschaft anwenden will, sollte folgendes beachten: Je intensiver man Methoden der Historischen Datenverarbeitung einsetzen will, desto genauer muß man seine Quellen kennen. Einen Ersatz für die Einarbeitung in das historische Material kann es nicht geben.

Aus dieser Erkenntnis entwickelt sich früher oder später die Forderung nach Quellennähe. Es hat keinen Zweck, historische Daten so stark zu formalisieren, daß sie sich einfach erfassen und äußerst effizient auswerten lassen. Der Preis für die daraus erwachsende Stromlinienförmigkeit ist zu hoch. Man schränkt seine eigenen und vor allem auch die Forschungen anderer unnötigerweise ein, wenn man die Daten nur auf eine spezielle Fragestellung bezogen erfaßt.

Die wesentliche Erkenntnis, die man sich vor einem möglichen Einsatz datentechnischer Methoden zu eigen machen sollte, lautet wie folgt:

*Society and culture produce fuzzy structures that are not easily accesible for methods of electronic data processing. The data of the Humanities and Social Sciences is often very complex and the records are, in various ways, incomplete. Computer programmes that are produced for commercial purposes are often very insufficient and have to be extended with programmes developed by the scientists themselves.*<sup>65</sup>

Es ist erforderlich, daß Historiker genau definieren, was sie von der Informatik für Lösungen für ihre Anforderungen erwarten. Von den drei Ebenen der Anwendung, die Thaller definiert – der rein werkzeughaften Verwendung (Textverarbeitung), der Übernahme methodischer Werkzeuge anderer Disziplinen und der Entwicklung eigener Techniken für die Geschichtswissenschaften<sup>66</sup> –, ist die letztere eindeutig weiter zu forcieren. Das Lamentieren über fehlende oder unzureichende Programme hilft wenig, wenn man nicht klar beschreiben kann, was man haben möchte.

Es scheint sich etwas zu bessern, wenn die bereits zitierten Autoren Recht behalten mit der Behauptung, der Trend ginge von der Betrachtung einzelner Softwareprodukte zurück *to the question of which basic model of knowledge representation on a computer is inherently best tuned to historical research.*<sup>67</sup> Eigene Beobachtungen sprechen leider dafür, daß außerhalb des Kreises der Entwickler von historischen Softwaresystemen, bei den reinen Anwendern, ein großes Wissensdefizit im Bereich Datenstrukturen, Datenmodelle, Daten-

und Informationsrepräsentation vorliegt. Oft äußert sich hier auch eine erschreckende Unsensibilität gegenüber historischen Quellen.

Da nicht davon auszugehen ist, daß irgendwann einmal alle Historiker dasselbe Softwaresystem benutzen werden, muß gewährleistet sein, daß die Daten als zentraler Teil der Historischen Datenverarbeitung möglichst ohne technische Hindernisse ausgetauscht werden können. Der in der Datenbanktechnik eingeführte Begriff der „Datenunabhängigkeit“, mit dessen Hilfe die Verarbeitung der Daten möglichst unabhängig von bestimmten Hardware- (etwa ganz speziellen Speichermedien) und Software-Komponenten (z. B. einem Programm, daß ihm eingegebene Daten nur für eine einzige Aufgabe verwenden kann), gemacht werden soll, müßte vermutlich etwas erweitert werden. In der Geschichtswissenschaft tritt nämlich der in kommerziellen Systemen wohl eher seltene Fall des kompletten Austauschs ganzer Datenbanken zwischen unterschiedlichen „Unternehmen“ auf. Hier scheint ein erheblicher Bedarf an theoretischer Forschungsarbeit zu liegen.<sup>68</sup>

