

Digitale Edition eines umfangreichen Quellenkorpus : Vorgehensweise und Probleme bei der Aufbereitung, Strukturierung und Kategorisierung des Quellenmaterials

Autor(en): **Kropa, Ingo H. / Boshof, Heidrun**

Objekttyp: **Article**

Zeitschrift: **Geschichte und Informatik = Histoire et informatique**

Band (Jahr): **11 (2000)**

PDF erstellt am: **20.06.2024**

Persistenter Link: <https://doi.org/10.5169/seals-8929>

Nutzungsbedingungen

Die ETH-Bibliothek ist Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Inhalten der Zeitschriften. Die Rechte liegen in der Regel bei den Herausgebern.

Die auf der Plattform e-periodica veröffentlichten Dokumente stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Einzelne Dateien oder Ausdrucke aus diesem Angebot können zusammen mit diesen Nutzungsbedingungen und den korrekten Herkunftsbezeichnungen weitergegeben werden.

Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die systematische Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber.

Haftungsausschluss

Alle Angaben erfolgen ohne Gewähr für Vollständigkeit oder Richtigkeit. Es wird keine Haftung übernommen für Schäden durch die Verwendung von Informationen aus diesem Online-Angebot oder durch das Fehlen von Informationen. Dies gilt auch für Inhalte Dritter, die über dieses Angebot zugänglich sind.

Digitale Edition eines umfangreichen Quellenkorpus: Vorgehensweise und Probleme bei der Aufbereitung, Strukturierung und Kategorisierung des Quellenmaterials

Ingo H. Kropač und Heidrun Boshof

Schon in den späten Siebzigerjahren wurde das Prinzip der quellenorientierten Datenverarbeitung formuliert und begonnen, die dazu notwendige Software zu entwickeln. In den Achtzigern bewirkte der Siegeszug des PCs, dass die ursprünglich sehr teuren und nur auf bestimmten Grosscomputern laufenden Lösungen von jedem Historiker unmittelbar auf seinem Schreibtisch angewendet werden konnten. In dieser Zeit entstanden auch verschiedene Ansätze zu einer computergestützten Quellenedition als logische Fortführung der quellenorientierten Datenverarbeitung, die aber im Wesentlichen immer das Ziel hatten, den Drucksatz zu unterstützen bzw. zu beschleunigen. Erst mit der Verbilligung der Hardware sowie schliesslich durch die rasante Entwicklung des Internets (speziell des World Wide Web) wurde es in der ersten Hälfte der Neunzigerjahre möglich, die Konzepte der Achtziger auch anhand grosser Quellenkorpora umzusetzen.¹

Wo stehen wir heute und welcher Mittel bedienen wir uns, wenn es gilt, digitale Editionen oder gar ganze virtuelle Archive zu produzieren? Was bedeuten diese Begriffe überhaupt bzw. welche methodischen und strategischen Implikationen bringen sie mit sich? Anhand eines Editionsunternehmens und der dahinterstehenden theoretischen Ansätze und methodischen Überlegungen soll in der Folge diskutiert werden, wie formale Verfahren

1 Um den vorgegebenen Rahmen dieses Beitrags nicht vollends zu sprengen, wurde im Folgenden darauf verzichtet, den Anmerkungsapparat massiv auszubauen, lediglich jene Werke wurden angegeben, die entweder zentral für die angesprochene Thematik sind oder wesentliche weitere Zitate enthalten. Zu einem (nicht mehr ganz aktuellen) Überblick zum Thema vgl. Kropač, Ingo H.: «Zur Konzeption von Informationssystemen in der Geschichtswissenschaft». In: Kaser Karl; Stocker Karl (Hgg.): *Clios Rache. Neue Aspekte strukturgeschichtlicher und theoriegeleiteter Geschichtsforschung in Österreich*. Wien-Köln-Weimar 1992, S. 87-129. – Kropač, Ingo H.: «Ad fontes. Von Wesen und Bedeutung der Integrierten Maschinellen Edition». In: Ebner, Herwig; Haselsteiner, Horst; Wiesflecker-Friedhuber, Ingeborg: *Geschichtsforschung in Graz. Festschrift zum 125-Jahr-Jubiläum des Instituts für Geschichte der Karl-Franzens-Universität Graz*. Graz 1990, S. 465-482.

die digitale Edition eines umfangreichen Quellenkorpus unterstützen und zu welchem Ende sie führen sollen.²

Wovon die Rede ist: die «Fontes Civitatis Ratisponensis»

Hinter dem Namen Fontes Civitatis Ratisponensis (FCR)³ verbirgt sich einerseits ein digitales Editionsunternehmen, das sich zur Aufgabe gestellt hat, die mittelalterlichen Quellen der Reichsstadt Regensburg zu edieren, andererseits ein Konzept der synergetischen Zusammenarbeit zwischen Archivaren, Historikern, Hilfs- und Informationswissenschaftlern zur Dokumentation und Erschließung unseres schriftlichen Kulturerbes. Um hierbei unterschiedlichen Benutzergruppen unterschiedliche Arten von Zugriffen zu ermöglichen, werden sowohl vernetzte Systeme als auch Einzelcomputer als Trägermedium für Faksimile- und kritische Textausgaben verwendet. Dies bedeutet, dass die Editionen aus einem Kernsystem in ein WWW-System exportiert und im Internet wie auch auf CDROMs angeboten werden.⁴

-
- 2 Um zu illustrieren, was als umfangreicher Quellenkorpus zu verstehen ist, seien einige Zahlen für den Zeitraum vom 13. Jahrhundert bis 1514 angeführt, jeweils nur die Regensburger Bürgerschaft betreffend: Wir schätzen auf ca. 60'000 urkundliche Quellen, ca. 30 «Cameralia» (Rechnungen, Zinsregister etc.) und über 300 «Literalien», in der Regel vermischte Amtsbücher mit einer grossen Zahl von überwiegend urkundlichen und rechtlichen Texten, ohne die neuzeitliche Rezeption zu berücksichtigen. – Ein vergleichbares Unternehmen stellt das «Duderstadt-Projekt» dar, obwohl der Ausgangspunkt dort die Bildung eines digitalen Archivs war, aus dem sich editorische Formen entwickelten, wogegen hier die Edition zuerst im Mittelpunkt steht und sich daraus das virtuelle Archiv ergibt. Zum «Duderstadt-Projekt» vgl. die Beiträge in: Ebeling, Hans-Heinrich; Thaller, Manfred: *Digitale Archive. Die Erschließung und Digitalisierung des Stadtarchiv Duderstadt*. Göttingen 1999.
 - 3 Die FCR sind das Ergebnis einer Kooperation zwischen dem Forschungsprojekt «Integrierte Computergestützte Edition (ICE)» am Forschungsinstitut für Historische Grundwissenschaften (Karl-Franzens-Universität Graz) und dem Amt für Archiv und Denkmalpflege der Stadt Regensburg. Die Finanzierung dieser Reihe von Forschungsprojekten erfolgt in der Hauptsache durch den österreichischen Fonds zur Förderung der wissenschaftlichen Forschung (Projekte P8798HIS, P10575HIS, P12083HIS, P14127OEK), die Digitalisierungskampagnen der letzten Jahre wurden von der Stadt Regensburg getragen. Das Forschungsinstitut für Historische Grundwissenschaften stellt die Infrastruktur der Projekte, von der Arbeitsplatzausstattung bis zu den Serversystemen. – Eine Übersicht zu den aktuellen Arbeiten, Demo-Versionen und ein bereits freigegebener Prototyp finden sich auf den Seiten der FCR-Online am Web-Server des Grazer Forschungsinstituts für Historische Grundwissenschaften (<http://www-fhg.kfunigraz.ac.at/fcr>).
 - 4 Die einzelnen Editionen der FCR werden nach der Methode der «Integrierten Computergestützten Edition» realisiert, die zwischen Produktions- und Anwendersystemen unterscheidet und im dritten Abschnitt dieses Artikels noch genauer diskutiert wird. Zu den Grundlagen vgl. Kropač, Ad fontes (wie Anm. 1) – Kropač, Ingo H.; Botzem, Susanne: «As You Like It. Archiving, Editing and

Um zu zeigen, auf welche Art und Weise der Benutzer auf die den Systemen zugrunde liegenden Datenbanken zugreifen und zu welchen Ergebnissen er so gelangen kann, scheint es sinnvoll zu sein, die Prototypen für zwei unterschiedliche Quellentypen vorzustellen, nämlich auf der einen Seite für serielle Quellen⁵ und auf der anderen Seite für urkundliche Quellen⁶, die für die Veröffentlichung sowohl auf CDROM als auch im Internet vorgesehen sind. Abgeleitet aus einem hilfswissenschaftlichen Expertensystem und aufbereitet für das World Wide Web sind beide Editionstypen mit dem Ziel produziert worden, eine möglichst grosse Zahl von Interessenten zu erreichen. Daher orientieren sich Struktur und Gestaltung des Systems an traditionellen Formen von Dokumentationen wie Urkundenbüchern und Regestenwerken. Für Kodices und kleinere gebundene Einheiten wie etwa Bürgeraufnahmebücher wurde als Zugriffsart die Seitenmetapher gewählt, für Urkundensammlungen hingegen die Regestenmetapher, um einen stückorientierten Zugriff zu ermöglichen.

