

Die Prosopographische Datenbank zur Geschichte der südöstlichen Reichsgebiete bis 1250 (PDB): Konzepte und Kurzdokumentation

INGO H. KROPAČ UND PETER BECKER

Ohne hier auf eine nähere Diskussion des Begriffs des Alltäglichen in der Geschichte beziehungsweise dessen Stellenwert in der Mediävistik eingehen zu wollen¹ soll die Grazer Prosopographische Datenbank (künftig: PDB) auch als Werkzeug zur Erforschung des mittelalterlichen Alltags vorgestellt werden. Durch den Anspruch, alle Personen, die in der Zeit bis 1250 in der vorgegebenen Region urkundlich genannt sind, zu erfassen, wird dem Benutzer die Möglichkeit geschaffen, neben den häufig vertretenen Mitgliedern der privilegierten Bevölkerungsgruppen auch die Unterschichten, die besonders in den Privaturkunden des Hochmittelalters vorkommen, gezielt zu analysieren. Damit wird auch jenen Vertretern der Alltagsgeschichte, die diese mit einer Geschichte von unten verbinden, Material bereitgestellt. Weiters wird es etwa möglich, das Interaktionsfeld der Gruppe der Hofbeamten über einen längeren Zeitraum hinweg zu studieren. Dadurch kann diese Datenbank zum wichtigen Hilfsmittel teils für begleitende Untersuchungen, teils für eigenständige Analysen werden.

Es liegt in der Natur von Werkzeugen, nicht Selbstzweck zu sein, sondern benutzt zu werden². Aus diesem Grund wendet sich diese Vorstellung bewußt an Fachhistoriker als potentielle Benützer, wobei versucht wird, weitgehend ohne informatische Fachterminologie auszukommen. Welche Inhalte in der Datenbank aufzufinden sind, wie sie organisiert sind und mit welchen Hilfsmitteln sie ausgewertet werden können, soll im folgenden beschrieben werden.

¹ Vgl. dazu Harry Kühnel: *Zum Geleit: Alltag und Fortschritt im Mittelalter*. In: *Alltag und Fortschritt im Mittelalter*. Internationales Round-Table-Gespräch, Krems an der Donau 1984. Wien 1986 (Veröffentlichungen des Instituts für mittelalterliche Realienkunde Österreichs 8), S. 5 - 6; Hans-Werner Goetz: *Alltag im Mittelalter*. In: *Archiv für Kulturgeschichte* 67, 1985, S. 207 - 225; Dietrich Schwarz: *Mittelalterliche Realienkunde: Standortbestimmung und Methode der Forschung*. In: *Die Erforschung von Alltag und Sachkultur des Mittelalters. Methode - Ziel - Verwirklichung*, Krems an der Donau 1982. Wien 1984 (Veröffentlichungen des Instituts für mittelalterliche Realienkunde 6), S. 9 - 13. - Zur Typisierung der Beiträge im Rahmen einer mittelalterlichen Alltagsgeschichte siehe die Studie von Manfred Thaller in diesem Band.

² Zu den grundlegenden Überlegungen, die speziell auch für die PDB gelten vgl. Ingo H. Kropač: *Vorschläge zur Standortbestimmung einer Historischen Fachinformatik*. In: *Die Geschichte und ihre Quellen. Festschrift für Friedrich Hausmann zum 70. Geburtstag*, hrsg.v. Reinhard Härtel et al. Graz 1987, S. 601 - 610. Zur Rolle von Quellenbanken in der historischen Forschung vgl. Manfred Thaller: *Vorschlag für einen internationalen Workshop über internationale Quellenbanken*. In: *Datenbanken und Datenverwaltungssysteme als Werkzeuge historischer Forschung*, hrsg.v. Manfred Thaller. St. Katharinen 1986 (Historisch-Sozialwissenschaftliche Forschungen 20), S. 9 - 30. - Zum Werkzeugcharakter solcher Datenbanken und ihrer Einbindung in eine *Historical Workstation* siehe den Beitrag von M. Thaller in diesem Band.