Die Ausbildung an den deutschen Universitäten muß wesentlich verbessert werden. Das Lehrangebot für Geisteswissenschaftler auf dem Sektor Datenverarbeitung ist meistens geradezu lächerlich. Was nötig wäre, ist eine Ausbildung, die fähig macht, unterschiedliche Quellen zu analysieren und die Bearbeitung mit den verschiedensten Methoden der Historischen Datenverarbeitung zu realisieren. Erforderlich ist nicht eine allgemeine Einführung in die Datenverarbeitung, sondern in die historische Fachinformatik als einer Disziplin, die sich mit der Natur der Daten und einer ihnen adäquaten Verarbeitung befaßt. Erfahrungen mit entsprechenden Kursangeboten liegen vor.<sup>69</sup>

Zum Aufbau einer sinnvoll nutzbaren Bibliothek an maschinell lesbaren Quellensammlungen sind zwei Voraussetzungen zu erfüllen:

1. eine urheberrechtliche Absicherung der Erzeugnisse<sup>70</sup> und
2. die Anerkennung der maschinellen Edition als einer der traditionellen Editionsarbeit gleichwertigen wissenschaftlichen Leistung.

### 7. Schlußbemerkung

Diese Darstellung ist nichts weniger als eine vollständige Übersicht über das jetzt oder in Zukunft mit der Datenverarbeitung Machbare. Hier kann auch kein Königsweg zur Historischen Datenverarbeitung gewiesen werden. Es gibt ihn nicht. Es kann nur im Ansatz gelingen, einen Ausschnitt aus der großen Bandbreite der Möglichkeiten darzustellen. Die Historische Datenverarbeitung hat mittlerweile eine technische Verfeinerung erreicht, die vor wenigen Jahren nicht denkbar war. Wer sich ihrer bedienen will, sollte sich sehr genau mit dem 'state of the art' vertraut machen, indem er sich von Anwendern und Entwicklern der Historischen Datenverarbeitung (und nicht von vermeintlichen „Computer-Spezialisten“) beraten läßt.<sup>71</sup> Illusionen sollte man sich aber keine machen: Vor dem geschichtswissenschaftlichen Erfolg steht – so lange es kaum maschinenlesbare Editionen gibt – die langwierige Dateneingabe.

Zumindest bei den Entwicklern, aber auch bei vielen Anwendern, ist die Historische Datenverarbeitung ein Auslöser zu einer intensiven Beschäftigung mit der historischen Methode und Anlaß zu grundlegenden quellenkundlichen Studien. Es kann durchaus als Vorteil angesehen werden, daß ein

erfolgreicher Einsatz Historischer Datenverarbeitung eine genaue Definition des geschichtswissenschaftlichen Vorgehens erfordert.