Als Beispiel für serielle Quellen sei die digitale Edition der Bürgeraufnahmebücher der Stadt Regensburg angeführt. Diese zur Gruppe der Amtsbücher zählenden Quellen sind von 1419 bis 1860 fast lückenlos überliefert und enthalten eine Auflistung der neu- und wiederaufgenommenen Bürger der Stadt. Nach dem Aufruf der Edition erscheint ein in drei Bereiche gegliedertes Fenster mit der Menüleiste, dem Anzeigebereich und der Navigationsleiste (siehe Abb. 1). Über die Schaltflächen der Menüleiste lassen sich in eigenen Pop-Up-Fenstern die Einleitung, Suchmasken, Register und Hinweise zur Systembenutzung aufrufen. Ähnlich wie eine traditionelle Edition enthält die Einleitung einen Überblick über die Geschichte Regensburgs, eine Beschreibung des Quellentyps Bürgerbuch,

Analysing Medieval Documents». In: Josef Smets (Hg.): *Histoire et Informatique. Ve Congrès «History & Computing», 4.-7.9.1990 à Montpellier*. Montpellier 1992, S. 301-313. – Kropač, Ingo H.; Botzem, Susanne; Kurschel, Henriette: «Das ICE-Projekt». In: *Informatik Forum*, Nr. 8/4, 1995, S. 159-164 – Kropač, Ingo H.; Kropač, Susanne: «Prolegomena zu einer städtischen Diplomatie des Spätmittelalters: das Beispiel Regensburg». In: Prevenier, Walter; de Hemptinne, Thérés: *La diplomatie urbaine en Europe au moyen âge. Actes du congrès de la Commission internationale de Diplomatie, Gand, 25-29 août 1998*. Louvain/Apeldoorn 2000 (= *Studies in Urban Social, Economic and Political History of the Medieval and Early Modern Low Countries* 9), S. 229-265.

5 Als Beispiel hier: *Das zweite Bürgeraufnahmebuch der Reichsstadt Regensburg*, nach Vorarbeiten des Projektteams «Regensburger Bürger/Häuserbuch» bearbeitet von Susanne Kropač. Regensburg-Graz 1999 (Fontes Civitatis Ratisponensis 2). [CDROM: ISBN 3-925753-68-0].

6 Als Beispiel hier: Die Urkunden des Regensburger Almosenamtes von den Anfängen bis 1400, bearbeitet vom Team des Forschungsprojekts «Integrierte Computergestützte Edition (ICE)». Graz 1998-2000 [URL: <http://www-fhg.kfunigraz.ac.at/fcr/ualm.htm>].

eine Beschreibung der Handschrift, eine Darstellung der Editionsgrundlagen und verschiedene Überblicksstatistiken.

Der Anzeigebereich ist, da die Edition dieser Quelle der Seitenmetapher folgt, in zwei Hälften geteilt: In der linken Hälfte wird das Faksimile der jeweiligen Seite der Originalquelle, in der rechten Hälfte dagegen der dazugehörige Editionstext dargestellt. Aus diesem kann per Mausklick der Fussnotenapparat angefordert werden, der dann in einem Pop-Up-Fenster erscheint. Sowohl die Fussnoten als auch sämtliche Textdateien, die hier im HTML-Format vorliegen, sind Exporte aus den Datenbanken und werden dort integriert verwaltet. Nur das Interface wurde in HTML bzw. Java und Javascript geschrieben, der gesamte Editionsbereich (also Texte, Erschliessungen und dgl.) jedoch wurde aus der Datenbank erzeugt. Mit Hilfe der Schaltflächen der Navigationsleiste im unteren Bildschirmbereich wird dem Benutzer die Möglichkeit gegeben, in der Edition zu blättern wie in einem Buch oder eine vergrösserte Darstellung einer bestimmten Seite anzufordern, die ebenfalls in einem eigenen Pop-Up-Fenster erscheint.

Durch einen Mausklick auf die Schaltfläche Volltextsuche wird ein Java-Applet geladen, das ein Fenster öffnet, in dem man verschiedene Möglichkeiten zu einer Volltextsuche vorfindet. Dieses Programm realisiert in eingeschränkter Weise auf lokaler Ebene jene Funktionalität, die über die Datenbanken besser angeboten werden könnte. Die Programmierung des entsprechenden Interfaces steht gerade in Bearbeitung und wird wesentlich über die momentanen Möglichkeiten hinausgehen.

Ebenso wie jede andere Edition verfügt auch dieser Prototyp über umfangreiche Register, die aus der Datenbank erzeugt wurden und sämtliche in der Quelle genannten Orte und Personen sowohl in Originalschreibweise als auch in einer nach bestimmten Regeln standardisierten Form enthalten.⁷ Um den Grad der Wahrscheinlichkeit der Entsprechung von erkannten Orten mit den in der Quelle genannten Namen darzustellen, wurden daher Lokalisierungen mit einer Wahrscheinlichkeit grösser als 50% *antiqua* und solche mit einer Wahrscheinlichkeit kleiner als 50% *kursiv* gedruckt.⁸

7 Um die in der Quelle genannten Orte identifizieren zu können, wurde eine Reihe von Kriterien festgelegt. Zunächst musste geprüft werden, ob die mittelalterliche Ortsnennung einem aktuellen Ortsnamen zuzuweisen ist. War dies der Fall, stellte sich die Frage, wie dieser aktuelle Ortsname zu beschreiben ist, wobei sich eine Beschreibung nach heutigen Verwaltungseinheiten anbot. Dabei musste nun beachtet werden, mit welcher Wahrscheinlichkeit der lokalisierte Ort und die Nennung der Quelle identisch sind bzw. welche anderen Ortsnamen dem Quellenbeleg entsprechen können.

8 Zur Veranschaulichung sei hier die Identifizierung des im Bürgerbuch 2 erwähnten Ortes «Ortt» erwähnt: Dieser lässt sich an zwei heutigen Orten lokalisieren, nämlich Mariaort in der Gemeinde Pettendorf und Mariaort in der Gemeinde Sinzing. Die Entscheidung der Bearbeiter ergab eine

Dabei wird die Entscheidung der Editoren, welcher aktuelle Ortsname mit welcher Wahrscheinlichkeit dem in der Quelle genannten Ort entspricht, angeboten, der Benutzer kann letztendlich auch anders entscheiden.⁹