1. GRUNDLEGENDE KONZEPTE

1.1. Zur Entstehung der Datenbank.

Ursprünglich wurde die PDB nicht als Hilfsmittel zur Erforschung des Alltagslebens im Mittelalter konzipiert, sondern vielmehr zur Unterstützung zweier großer Editions-vorhaben – das *Urkundenbuch der Steiermark und ihrer Regenten* (Bearbeitung: Friedrich Hausmann) und das *Urkundenbuch des Patriarchats Aquileia* (Bearbeitung: Reinhard Härtel und Friedrich Hausmann), – am Forschungsinstitut für Historische Grundwissenschaften an der Karl-Franzens-Universität Graz, die mit Hilfe des Computers durchgeführt werden³. Dabei sollten die in der Datenbank enthaltenen Personen vor allem für Datierungszwecke und Echtheitsüberprüfung im Rahmen des *discrimen veri ac falsi* eingesetzt werden. Dieser Wunsch war die Geburtsstunde der *Prosopographischen Datenbank zur Geschichte der südöstlichen Reichsgebiete bis 1250*. Im Rahmen eines Forschungsprojekts⁴ wurde die Arbeit an der Datenbank in Angriff genommen, die nun ein Stadium erreicht hat, das Vorstellung und Beschreibung als wünschenswert erscheinen läßt. Die PDB wird durch ihre neutrale Konzeption nicht nur zum Hilfsmittel für Datierungsanliegen, sondern kann in vielerlei Hinsicht von der historischen Forschung verwendet werden, eben auch zur Analyse bestimmter Aspekte mittelalterlichen Alltagslebens.

Daraus ist klar erkennbar, daß es sich bei der PDB nicht um eine Prosopographie⁵ an sich handelt, sondern um eine Materialsammlung sämtlicher urkundlich genannter Personen, ohne Einschränkung auf eine bestimmte Schicht oder Gruppe. Der gleiche Grundgedanke findet sich auch bei ähnlich gelagerten Forschungsprojekten⁶, die jedoch zum Teil

³ Zum Konzept der maschinellen Edition und deren Voraussetzungen vgl. Reinhard Härtel und Ingo H. Kropač: Edition und Auswertung mittelalterlicher Urkunden. Probleme bei Standardisierung und Transfer fortlaufender Texte. In: *Datennetze für die Historischen Wissenschaften? Probleme und Möglichkeiten bei Standardisierung und Transfer maschinenlesbarer Daten*, hrsg.v. Friedrich Hausmann et al. Graz 1987, S. 100 - 112.

⁴ Das vom Bundesministerium für Wissenschaft und Forschung unterstützte und nunmehr abgeschlossenen Projekt trägt den Titel *Untersuchung der Einsatzmöglichkeiten der EDV für die historische Fak-tendokumentation im Bereich der Historischen Grundwissenschaften* und steht unter der Leitung von Univ.-Prof. Dr. Friedrich Hausmann. Zu den konzeptuellen Grundlagen und allgemeinen Beschreibungen der Teilprojekte vgl. Ingo H. Kropač: Computergestützte Projekte am Forschungsinstitut für Historische Grundwissenschaften der Karl-Franzens-Universität Graz. In: *Fakten – Daten – Zitate* 6/1, (Wien) 1986, S. 9 f.

⁵ Es ist hier nicht der Ort, eine umfassende Diskussion des Begriffs Prosopographie abzuführen. Zur Begrifflichkeit und zu den besonderen Problemen der Prosopographie in der Mediaevistik vgl. Karl Schmid: Prosopographische Forschungen zur Geschichte des Mittelalters. In: *Aspekte der historischen Forschung in Frankreich und Deutschland*, hrsg. von Gerhard A. Ritter und Rudolf Vierhaus. Göttingen 1981 (Veröffentlichungen des Max-Planck-Instituts für Geschichte 69), S. 54 - 78. – Zum Stand der Forschung im österreichischen Raum vgl. Gerhard Jaritz und Albert Müller: *Medieval Prosopography in Austrian Historical Research: Religious and Urban Communities*. In: *Medieval Prosopography* 7/1, 1986, S. 57 - 86.