#### Anmerkungen

- 1 Als einführende Literatur seien vorab vor allem genannt: Karl Heinrich Kaufhold, Jürgen Schneider (Hrsg.): *Geschichtswissenschaft und elektronische Datenverarbeitung*, Stuttgart 1988; Albert Müller, Manfred Thaller (Hrsg.): *Computer in den Geisteswissenschaften*, Frankfurt a. M., New York 1989.
- 2 Erwin Riedenauer: *Elektronische Datenverarbeitung im Dienst von Landes- und Gesellschaftsgeschichte*. – In: *ZBayrLDG* 35,2 (1972), S. 379–435.
- 3 Das Beispiel ist entnommen: Manfred Thaller: *Gibt es eine fachspezifische Datenverarbeitung in den historischen Wissenschaften?* – In: Kaufhold, Schneider: *Geschichtswissenschaft und elektronische Datenverarbeitung*. S. 45–83. S. 57 ff. (Dieser Aufsatz ist äußerst lesenswert!)
- 4 Textverarbeitung und andere, eher der Büroorganisation zuzurechnende Verfahren, werden nicht zur Historischen Datenverarbeitung gezählt.
- 5 Riedenauer, *Elektronische Datenverarbeitung*, S. 387.
- 6 Wilhelm Heinz Schröder: *Historische Sozialforschung: Forschungsstrategie, Infrastruktur, Auswahlbibliographie*. – In: *Historical Social Research*, Suppl. 1, 1988.
- 7 Sehr gut zur Einführung: Gerhard Armingier, Konrad H. Jaraus, Manfred Thaller: *Quantitative Methoden in der Geschichtswissenschaft. Eine Einführung in die Forschung, Datenverarbeitung und Statistik*. Darmstadt 1985; Helmut Thome: *Grundkurs Statistik für Historiker. Teil I: Deskriptive Statistik* (*Historical Social Research*, Suppl. 2), 1989; ders.: *Grundkurs Statistik für Historiker. Teil II: Induktive Statistik und Regressionsanalyse* (*Historical Social Research*, Suppl. 3), 1990. Ältere Einführungen sind nur noch eingeschränkt zu empfehlen und werden deshalb hier nicht aufgeführt.
- 8 Schröder, *Historische Sozialforschung*, S. 23.
- 9 Ebd.
- 10 SPSS ist z. Zt. das verbreitetere. Zudem gibt es eine Einführung für Historiker (die sich aber auf eine alte Version von SPSS bezieht): Manfred Thaller: *Numerische Datenverarbeitung für Historiker*. Wien, Köln 1982. – Es scheint, daß die Mehrzahl der bis jetzt in Deutschland archivierten Datensätze mit SPSS bearbeitet wurden. Siehe: *A Guide to Historical Datasets in U.S. and European Social Science Data Archives, Part VII.: Central Archive for Empirical Social Research* (Center for Historical Social Research), Cologne (Germany). – In: *Historical Social Research* Vol. 16 (1991) No. 2, S. 182–189 and No. 3, S. 152–158 and No. 4, S. 135–143.