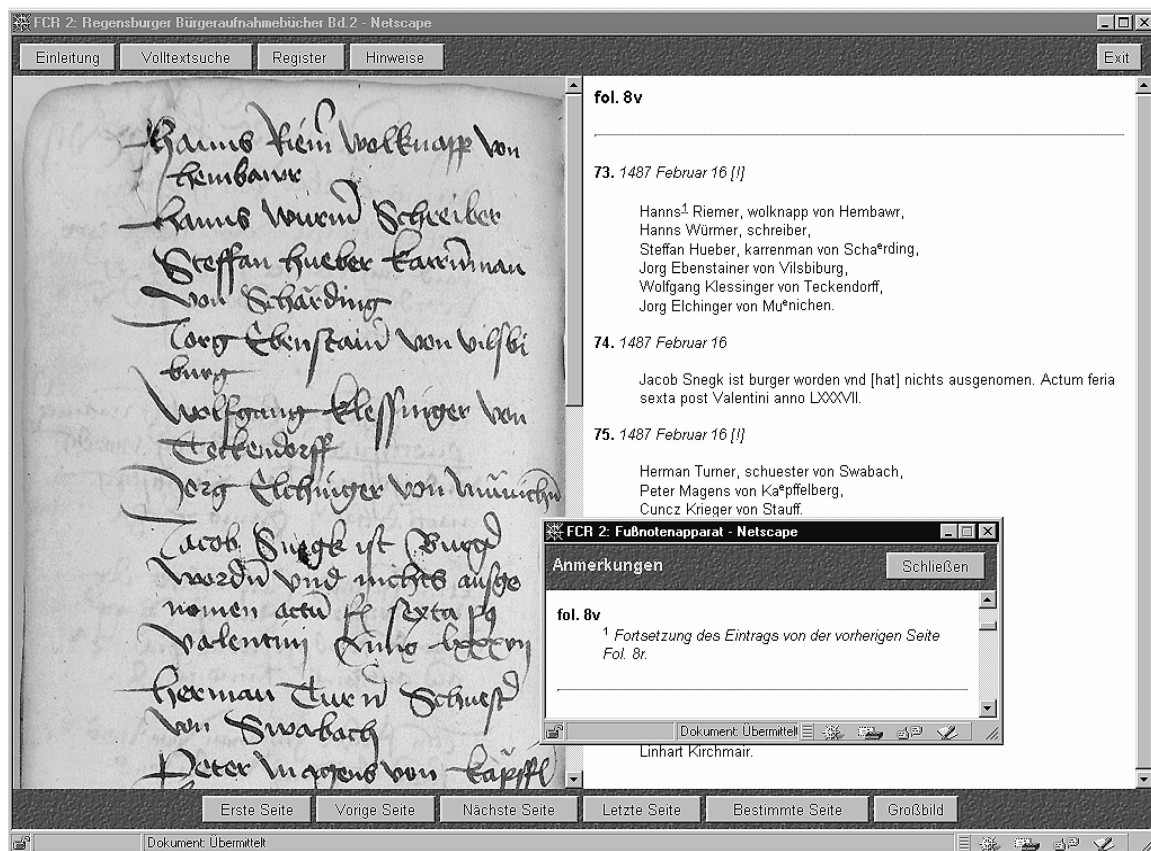


Abb 1: Das zweite Bürgeraufnahmebuch der Reichsstadt Regensburg: Faksimileedition mit Anmerkungenfenster.

Die Personenregister enthalten sämtliche in der Quelle genannten Personen, gegliedert nach den Vornamen in standardisierter Form, den Vornamen in Originalschreibweise und nach Beschreibungen. Dabei werden bekannte Verwandtschaftsbeziehungen der Personen untereinander kursiv in Klammern hinter den Namen gesetzt. Enthalten die Personennennungen sowohl die Berufsbezeichnung als auch den Herkunftsort, so sind die Personen im Register nach Beschreibungen sowohl unter dem Beruf als auch unter dem Ort verzeichnet.

höhere Wahrscheinlichkeit der Zuordnung zu Pettendorf, womit diese Möglichkeit antiqua, Sinzing hingegen kursiv ausgegeben wird. Ein Link führt jedoch zur jeweils anderen Möglichkeit.

⁹ Diese Angaben der Wahrscheinlichkeit beruhen auf evaluierten Sichtbarkeiten, die beim Record Linkage zwischen den Quellenbelegen und dem Ortskatalog (beides sind natürlich Datenbanken) zugewiesen wurden.

Als Beispiel für urkundliche Quellen soll die Edition der Urkunden des Regensburger Almosenamtes dienen. Sie ist als Prototyp für die Edition urkundlicher Quellen gedacht und folgt – anders als bei der Edition serieller Quellen – nicht der Seiten-, sondern der Regestenmetapher. Auch die Inhalte dieses Systems wurden nicht in HTML geschrieben, sondern aus den Datenbanken in definierte Benutzerschnittstellen migriert.

Auch beim Aufruf des Systems der urkundlichen Quellen erscheint ein Basisfenster, das mit einem Menüfeld auf der linken Seite, einer Navigationsleiste im oberen Teil und einem Anzeigebereich im unteren Teil der rechten Fensterseite in drei Bereiche geteilt ist. Über die Schaltflächen des Menüfeldes kann der Benutzer die Einleitung, verschiedene Arten des Zugriffs, Personen- und Ortsregister, mehrere Suchmöglichkeiten und wichtige Hinweise zum System anfordern, die wiederum in eigenen Pop-Up-Fenstern erscheinen. Die Schaltfläche «Datenbank-Zugriff via DML» (Data Manipulation Language) erlaubt direkte (beschränkte) Retrieval-Prozeduren im Kernsystem, wogegen die Schaltfläche «Datenbank-Zugriff via GUI» auf eine in Entwicklung befindliche und deshalb nicht freigegebene graphische Retrieval-Umgebung verweist.

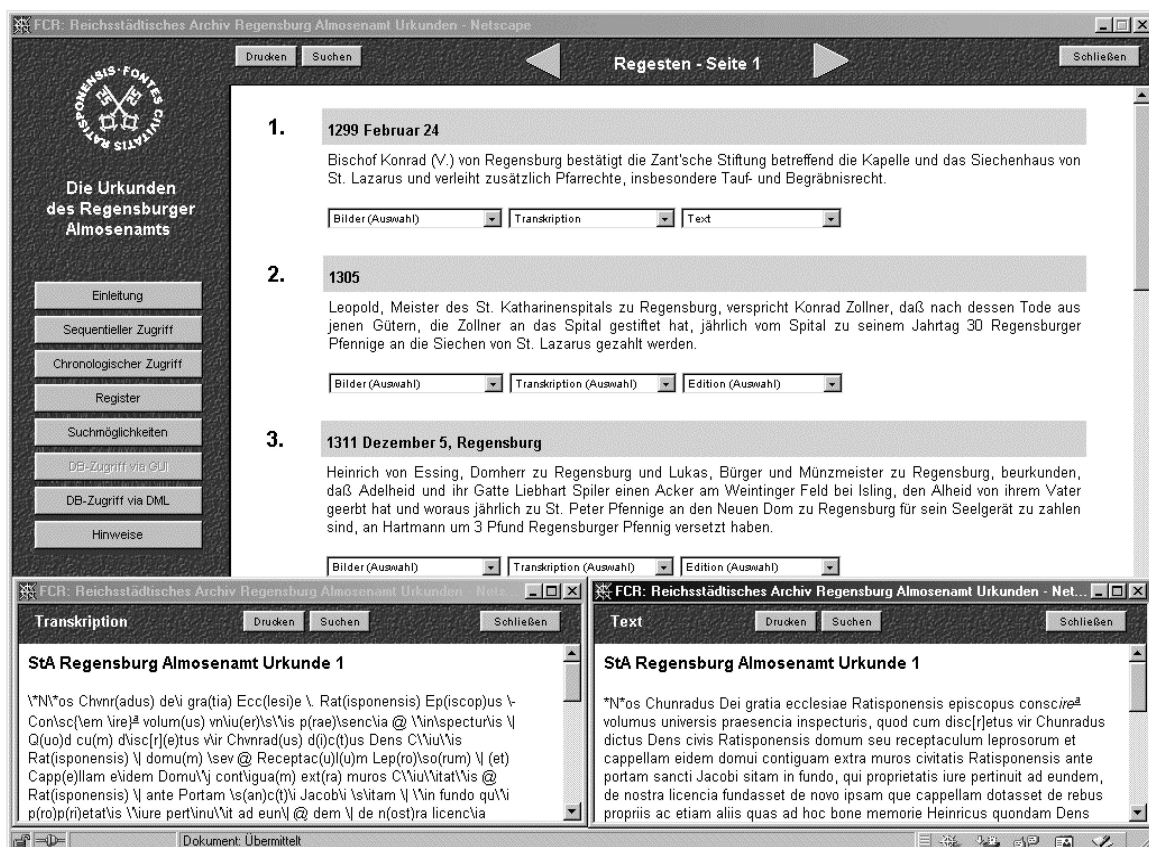


Abb. 2: Urkunden des Almosenamtes (StA Rgbg): Basisfenster mit Transkriptions- und Editionsfenster.