⁶ Hier seien als Beispiele nur die Freiburger Datenbank mittelalterlicher Personen und Personengruppen, die Prosopographia Orbis Latini (PROL) und das Prosopographische Lexikon der Palaiologenzeit genannt. Zu diesen Projekten vgl. Karl Schmid: Zum interdisziplinären Ansatz, zur Durchführung und zum Anliegen des Fulda-Projekts. In: *Die Klostersgemeinschaft von Fulda im früheren Mittelalter*, hrsg. v. Karl Schmid.

auf anderen Quellenbeständen basieren oder unterschiedliche Zielsetzungen verfolgen. Somit versteht sich die PDB sehr wohl als Vorstufe oder Hilfsmittel zu einer Prosopographie des Südostens des Römisch-deutschen Reichs im Hochmittelalter, soll aber darüber hinaus auch Funktionen für andere historische Disziplinen erfüllen können.

1.2. Inhalt und Zielsetzung.

Das bereits angesprochene neutrale Erfassungskonzept der Datenbank bedingt eine Erfassung sämtlicher in urkundlichen Quellen genannter Personen samt ihren unmittelbaren Beschreibungen. Wegen der nahezu unüberschaubaren Menge an Quellen wurde der Zeitraum der Erfassung auf das Hochmittelalter eingegrenzt, die geringe Überlieferung aus dem Frühmittelalter wird jedoch ebenfalls berücksichtigt. Neueste Forschungen⁷ haben gezeigt, daß durch die relativ hohe regionale Mobilität verschiedener sozialer Schichten eine Eingrenzung auf einzelne Regionen nicht sinnvoll sein kann. Um zumindest dem Lebens- und Wirkungsbereich des Adels zu entsprechen – diese Schicht ist wohl die in Urkunden meistgenannte – müssen neben den Personen der zur Bearbeitung anstehenden Regionen – die Steiermark und das Patriarchat Aquileia – auch jene des gesamten österreichischen Raumes ohne Vorarlberg sowie Personen aus Teilen Bayerns, des heutigen Slowenien, Istriens und Venetiens Aufnahme in die Datenbank finden.

Trotz dieser Eingrenzung in räumlicher, zeitlicher und auch quellentypologischer Hinsicht erfordert allein die Erfassung aller Personennennungen einen so großen Arbeitsaufwand, daß die Errichtung der Datenbank "nur" für die beiden Editionsprojekte wirtschaftlich kaum zu rechtfertigen wäre. Daneben erfordert auch die Verwaltung eines solchen Unternehmens in datentechnischer Hinsicht oder die Entwicklung spezieller Hilfsmittel zur Auswertung (etwa zur Identifizierung von Einzelpersonen in Suchprofilen) einen nicht unwesentlichen Bedarf an Arbeitskapazität. Aus diesen Gründen muß eine solche Datenbank so konzipiert sein, daß sie nicht nur den Anforderungen eines Grundwissenschaftlers genügt, sondern die Bedürfnisse der historischen Forschung auf möglichst breiter Basis abdeckt. Forscher aus dem Bereich der Rechts- und Sozialgeschichte, der Landes- und vergleichenden Regionalgeschichte, den verschiedenen historischen Grundwissenschaften, aber auch nicht zuletzt der Alltagsgeschichte – um nur einige Fachgebiete anzuführen – sollen in die Lage versetzt werden, nach den in ihrer Disziplin verwendeten Methoden mit der PDB arbeiten zu können.

Diese Vorgaben stellen nicht nur hohe Ansprüche an Design⁸ und Organisation der

Bd. 1, München 1978 (Münstersche Mittelalter-Schriften 8/1), S. 11 - 36. – Karl F. Werner: Problematik und erste Ergebnisse des Forschungsvorhabens "PROL" (Prosopographia Regnorum Orbis Latini). Zur Geschichte der ost- und mitteleuropäischen Oberschichten bis zum 12. Jahrhundert. In: Quellen und Forschungen aus italienischen Archiven und Bibliotheken 57, 1977, S. 69 - 87. – E. Trapp: Probleme der Prosopographie der Palaiologenzeit. In: Jahrbuch der österreichischen Byzantinistik 27, 1978, S. 181 - 201.