- 11 Der Unterschied zwischen einem Datenbanksystem und einer – weniger leistungsfähigen – Dateiverwaltung kann hier nicht dargestellt werden. Zu Datenbanksystemen siehe etwa: G. Schlageter, W. Stucky: *Datenbanksysteme. Konzepte und Modelle*. 2. Aufl. Stuttgart 1983. / P. C. Lockemann, J. W. Schmidt (Hrsg.): *Datenbankhandbuch*. Berlin u. a. 1987 (setzt etwas mehr Kenntnis voraus als Schlageter/Stucky).
- 12 Der Begriff wurde von Manfred Thaller eingeführt. Die Darstellung folgt weitgehend seinen in verschiedenen Aufsätzen niedergelegten Ausführungen hierzu. Siehe etwa: Manfred Thaller, *Data Bases v. Critical Editions*. In: *Historical Social Research* Vol. 13 (1988) No. 3, S. 129–139.
- 13 Aufzählung nach Manfred Thaller: *The Historical Workstation Project*. In: *Historical Social Research*, Vol. 16 (1991) No. 4, S. 51–61. S. 53. Der aktuelle Diskussionsstand ist gesammelt in mehreren Aufsätzen nachzulesen in: *Historical Social Research* Vol. 16 (1991) No. 4.
- 14 Thaller nimmt sie wohl deswegen nicht explizit in seine Aufzählung, weil sie nicht zum Entwicklungsbereich seines Projekts gehören.
- 15 Verf. bevorzugt T<sub>E</sub>X resp. L<sub>A</sub>T<sub>E</sub>X, was sich gerade in Verbindung mit κλειω anbietet. Siehe hierzu: Donald E. Knuth: *The T<sub>E</sub>Xbook*. 11. Aufl., Reading Mass. (etc.) 1991. / Norbert Schwarz: *Einführung in T<sub>E</sub>X*, 3. Aufl. 1991. / Leslie Lamport: *L<sub>A</sub>T<sub>E</sub>X – A Document Preparation System*, Reading Mass., 1985. / Helmut Kopka: *L<sub>A</sub>T<sub>E</sub>X – Eine Einführung*. 3. Aufl., 1991. T<sub>E</sub>X gibt es als Shareware-Version (gegen Unkostenerstattung) für Mitglieder der deutschen T<sub>E</sub>X-Anwendervereinigung DANTE e. V., Heidelberg.
- 16 Die Behauptung von Frank M. Bischoff, es erfülle *auch gewisse Funktionen eines Datenbanksystems* trifft als Einschätzung auch auf die Vorgängerversion CLIO nicht zu. Siehe Frank M. Bischoff: *EDV-Einsatz im Fachbereich Geschichtswissenschaften der Philipps-Universität Marburg*. Marburg 1988, S. 26. Zu κλειω vor allem: Manfred Thaller, κλειω 3.1.1. Ein Datenbanksystem. (Halbgraue Reihe zur historischen Fachinformatik, B1) St. Katharinen 1989; Peter Bekker, Thomas Werner: κλειω. Ein Tutorial. (Halbgraue Reihe zur historischen Fachinformatik, Band A1), 2. erweiterte Auflage St. Katharinen 1991.