Die vielfältigen Suchmöglichkeiten dieser Edition sind ebenso wie die Volltextsuche der Edition serieller Quellen in Java programmiert und nehmen nur den in Entwicklung befindlichen Datenbank-Zugriff über ein Web-basiertes Graphical User Interface voraus, das derzeit in Entwicklung steht.¹⁰ Eine Besonderheit der Suche in diesem System stellt die Möglichkeit der Kombination verschiedenster Kriterien, nach denen die Urkunden durchsucht werden sollen, dar. Über Karteikarten kann der Benutzer Datum, Textsorte, Sprache, Inhalt und äussere Merkmale einer Urkunde bestimmen und so seine Suchabfrage in hohem Masse präzisieren.¹¹

Im Anzeigebereich, in dem man über die Navigationsleiste blättern kann, werden pro Seite zehn Regesten dargestellt, wobei sich unter dem einzelnen Regest nebeneinander drei Pull-Down-Menüs mit verschiedenen Auswahlmöglichkeiten befinden. Durch einen Mausklick auf das erwünschte Thema wird ein neues Fenster aktiviert, das im Anzeigebereich unter der Menüleiste das gewählte Bild bzw. den gewählten Text enthält. So kann der Benutzer entscheiden, ob er unter dem Punkt «Bilder» die Gesamtansicht der Urkundenvorder- und Rückseite, die Teilansicht des Textes wie auch der Dorsalvermerke, oder die Abbildungen sämtlicher vorhandener Siegel betrachten will. Im Bereich «Transkription» besteht die Möglichkeit, zwischen dem transkribierten Text, der Beschreibung der diskriminierenden Buchstaben und der verbalen Beschreibung des Duktus zu wählen. Anhand der Transkription soll versucht werden, den Text der Quelle möglichst originalgetreu darzustellen. Dies bedeutet, dass neben den mittelalterlichen Satz- und Pausenzeichen auch jene Buchstaben, die über das moderne lateinische Alphabet hinausgehen wie beispielsweise das lange S, ein gestrichenes i, ein Buchstabe mit Dach etc., dargestellt werden. Dazu war es notwendig, ein Codesystem zu entwickeln, das sich an die Eingabekonventionen des Satzsystems TeX anlehnt.¹²

10 Gunter Vasold entwickelt im Rahmen seiner laufenden Dissertation mit dem Titel *Netzbasierte Zugriffsformen auf komplexe geschichtswissenschaftliche Informationssysteme* die entsprechenden Interfaces, zum Teil nach der Assistentenlogik.

11 Stellt man sich beispielsweise im Rahmen einer Untersuchung von Privaturkunden die Aufgabe, alle von Bürgern ausgestellten Siegelurkunden, die auch Zeugen enthalten, zu suchen, so kann man diese Kriterien in der Suchabfrage formulieren und erhält ein präzises Ergebnis mit exakt jenen Urkunden, die diese Kriterien erfüllen.

12 Diese stark kodierte Form ist eine unmittelbare Ausgabe aus der Datenbank und dient weniger zur Lektüre als für Suchprozesse und Analysen. Zur Syntax vgl. Knuth, Donald E.: *The TeX-Book*. Reading 1984, bzw. die Systemdokumentation am Web-Server des Forschungsinstituts für Historische Grundwissenschaften [URL: <http://www-fhg.kfunigraz.ac.at/ice>].

Über das Auswahlm Menü «Edition» gelangt der Benutzer zu dem eigentlichen Editionstext, der ebenso wie der des Bürgerbuches weitgehend den Richtlinien der Monumenta Germaniae Historica und der Deutschen Reichstagsakten folgt. Zur Zeit steht unter diesem Auswahlm Menü nur der Punkt Text zur Verfügung, der den Quelltext nach Art einer traditionellen Edition wiedergibt.

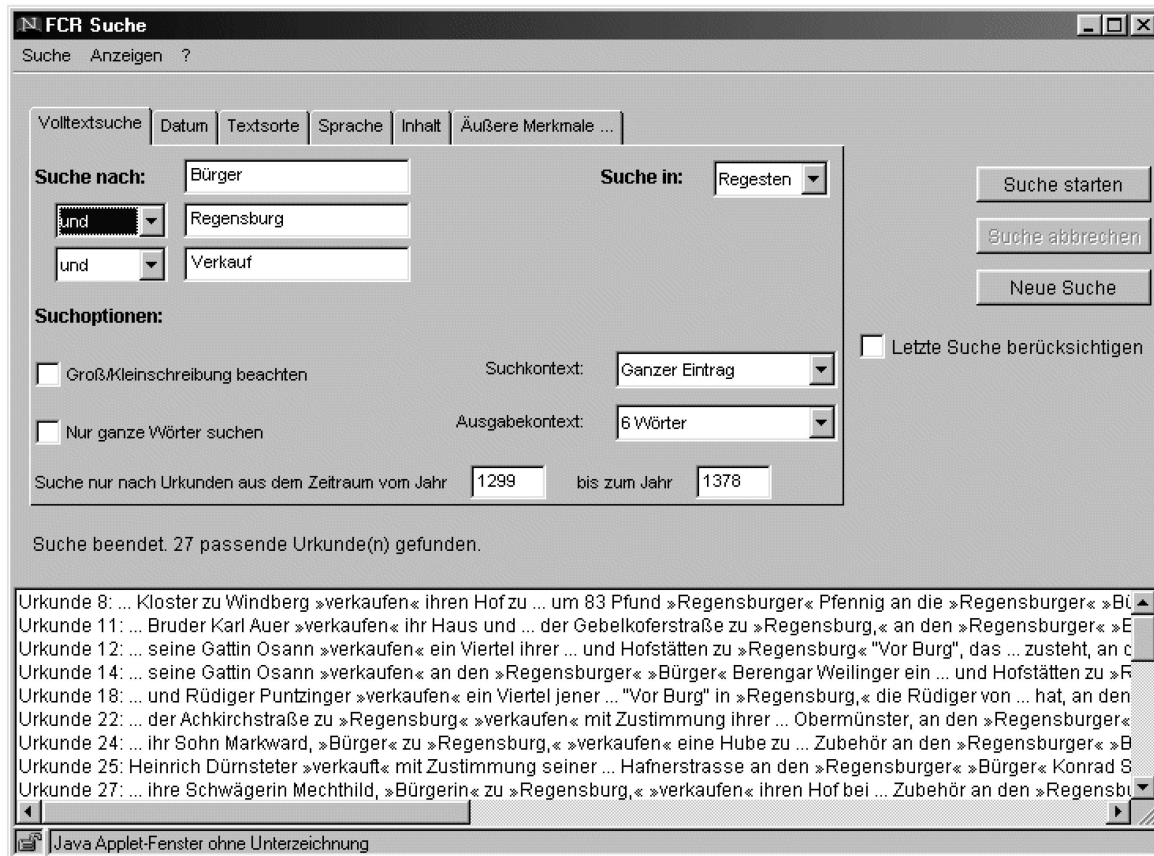


Abb. 3: Das Suchfenster

Was dahinter steht: Formen «historischer» Datenverarbeitung

Ausgangspunkt für Unternehmen in der oben beschriebenen Art sollten immer Entscheidungen im theoretisch/methodischen Bereich sein, um eine konsistente Arbeitsweise zu gewährleisten. Im Bereich der typisch «historischen», also geschichtswissenschaftlichen Datenverarbeitung lassen sich drei grundlegend verschiedene Formen ausmachen. Die methodenorientierte Datenverarbeitung konzentriert sich in der Regel auf einen Lösungs-

weg, ohne die Spezifika der Quelle zu reflektieren.¹³ Dazu zählen etwa die diversen Konkordanzbildungen im linguistischen Methodenkanon oder die statistischen Verfahren der quantifizierenden Methoden in der Geschichtswissenschaft. Dagegen nimmt die problem- oder phänomenorientierte Datenverarbeitung die effiziente Problemlösung unter Berücksichtigung der Methodenpluralität zum Ausgangspunkt, was nichts anderes bedeutet, als dass ein zugrundeliegendes Problem oder Phänomen die Informationsstruktur und die Datenmodellierung im Abbildungsprozess bestimmt.¹⁴ Damit stellt sie im Wesentlichen ein arbeitsökonomisches Modell dar, indem inhaltlich-thematische Einschränkungen und Auswahlprozesse aus den Quellen vorangeschaltet werden. Wann immer von Edition die Rede ist, spielt die quellenorientierte Datenverarbeitung die entscheidende Rolle; auf sie soll in der Folge näher eingegangen werden.¹⁵

Theorien über das Verhältnis zwischen einer überlieferten Vergangenheit und den sie beschreibenden Quellen müssen immer die Basis der formalen Analyse bilden. Zur Durchführung einer korrekten Quellenkritik und -analyse ist es meist erforderlich, sowohl innere als auch äussere Merkmale einer Quelle abzubilden. Zudem liegt es im Wesen der Quellen, zugleich mehrere, zum Teil überlappende Strukturen zu enthalten. Als Beispiel sei eine mittelalterliche Urkunde angeführt: Eine erste und einfachste Struktur lässt sich in der Einteilung nach Vorder- und Rückseite, Zeilen und Worten erkennen, eine weitere im Formular der Urkunde, eine dritte etwa im Sinne einer semantischen Analyse in der Einteilung nach Rechtssubjekten, Rechtsobjekten und Rechtshandlungen samt dem Beziehungsgeflecht, in das sie eingebettet sind. Da ein Historiker in der Regel nicht einen Typus von Quelle analysiert, sondern das von ihm zu untersuchende Phänomen in

13 Thaller, Manfred: «Ungefähre Exaktheit. Theoretische Grundlagen und praktische Möglichkeiten einer Formulierung historischer Quellen als Produkte <unscharfer> Systeme». In: Nagl-Docekal, Herta; Wimmer Franz: *Neue Ansätze in der Geschichtswissenschaft*. Wien 1984 (Conceptur-Studien 1), S. 77-100, hier: S. 83.

14 Kropač, Konzeption, S. 116 f. (wie Anm. 1) – Vgl. auch die Ansätze bei Schijvenaars, Toine: «Data Modelling of Sources in Dutch Historical Research Projects.» In: Bocchi, Francesca; Denley, Peter (Hgg.): *Storia & Multimedia. Atti del Settimo Congresso Internazionale di Association for History & Computing*. Bologna 1994, S. 501-522.

15 Der Begriff «quellenorientierte Datenverarbeitung» wurde im Wesentlichen von M. Thaller geprägt und in der Folge mit sehr unterschiedlicher Bedeutung in der Literatur verwendet. Zur ursprünglichen Begrifflichkeit vgl. Thaller, Manfred: «Vorüberlegungen für einen internationalen Workshop über die Schaffung, Nutzung und Verbindung grosser interdisziplinärer Quellenbanken in den historischen Wissenschaften.» In: Thaller, Manfred (Hg.): *Datenbanken und Datenverwaltungssysteme als Werkzeuge historischer Forschung*. St. Katharinen 1986 (Historisch-Sozialwissenschaftliche Forschungen 20), S. 9-30, hier bes. 10-12.

verschiedenen Quellen oft stark unterschiedlicher Informationsstruktur und in ebenso unterschiedlicher Informationsdichte beschrieben ist, potenziert sich das Problem: Eine Menge von einzelnen, in Informationsstruktur, -gehalt und -dichte unterschiedlichen Quellentypen, die ihrerseits eine a priori nicht festzulegende Menge von Strukturen aufweisen, muss gemeinsam verwaltet und analysiert werden. Als Ziel gilt das «virtuelle Archiv», eine Summe von sogenannten «Quellenbanken», Methodenbanken und historisch/hilfswissenschaftlichem Hintergrundwissen über die Quellen. Zusammenfassend lässt sich definieren, dass wir dann von quellenorientierter Datenverarbeitung sprechen, wenn nach ausführlicher Analyse von Informationsstruktur und Informationsgehalt quellenimmanente Strukturen oder Teilstrukturen in einer Datenbank abgebildet werden. Wenn die daraus entstehenden Quellenrepräsentationen durch die entsprechenden Repräsentationen von Hintergrundwissen ergänzt werden, sprechen wir von der Dokumentation historischen Quellenmaterials, wobei dieses Wissen durch Begriffsbildung aus den Quellenbelegen (möglichst transparent) zu entwickeln ist.¹⁶

Wo bleibt die (textkritische) Edition? Im Sinne der Historischen Fachinformatik könnte eine Definition wie folgt getroffen werden: Eine Quelle zu edieren bedeutet, ihre Informationsstruktur, ihren Informationsgehalt, ihre Informationsdichte und ihre Überlieferungsstruktur auf eine Weise zugänglich zu machen, die so vollständig als möglich sein soll und den Benutzer der Edition in die Lage versetzt, diese analytisch zu verwenden, ohne die eigentliche Quelle einsehen zu müssen. Damit wird der quellenorientierten Dokumentation zusätzlich das Element der Überlieferung beigefügt.¹⁷ Die Überlieferungssituation historischer Quellen jedweder Art entlang der Zeitachse begründet das diachrone Verfahren der Historischen Fachinformatik und in diesem Zusammenhang die Methode der «Integrierten Computergestützten Edition», die durch iterative Forschungs- bzw. Editionsprozesse und einem Teamwork von Archivaren, Hilfswissenschaft-

16 Vgl. Kropač, Konzeption, S. 115f. (siehe Anm. 1).

17 Zu einer formalen Betrachtungsweise des Überlieferungsprozesses und der daraus resultierenden Folgen vgl. Kropač, Ingo H.: «Processing the Delivered Past: From Primary Sources to Secondary Analyses». In: Hockey, Susan; Die, Nancy (Hgg.): *Research in Humanities Computing 4. Selected Papers from the ALLC/ACH Conference, Christ Church, Oxford, April 1992*. Oxford 1996, S. 156-170 – Kropač, Ingo H.: «Cognition, Imagination or Illusion? History From a Formal Point of View.» In: Borodkin, Leonid; Doorn, Peter (Hgg.): *Data Modelling – Modelling History. Proceedings of the XI International Conference of the Association for History and Computing, Moscow 1996*, Moskau 2000, S. 50-65.

tern, Historikern und Informationswissenschaftlern eine transparente Dokumentation und Edition zum Ziel hat.

Wie es gemacht wird: die Methode der «Integrierten Computergestützten Edition»

Das Ziel der Methode der integrierten computergestützten Dokumentation und Edition¹⁸ ist die Generierung eines Informationssystems, welches seinem Benutzer die Quellen in ihren unterschiedlichen Repräsentationsebenen, alle (automatisierten) Transformationsverfahren von einer Ebene zur nächsten und alle Entscheidungen und Interpretationen des Bearbeiters zur Wertung der Quelle zur Verfügung stellt. Die Verwaltung der unterschiedlichsten Repräsentationsformen der Quelle einerseits und des deklarativen und prozeduralen Wissens des Bearbeiters andererseits soll jene Transparenz gewährleisten, die der Benutzer benötigt, um auf intersubjektive Weise alle Entscheidungen des Bearbeiters rekonstruieren, überprüfen und gegebenenfalls korrigieren zu können. Unterstützt wird dies durch die gemeinsame und gleichzeitige Verwaltung der verschiedenen Repräsentationsebenen der Quelle und alles Wissens über dieselbe; das gilt nicht nur für den eigentlichen Text, sondern für alle Erschliessungen.

Unter Beachtung eines Modells, nach dem sämtliche Informationen mit Berücksichtigung aller Entscheidungen des Editors klassifiziert werden können, wurde ein System geschaffen, das Fakten als unmittelbare Information und Ableitungen als mittelbare Information in unterschiedlichen Datenbasen sowie Interpretationen als paradigmengestützte Information in einer speziellen Form von Wissensbasen repräsentiert. Dabei kommen derzeit je nach Quellentyp (Textsorte) drei unterschiedliche Strukturen zur Anwendung: für Urkunden und urkundliche Bücher, für vermischte Amtsbücher mit urkundlichem und seriellem Inhalt sowie für ausschliesslich serielle Quellentypen.

Der Bereich der urkundlichen Quellen bildete den Ausgangspunkt des Projektes, das von Beginn an den Anspruch erhob, hilfswissenschaftliche Forschung zu unterstützen. Aus diesem Grunde mussten alle jene Informationen abgebildet werden, die sowohl für diplomatische und kanzleigeschichtliche Studien als auch für Editionen relevant sind: äussere Merkmale, innere Merkmale und Überlieferung (siehe Abb. 4).