- 17 Sie ist in den Handbüchern unzureichend erläutert. Wer es nicht aufgrund der Systemleistungen bemerkt, findet den entsprechenden Hinweis, daß es ein semantisches Datenbanksystem ist, nur an sehr versteckter Stelle. Thaller, κλειω 3.1.1, Abschnitt 5.8.
- 18 Zitat nach Ahasver v.Brandt : Werkzeug des Historikers, 10. Aufl. Stuttgart, Berlin, Köln, Mainz 1983, S. 48.
- 19 Carl August Lückerath, Rolf Gundlach : Historische Wissenschaften und elektronische Datenverarbeitung, Frankfurt a.M., Berlin, Wien 1976.
- 20 Siehe auch den Kommentar von M. Thaller : ... *man sollte aber nicht glauben, die Anwendung der EDV auf historische Fragestellungen müsse so kompliziert sein.* (Thaller : Numerische Datenverarbeitung für Historiker. S. 159).
- 21 Anschaulich : Bildarchiv Foto Marburg (Hrsg.) : Illustrierter Thesaurus der Datenbank Kunst in Deutschland. Umdruck 3 : Thesaurus antiker keramischer Gefäße. Marburg 1987.
- 22 Im Bereich der archäologischen Dokumentation meint man, daß noch nicht einmal ein grundlegender Konsens für eine einheitliche Terminologie vorhanden ist und somit auch jede wissenschaftliche Grundlage für einen Thesaurus fehlt. Ulrich Kampffmeyer : Das Archäologische Computer System „ARCOS“. – In : Manfred Thaller, Albert Müller (Hrsg.) : Computer in den Geisteswissenschaften, Frankfurt a.M./New York 1989. S. 139–215. S. 206.
- 23 Man geht im Landesmuseum für Technik und Arbeit, Mannheim, wohl zu recht davon aus, daß es bei der großen technischen Bandbreite, die für die dortige EDV-gestützte Inventarisierung erforderlich ist, unmöglich ist, für alle betroffenen Fachbereiche jemals Thesauren verfügbar zu haben. Arbeitspapier für die 5. Regionalkonferenz technikgeschichtlich orientierter Museen Baden-Württemberg, 1986.
- 24 Gerhard Jaritz : The Image as Historical Source or Grabbing Contexts. – In : Historical Social Research Vol. 16 (1991) No. 4. S. 100–105. S. 102.
- 25 Die Arbeiten werden am Institut für mittelalterliche Realienkunde Österreichs durchgeführt. Das dortige Bildarchiv umfaßt 20000 Photographien von mittelalterlichen Bildquellen, die durch die Datenbasis REAL beschrieben und erschlossen werden. Im Rahmen des Projekts wird die Komponente zur Bildverarbeitung mit κλειω entwickelt. G. Jaritz : Finding the Signs. Pictures of Medieval Life. – In : Heinrich Best, Ekkehard Mochmann, Manfred Thaller (eds.) : Computers in the Humanities and the Social Sciences. München et al. 1991, S. 61–67.
- 26 Jaritz, Images as Historical Source, S. 101.
- 27 Lutz Heusinger, Marburger Informations-, Dokumentations- und Administrationssystem (MIDAS) – Handbuch, München u. a. 1989. S. 7.
- 28 Der sich zugegebenermaßen erweitern läßt.
- 29 Was man von ungedruckten Arbeitspapieren, Magisterarbeiten oder internen Forschungsberichten bzw. Dokumentationen nun gerade nicht behaupten kann.
- 30 Eine Schwäche ist der Verzicht auf die demographischen Forschungen von A. E. Imhof. Siehe hier : Arthur E. Imhof (Hrsg.) : Historische Demographie als Sozialgeschichte. Gießen und Umgebung vom 17. zum 18. Jahrhundert. (Quellen und Forschungen zur hessischen Geschichte, 31) Darmstadt, Marburg 1975. Hier wäre aber eine Darstellung der historischen Demographie erforderlich gewesen, was zu umfangreich wäre. Zudem sind die Arbeiten sehr leicht zugänglich. Das an sich hinreichend gut dokumentierte Projekt von Barbara Hoffmann und Klaus Horn : Ehepaare, Eheverläufe und Lebenslauf in Leipzig 1580–1730 (In : Historical Social Research Vol. 15 (1990) No. 3. S. 171–198.) an der Gesamthochschule Kassel wurde nicht herangezogen, da für die Darstellung der Arbeit mit κλειω insbesondere im Zusammenhang mit der Historical Workstation die Regensburger Arbeiten wesentlich aussagefähiger sind.
- 31 Hessische Truppen im amerikanischen Unabhängigkeitskrieg (HETRINA) (Veröffentlichungen der Archivschule Marburg) Marburg 1972 ff.
- 32 Hessische Truppen im amerikanischen Unabhängigkeitskrieg (HETRINA), Bd. 1, (Veröff. der Archivschule Marburg, Nr. 10) Marburg 1972.
- 33 Ebd. S. 4.
- 34 Ebd. S. 3–17.
- 35 Ebd. S. 4.
- 36 Stand 1989. Inge Auerbach : Hessen-Kassels Rußlandauswanderung im 19. Jahrhundert. Zweck und Grenzen des HESAUS-Projektes. – In : Hess. Jahrbuch f. Landesgeschichte. 39. Bd. (1989) S. 251–269. Anm. 1.
- 37 Inge Auerbach : Hessische Auswanderer (HESAUS). Index nach Familiennamen. Bd. II : Auswanderer aus Hessen-Kassel 1840–1850 (Veröff. d. Archivschule Marburg, Nr. 12) Marburg 1988. S. 6.
- 38 Auerbach, Hessen-Kassels Rußlandauswanderung. S. 1.