18 Siehe Anm. 4.

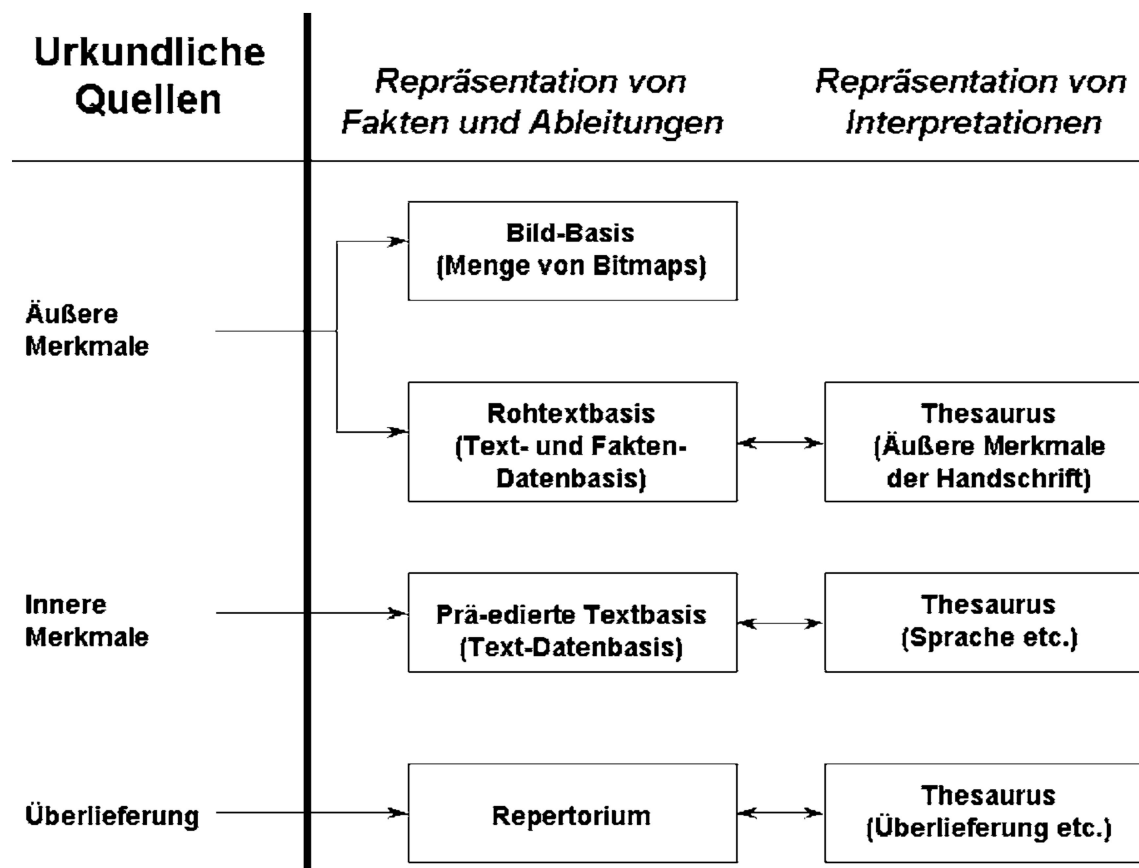


Abb 4: Abbildung urkundlicher Quellen

Die äusseren Merkmale einer Urkunde sind in drei Repräsentationsformen festgehalten:

1. Im digitalisierten Faksimile werden nahezu alle äusseren Merkmale der ursprünglichen Manuskripte in Form einer Menge von Bitmaps gezeigt.
2. Der transkribierte Text, der in einer «Rohtextbasis» verwaltet wird, gibt den Text möglichst originalgetreu wieder, was bedeutet, dass mit Hilfe eines Kodierungssystems¹⁹ Interpunktion, Abkürzungen, Korrekturen, wortspezifische Buchstabenformen etc. wiedergegeben werden.
3. Detaillierte Dokumentationen der Schrift, der schreibertypischen Buchstaben, des Textzustands, der Beglaubigungsmittel etc. sind ebenfalls in der «Rohtextbasis» abgespeichert.

Die inneren Merkmale hingegen werden in der «Prä-edierten Textbasis» verwaltet, wobei der Textbestand dem einer gedruckten Edition gleicht. Die Überlieferungssituation der Quellen ergibt sich aus einem Netz von Faktendatenbanken, die mit einem Archivrepertorium vergleichbar sind und den Lagerort, das Kopfrege und andere die Quelle indizierende

¹⁹ Vgl. Anmerkung 12.

Informationen wie etwa Belege zu Personen, Institutionen und Orten enthalten.

Damit ergeben sich für die Datenbasen der urkundlichen Quellen folgende Strukturmodelle:

«Rohtextbasis:

- D(okument): Zitat
- |
- |--- T(ext): Text
- |
- |--- R(ückseite): Text
- |
- |--- N(otiz): Notiz
- |
- |--- B(uchstabe): Buchstabe, Schrifttyp, Beschreibung, Besonderes
- |
- |--- L(igatur): Ligatur
- |
- |--- S(chriftduktus):Duktus

«Prä-edierte Textbasis:

- D(okument): Zitat
- |
- |--- T(ext): Invocatio, Intitulatio, Inscriptio, Arenga, Publicatio, Narratio, Dispositio, Sanctio, Corroboratio, Subscriptiones, Datierung

«Repertorium»:

- Dokument: Zitat, Datum, Ort, PDB
- |
- |--- Bearbeitung: Bearbeiter, Datum, Status
- |
- |--- Bild: Vorderansicht, Rückansicht, Text, Dorsal, Siegel
- |
- |--- Regest: Regest
- |
- |--- Person: Identifikation, Name, Beschreibung, Funktion, Funktionszahl, Status, Verwandtschaft
- |
- |--- Institution: Identifikation, Funktion, Typ, Name, Ort
- |
- |--- Ort: Identifikation, Beschreibung
- |
- |--- Objekt: Identifikation, Beschreibung

Ebenso wie die Abbildung urkundlicher Quellen erfolgt auch die serieller Quellen in mehreren Repräsentationsformen. In einer Bildbasis werden die digitalisierten Faksimiles der Quelle, denen eindeutige Namen zugewiesen wurden, verwaltet. Dabei stellt jedes Bild genau eine Seite der Quelle dar.

Bei der Quellengruppe der vermischten Amtsbücher mit überwiegend seriellen Charakter wird der (prä-edierte) Text der Quelle in einer «Textdatenbasis» abgebildet, indem der Wortlaut der einzelnen Einträge ohne weitere Spezifikation wiedergegeben wird.²⁰ Durch die Identifikations-

20 Dies bedeutet, dass im Gegensatz zu den urkundlichen Quellen die «neutrale» Form der reinen Transkription aus rein ökonomischen Gründen entfällt und der Text bereits implizit nach

nummer wird jedem Dokument das entsprechende Bild in der Bildbasis zugewiesen. Jeder Eintrag der Texdatenbasis wird durch eine fortlaufende Nummer gekennzeichnet, die wiederum der Eintragsnummer in der «Strukturdatenbasis» entspricht. Im Gegensatz zu den beiden vorgenannten Repräsentationen, welche die äusseren Merkmale der Quelle abbilden, werden hier die inneren Merkmale beschrieben, so etwa die Art des Eintrags, die in ihm genannten Orte und Personen etc. Diese Datenbasis dient zur Erstellung der Personen-, Orts- und Sachkataloge und des Glossars. Erst auf dieser Ebene finden Erschliessung und Analyse statt, Bildbasis und Texdatenbasis dienen «nur» der Darstellung der Quelle (siehe Abb. 5).

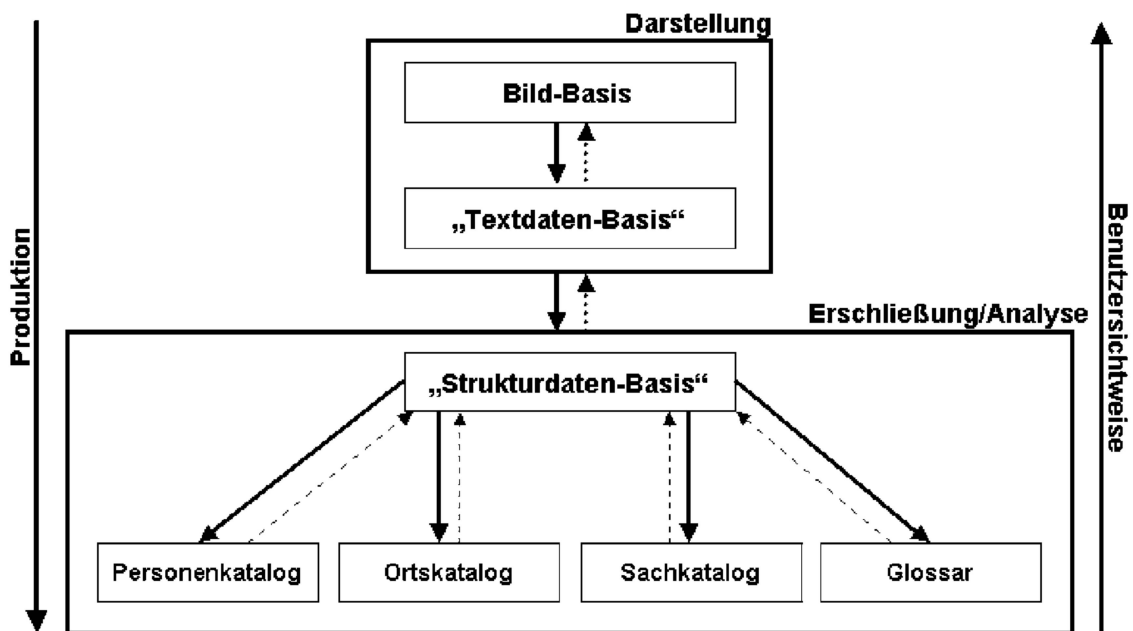


Abb 5: Abbildung serieller Quellen 1

Die Struktur vermischter Amtsbücher mit überwiegend seriellem Inhalt sieht demnach folgendermassen aus:

festgelegten Editionsregeln aufgenommen wird. Urkundliche Bücher und Rechtsbücher werden dagegen wie urkundliche Quellen behandelt.