- 39 Inge Auerbach, Elektronische Datenverarbeitung im Staatsarchiv Marburg. – In: Hess. Jahrbuch für Landesgeschichte, 28. Bd., Marburg 1978, S. 237–244. S. 238.
- 40 Die Daten sind Bestandteil der hessischen Datenbank HADIS. Über die Regeln für eine Benutzung dieser Daten ist noch eine Entscheidung der hessischen Landesregierung erforderlich. (Briefliche Mitteilung von Frau Archivoberrätin Prof. I. Auerbach, Staatsarchiv Marburg vom 16.1.1992).
- 41 Inge Auerbach: Auswanderung aus Kurhessen 1832–1866. – In: Hess. Blätter f. Volks- und Kulturforschung NF 17, 1985, S. 19–50.
- 42 Die folgenden Ausführungen basieren im wesentlichen auf Gesprächen mit Dr. Dieter Rödel, einigen eigenen Erfahrungen mit dem Projekt und den Aufsätzen von Dieter Rödel: „LEHNBU“. Ein EDV-Projekt zur Wirtschafts- und Sozialgeschichte Frankens im Spätmittelalter. – In: Jahrbuch der historischen Forschung in der Bundesrepublik Deutschland, Berichtsjahr 1989, S. 19–25; und Hans-Peter Baum, Rolf Sprandel: Die Erforschung von Lehensregistern in Verbindung mit der EDV. – In: Jahrbuch der historischen Forschung in der Bundesrepublik Deutschland, Berichtsjahr 1980, S. 49–55.
- 43 Bücher unterschiedlichen Inhalts.
- 44 Dieter Rödel: „LEHNBU“. Ein EDV-Projekt zur Wirtschafts- und Sozialgeschichte Frankens. S. 20.
- 45 Verf. hat sich selbst mit fränkischen Getreidepreisen im Spätmittelalter anhand dieser Datenbasis befaßt.
- 46 Genannt sei etwa Dieter Rödel: Das erste Salbuch des Hochstifts Würzburg (Studien zur Bayerischen Verfassungs- und Sozialgeschichte, XIII) München 1987, zugl. Diss. Würzburg 1984. Die Arbeit ist im übrigen ein sehr schönes Beispiel dafür, daß zwischen „klassischer“ Geschichtswissenschaft und „harter“ Quantifizierung sehr viele Übergänge existieren. Weitere genannt in Rödel, LEHNBU.
- 47 Als Schwachpunkt erweist sich die Übertragung aller Namen ins Neuhochdeutsche. Hier hätte man – auch bei nicht vorhandenem Soundex-System – die Originalschreibweise grundsätzlich mit aufnehmen sollen.
- 48 Zitat nach Hans-Peter Baum, Rolf Sprandel, Die Erforschung von Lehensregistern. S. 50.
- 49 Es mag auch bemerkenswert erscheinen, daß die Datenbasis bereits eine Umstellung der Rechensysteme am Rechenzentrum schadlos überstanden hat und jetzt auf ein drittes System (UNIX) portiert werden soll.
- 50 Allerdings muß man in Rechnung stellen, daß man zur Realisierung neben den nicht unbedeutlichen finanziellen Mitteln auch eine Persönlichkeit benötigt, wie sie R. Sprandel sicherlich darstellt. Die Daten sind für die Forschung über das Zentrum für Historische Sozialforschung, Köln zu beziehen: R. Sprandel: Economic and Social History of Franconia during the Late Middle Ages. ZA-Nr. 8036.
- 51 Literatur in Auswahl: Bettina Callies, Heinrich Wanderwitz: Das Regensburger Bürger- und Häuserbuch, eine Projektstudie. – In: Regensburg und Bayern im Mittelalter, Festschrift für Kurt Reindel zum 60. Geburtstag. (Studien und Quellen zur Geschichte Regensburgs, Band 4), Regensburg 1987, S. 131–147; Robert Dietz, Walburga Knorr: Die Bearbeitung von Inschriften am Stadtarchiv Regensburg mit TEX, κλειω und StanFEP. Die Epigraphische Datenbank. – In: Thomas Engelke, Jürgen Nemitz, Carolin Trenkler (Hrsg.): Historische Forschungen mit κλειω (Halbgraue Reihe zur historischen Fachinformatik, Band A8), St. Katharinen 1990, S. 99–112; Bettina Callies, Lothar Kolmer: A Computerised Medieval City Archive: The Project „Regensburger Bürger- und Häuserbuch“. – In: Peter Denley u. a. (Hrsg.): History and Computing II, Manchester, New York 1989, S. 266–272; Bettina Callies: Dokumentation im Projekt „Regensburger Bürger- und Häuserbuch“. – In: Friedrich Hausmann u. a. (Hrsg.): Datennetze für die Historischen Wissenschaften. Probleme und Möglichkeiten bei Standardisierung und Transfer maschinenlesbarer Daten. Graz 1987, S. 46–50; Steffen Wernicke, Martin Hoernes: Auswertung der spätmittelalterlichen Urfehdeurkunden Regensburgs. – In: Engelke/Nemitz/Trenkler (Hrsg.): Historische Forschungen, S. 17–22; dies.: „Umb die unzucht die ich handelt han ...“ Quellen zum Urfehdewesen. (Halbgraue Reihe zur historischen Fachinformatik, Band A9), St. Katharinen 1990; Thomas Engelke: „... und hiezen es schreiben in der stat puoch“. Das Projekt Regensburger Amtsbücher des 14. Jahrhunderts. – In: Engelke/Nemitz/Trenkler: Historische Forschungen, S. 9–16.
- 52 Callies/Wanderwitz, Regensburger Bürger- und Häuserbuch, S. 138.
- 53 Ebd. S. 131.
- 54 Informationspapier des Stadtarchivs Regensburg zur 1. Tagung der κλειω-user-group, Regensburg, 18.–21. März 1990.
- 55 Callies, Dokumentation, S. 47.

- 56 Ebd.
- 57 Um hier terminologischen Schwierigkeiten auszuweichen, soll unter diesem Begriff einmal alles zusammengefaßt werden, was bei C. H. Hauptmeyer (Hrsg.): *Landesgeschichte heute*. Göttingen 1987, an Teilbereichen vertreten ist. Für Stadt-, Regional- oder Landesgeschichte ergeben sich im Prinzip aus der hier gewählten Sicht der Historischen Datenverarbeitung keine wesentlichen Unterschiede.
- 58 Was grundsätzlich sehr zu empfehlen ist, wenn die Überlieferung dies zuläßt.
- 59 Ernst Hinrichs: *Regionalsgeschichte*. – In: Hauptmeyer, *Landesgeschichte heute*. S. 16–34. S. 20.
- 60 M. Thaller schätzt den Speicheraufwand der unstrukturierten Daten (d. h. als fortlaufender Text, nicht als Volltextdatenbank) auf etwa 0,5 Gigabyte. M. Thaller: *Vorüberlegungen für einen internationalen Workshop über die Schaffung, Verbindung und Nutzung großer interdisziplinärer Quellenbanken in den historischen Wissenschaften*. – In: Thaller, *Datenbanken und Datenverwaltungssysteme*. S. 9–30. S. 12.
- 61 Reinhard Härtel, Ingo H. Kropač: *Edition und Auswertung mittelalterlicher Urkunden: Probleme bei Standardisierung und Transfer fortlaufender Texte*. In: Friedrich Hausmann u. a., *Datennetze*, S. 100–112. S. 110.
- 62 Ebd. S. 108 f.
- 63 Härtel und Kropač räumen allerdings urheberrechtliche Probleme ein, deren Lösung aber nicht primär Aufgabe der Editoren sein kann; ebd. S. 112.
- 64 Ebd. S. 110.
- 65 Heinrich Best, Ekkehard Mochmann, Manfred Thaller: *Computers and Methodology: Promises of the Past, Reality of the Ninties?* – In: *Historical Social Research*, Vol. 14 (1989) Nr. 3. S. 5–8, S. 5 f.
- 66 Thaller, *Fachspezifische Datenverarbeitung*, S. 47.
- 67 Best, Thaller, Mochmann, *Computers and Methodology*, S. 7.
- 68 Überlegungen zu Datenmodellen aus dem Kreis der Geschichtswissenschaft sind eher selten. (Siehe etwa Jörn Sieglerschmidt: *Probleme des Aufbaus und Umfangs einer Datenbank 'Historische Statistik'*. – In: *Historical Social Research*, Vol. 13 (1988) No. 1. S. 89–110. S. 91 ff. oder Thaller, *Fachspezifische Datenverarbeitung*, S. 67 ff.)
- 69 Dargestellt z. B. von Ingo H. Kropač: *Von der Quelle zum Datensatz*. – In: Thaller, *Datenbanken und Datenverwaltungssysteme*, S. 199–215.
- 70 Dies kann nicht von der Geschichtswissenschaft geleistet werden, sondern steht im Zusammenhang mit allen ähnlich gearteten Verlagsprodukten.
- 71 Aus mehreren Möglichkeiten seien die zwei (in Deutschland) wichtigsten herausgegriffen: Für allgemeine Probleme und speziell Statistik und Quantifizierung das Zentrum für historische Sozialforschung, Bachemerstraße 40, 5000 Köln 41 (hier ist auch das zentrale Datenarchiv der historischen Sozialforschung in Deutschland) und für κλειω und eher „qualifizierende“ Forschung: Max-Planck-Institut f. Geschichte, – kleio –, Postfach 2833, 3400 Göttingen.