«**Textdatenbasis:**

Seite: Nummer, Bild
|
|---Eintrag: Nummer, Text

«**Strukturdatenbasis:**

Textabschnitt: Identifikation, Typ, Regest
|
|----Eintrag: Nummer, Art, Streichung
|
|----Person: Identifikation, Vorname, Familienname, Beruf, Funktion, Eintragszeit
|
|----Abgabe: Menge, Objekt, Einzelabgabe, Gesamtabgabe, Eintragszeit
|
|----Ort: Identifikation, Bezeichnung
|
|----Terminus: Identifikation, Bezeichnung

Im Gegensatz zu diesen eher komplexen seriellen Quellen werden bei der Abbildung von Bürger- und Rechnungsbüchern Textdatenbasis und Strukturdatenbasis in eine Datenbasis integriert; neben der Bildbasis existiert also nur eine Repräsentationsform, die sowohl den transkribierten Text als auch die hierarchisch untergeordnete Darstellung der inneren Merkmale beinhaltet (siehe Abb. 6).

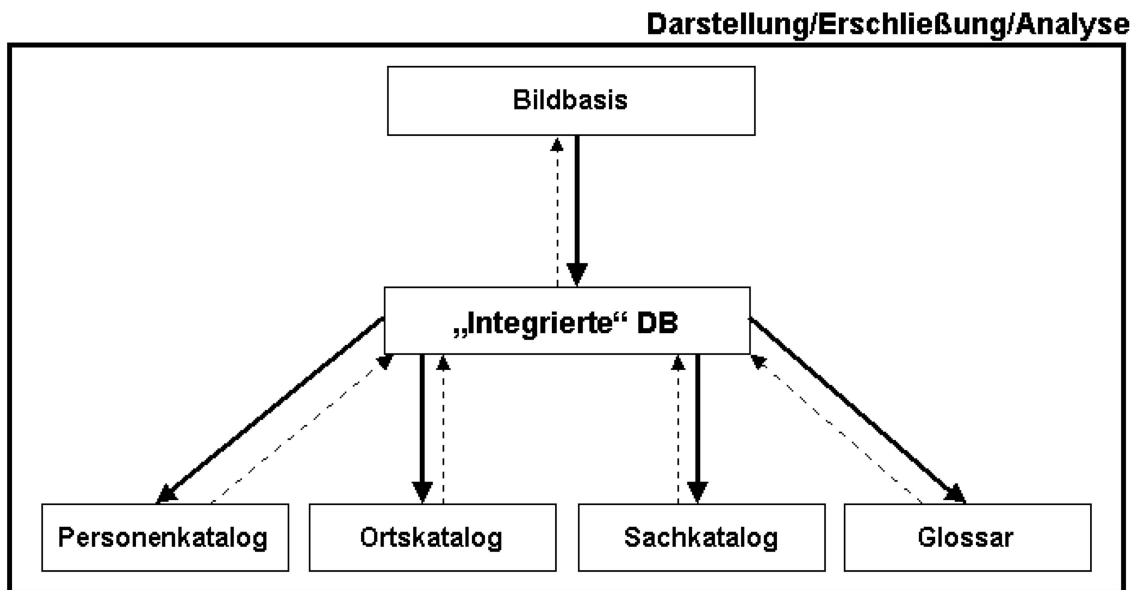


Abb 6: Abbildung serieller Quellen 2

Somit geschehen Darstellung, Erschließung und Analyse der Quelle auf der gleichen Ebene, woraus sich folgendes Strukturmodell ergibt:

«Integrierte» Datenbasis:

- | Rubrik: Rubrik-Identifikation, Titel
- | |
- | |----Satz: Satz-Identifikation, Text
- | | |
- | | |----Ausgaben f. Person: AP-Identifikation, Art, Anzahl, Empfänger, Einzelausgabe,
| | | Gesamtausgabe, Zweck, Tätigkeit, Ziel, Datum, Dauer
- | | |
- | | |----Ausgaben f. Sachen: AS-Identifikation, Art, Menge, Objekt, Einzelausgabe,
| | | Gesamtausgabe, Zweck, Nehmer, Datum, Dauer
- | | |
- | | |----Einnahme v. Person: EP-Identifikation, Art, Anzahl, Empfänger, Einzeleinnahme,
| | | Gesamteinnahme, Zweck, Tätigkeit, Ziel, Datum, Dauer
- | | |
- | | |----Einnahme v. Sachen: ES-Identifikation, Art, Menge, Objekt, Einzeleinnahme,
| | | Gesamteinnahme, Zweck, Geber, Datum, Dauer
- | | |
- | | |----Satzsumme: Summe
- | | |
- | | |----Randnotiz: Text
- | | |
- | |----Rubriksumme: Text, Summe

Für jede Urkundensammlung und jeden Kodex, von der Urkundenreihe bis zum städtischen Amtsbuch, werden solche «Kernsysteme» aus logisch miteinander verbundenen Modulen errichtet. Eine darüber liegende Verknüpfung der unterschiedlichen Erschliessungsmodule untereinander bewirkt in der Folge ein Netzwerk von Informationen aus miteinander in Beziehung stehenden Quellen, ohne die Bestände, aus denen diese Quellen ursprünglich stammen, zu verändern oder in ihrer Ordnung zu zerstören. Kataloge von Deskriptoren und Registern erschliessen und ergänzen diese Basismodule. Abb. 7 zeigt ein Schema des Gesamtsystems beispielhaft nur für urkundliche Quellen, wobei die eigentliche Edition im Sinne eines Anwendersystems zwischen dem Kernsystem, aus dem sie abgeleitet wird, und dem virtuellen Archiv, aus dem die Informationen über die Überlieferungssituation der Einzeltexte stammt, anzuordnen ist. Dies gilt analog natürlich auch für serielle oder gemischte Quellentypen. Die Gesamtheit *aller* Quellenrepräsentationen und die Kumulierung *aller* verfügbarer Kataloge ergeben im Endausbau ein virtuelles Archiv zum Betrachtungsgegenstand, in unserem Falle zur Geschichte der Regensburger Bürgerschaft im Spätmittelalter.

Dieses Konzept strebt nicht das monolithische Nebeneinander von wie auch immer gespeicherten Bildern und fertig edierten Texten an, sondern ein integriertes Informationssystem, in dem Bilder und Texte logisch verknüpft und dynamisch verwaltet werden und die Abbildung sämtlicher Bearbeitungs- und Realisierungsstufen den Weg vom digitalisierten Bild

zum prä-edierten Text dokumentiert, ohne dass dabei die originalen Bilder oder einzelne Textstufen verloren gehen. Es gestattet ein provenienzorientiertes Vorgehen, wobei es methodisch wie auch arbeitstechnisch unerheblich ist, ob aus Beständen des virtuellen Archivs ein Kernsystem erarbeitet oder ein Kernsystem in das virtuelle Archiv eingegliedert wird. Aufgrund dieser Dynamik kann auch den jeweilig vorhandenen Ressourcen, ob nun speziell ausgebildete Forscher oder zur Verfügung stehende Geldmittel, voll entsprochen werden.

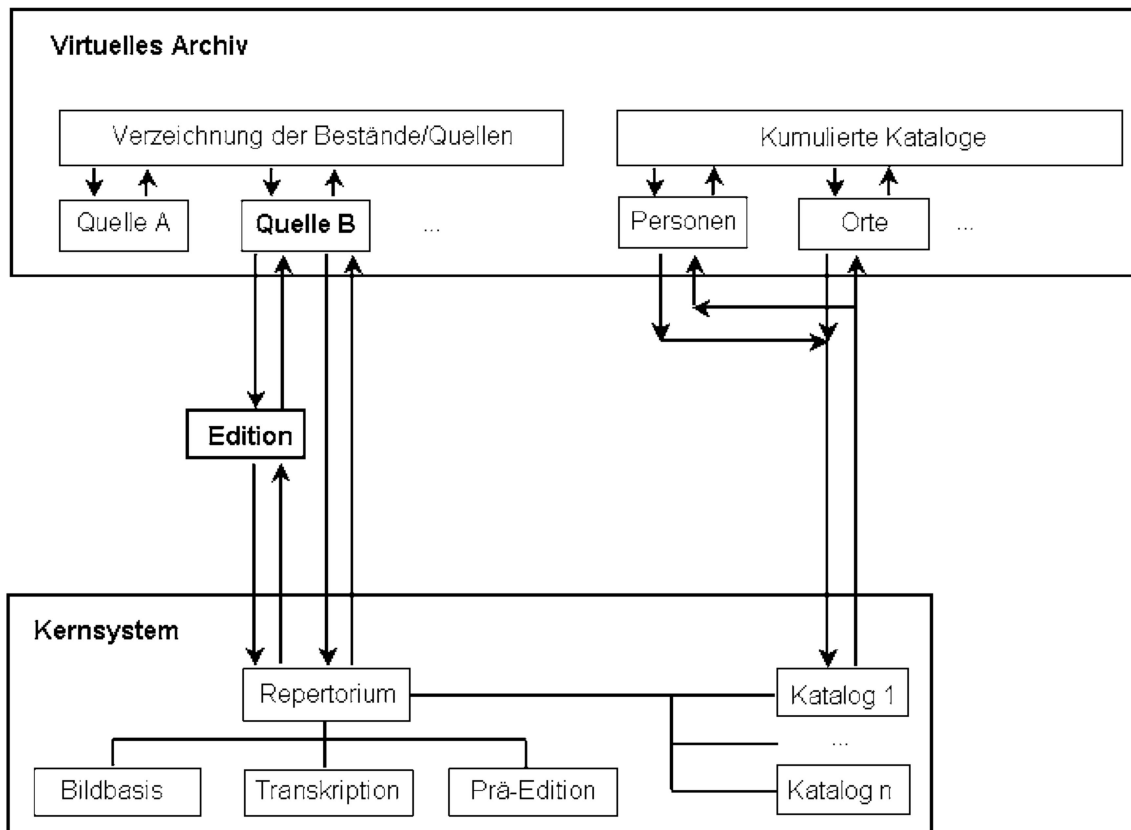


Abb 7: Schema des Gesamtsystems unter Berücksichtigung der urkundlichen Quellen

Damit erhebt sich letztendlich die Frage nach der für ein solches System geeigneten Software. Obwohl durch einen durchaus angestrebten hohen Abstraktionsgrad der Einzelmodule Aufbau und Inhalt des Systems trivial erscheinen mögen, ist dies bei der konkreten Abbildung der Quellen nicht mehr so, und jeder Spezialfall benötigt eine spezielle Lösung. Nachdem historische Quellen – wie auch die meisten Tatsachen des «normalen» Lebens – nicht in Tabellen organisiert sind und auch nur unter Verzicht auf zentrale Informationsgehalte in solchen modelliert werden können (wenn auch Verfechter der relationalen Datenbanken nicht müde werden, das

Gegenteil zu behaupten), brauchen wir erheblich komplexere Datenmodelle und Softwarelösungen. Das skizzierte Kernsystem der FCR basiert deshalb auf einem speziell für die quellenorientierte Datenverarbeitung in den historischen Wissenschaften entwickelten und unter dem Namen «Kleio» bekannten Programmpaket.²¹

Diese speziell für die Bedürfnisse von Historikern entwickelte Software ist ein Ergebnis des «Historical Workstation Project» und wurde von M. Thaller zuerst am Max-Planck-Institut für Geschichte in Göttingen in Zusammenarbeit mit mehreren europäischen Forschungsinstitutionen entwickelt. Es steht wissenschaftlichen (nicht kommerziellen) Anwendern kostenlos zur Verfügung. Diese Software bietet eine Reihe von Besonderheiten, die bei kommerziellen Datenbankprogrammen nicht zu finden sind, jedoch der Art von Informationsverarbeitung entspricht, die typisch für Anwendungen im Bereich der Historischen Fachinformatik ist:

1. Für «Kleio» wurde ein kontextsensitives Datenmodell entwickelt, das seinerseits auf dem Konzept semantischer Netze basiert. Für den Benutzer ergibt sich je nach Anwendung eine hierarchische, eine netzwerkorientierte oder in speziellen Fällen eine «relationale» Sichtweise des Datenmodells, das Programm ist «von sich aus» in der Lage, bestimmte Kontexte der Quelle zu berücksichtigen.
2. Das Programm bietet datentechnische Hilfsmittel, der extremen Variabilität historischen Quellenmaterials und dessen inhärenten Unschärfen effizient zu begegnen. So können etwa sowohl stark strukturierte als auch weitgehend unstrukturierte Daten verwaltet, Konsistenzprobleme gelöst und definitorische Unschärfen durch die Implementation der «fuzzy-set-logic» disambiguiert werden.
3. In historischen Quellen finden sich eine Reihe von Informationen, die durch die klassischen Datentypen nicht fassbar sind oder nicht unmittelbar interpretiert werden können. Daher stellt «Kleio» Mittel bereit, solche Informationen in ihren Merkmalen zu definieren und so neben den Quellen auch Wissen über sie in sogenannten «logischen Objekten» zu verwalten. Dazu zählen etwa Definitionsumgebungen für erweiterte Datentypkonzepte, der Ausgleich orthographischer Varianten, die Lemmatisierung des Lateinischen, numerische und nichtnumerische Klassifikationssysteme, definierbare Verbindungen zwischen Entitäten

21 Vgl. Thaller, Manfred: «Kleio. A Database System.» St. Katharinen 1993 (Halbgraue Reihe zur Historischen Fachinformatik B 11). – Woollard, Matthew; Denley, Peter: «Source-Oriented Data Processing for Historians: a Tutorial for «Kleio».» St. Katharinen 1993 (Halbgraue Reihe zur Historischen Fachinformatik A 23).

mehrerer Datenbasen (Nominal Record Linkage), die Verwaltung nicht-dezimaler Mass- und Währungsangaben, die auch temporären Schwankungen unterworfen sein können oder auch die Interpretation historischer Datierungsangaben.

4. Daneben enthält «Kleio» auch hochspezialisierte Komponenten wie etwa Mittel zur automatisierten Produktion von thematischen Karten, ein Modul zur digitalen Bildverarbeitung (das sogenannte «Image Analysing System») sowie Interfaces zu Statistikprogrammen, zum Satzprogramm «TeX» und zu Auszeichnungssprachen (HTML/XML).
5. In den letzten Entwicklungsphasen entstand eine Programmversion, die eine Schnittstelle zu HTTP-Servern bietet und damit den direkten Datenbank-Zugriff aus dem World Wide Web ermöglicht.

Insgesamt betrachtet, präsentiert sich «Kleio» nicht nur als bloße Datenbank-Software (DBMS), wenn es auch schon in diesem Bereich für Historiker wesentlich mehr zu leisten imstande ist als herkömmliche kommerzielle Produkte. Darüber hinaus bietet die Software die Möglichkeit, komplexe Informationssysteme aufzubauen und deren Leistungsfähigkeit aufgrund der offenen Systemarchitektur ständig zu erweitern.

Damit ermöglicht «Kleio» sowohl die Verwaltung des mittlerweile bereits ziemlich komplexen Informationssystems als auch sämtliche notwendigen Analyseprozesse sowie die Migration der Inhalte in (statische) Web-Systeme und den dynamischen Zugriff aus Web-basierten Systemen unmittelbar auf das Kernsystem. Bis auf die eigentliche Digitalisierung werden damit alle Schritte des iterativen Editions- und Dokumentationsprozesses unterstützt.

Zusammenfassung

Die Methode der Integrierten Computergestützten Edition beschreibt einen iterativen und integrativen Weg von der Dokumentation historischen Quellenmaterials zur dessen (textkritischer) Edition. Als höchstes Ziel gilt dabei die Transparenz der dokumentarischen bzw. editorischen Arbeitsprozesse, die durch Modularität und Offenheit des Systems gewährleistet wird; Experten- und Anwendersysteme stehen zumindest aus der Sichtweise der Benutzer getrennt nebeneinander. Zweck dieser Methode ist letztendlich die Initialisierung des auf Quellen beruhenden historischen Forschungsprozesses unter Beachtung der Methodenpluralität, wenn auch die Arbeit mit Quellen in der heutigen historischen Forschung weit hinter einer spekulativen Geschichtsbetrachtung zurückzuliegen scheint.

Die Edition grosser Quellenkorpora kann u.E. nur unter Anwendung der so genannten «neuen Medien» und unter Beachtung informationswissenschaftlicher Grundsätze effizient und ökonomisch durchgeführt werden. Dazu bedarf es nicht nur interdisziplinärer Ansätze, sondern eines intensiven Teamworks von Archivaren, Hilfswissenschaftlern, Historikern und Informationswissenschaftlern. Die Trennung von Quellen- und Wissensrepräsentation und die Unterstützung iterativer Forschungsprozesse soll jene Transparenz gewährleisten, die sicherstellt, dass unser schriftliches Kulturerbe durch digitale Faksimiles gesichert und der Zugang zu den Quellen zunehmend demokratisiert werden kann.