⁷ Das gilt in hohem Maß für den Adel im Südosten des alten Reichs, reicht aber auch darüber hinaus. Vgl. dazu jüngst Friedrich Hausmann: Carinsiani e Stiriani in Friuli. In: Il Friuli dagli Ottoni agli Hohenstaufen, hrsg. v. Guiseppa Fornasir, Udine 1984, S. 547 - 596, sowie Heinz Dopsch: Salzburg und Aquileia. In: ebd., S. 509 - 545.

⁸ Darunter ist in diesem Fall die gesamte Gestaltung der Datenbank zu verstehen. Das bedeutet, daß Bedürfnisse und Vorgaben analysiert werden müssen, damit eine Entscheidung für jene Form, in der die

Datenbank als solcher, sie setzten auch strenge und durchdachte Konzeptionen auf dem ureigensten Gebiet des Historikers voraus, nämlich der Quellenkritik und -analyse. Deren Ergebnis ist die unmittelbare Beschreibung einer Person, wie sie uns in der Quelle entgegentritt (s.u. Punkt 3). Wenn Informationsgehalt und -struktur der Quelle zum speziellen Phänomen – in diesem Falle der einzelnen Person – abgeklärt und gesichert ist, muß ein nächster Schritt folgen, der wieder eine Vielzahl von Problemen aufwirft: die Abbildung der Informationen in der Datenbank. Um unterschiedlichste Auswertungen und Auswertungsmethoden zu gewährleisten, wurde die neutrale Aufbereitung der Daten zum obersten Prinzip gemacht, d.h., "falls eine Reinkarnation Rankes den Wunsch haben sollte, sie zu benutzen, so sollten sie von der Konzeption her genauso dazu in der Lage sein, wie wenn einer der Cliometriker der 60er Jahre den selben Wunsch haben sollte"⁸. Dies wird durch die Abbildung jener Informationsstruktur, welche die Quelle über eine Person enthält, in ein formales System erreicht, welches seinerseits ausschließlich den Informationsgehalt der Quelle widerspiegelt. Auf diese Weise entsteht ein abstraktes Modell, das gleichsam als Axiom für das formale System der Datenbank steht.

1.3. Design der Datenbank.

Die umfangreichen Vorgaben, die an die PDB gestellt werden müssen, lassen sich in einige wenige Punkte zusammenfassen:

- Wegen des hohen Arbeitsaufwands muß die Eingabearbeit so ökonomisch als möglich organisiert sein.
- Die Informationen müssen unmittelbar und neutral aus der Quelle in die Datenbank übertragen werden.
- Für Forschungen, denen der Informationsgehalt der Datenbank nicht ausreicht, muß die Möglichkeit geschaffen werden, entweder weitere Informationen in die Datenbank einzubringen oder sie mit einer weiteren, anders strukturierten Datenbank, welche die zusätzlichen Informationen enthält, zu verknüpfen. Dies muß sowohl für formalisierte Quellenteile als auch für Volltexte (beispielsweise ein Regest oder eine Urkunde, bzw. Teile davon) gelten.
- In der Datenbank genügt es nicht, herkömmliche Methoden des Data Retrieval mit Suchprofilen nach der Bool'schen Logik zu verwenden (s.u., Punkt 2.2.), zusätzlich müssen die Unschärfe der aus historischen Quellen extrahierbaren Kategorien und die unterschiedlichen Schreibweisen derselben berücksichtigt werden können¹⁰.

Informationen organisiert sein sollen, gefällt werden kann und notwendige Bearbeitungen durchführbar sind. Auch die Nutzung der Datenbank in Hinsicht auf eine bestimmte Hard- und Softwaresituation sowie eine Zielgruppenanalyse sind dem Design zuzurechnen.

⁹ Manfred Thaller: Praktische Probleme bei der interdisziplinären Untersuchung von Gemeinschaften "langer Dauer". In: Aspekte der historischen Forschung in Frankreich und Deutschland, hrsg. v. Gerhard A. Ritter und Rudolf Vierhaus. Göttingen 1981 (Veröffentlichungen des Max-Planck-Instituts für Geschichte 69), S. 177.

¹⁰ Zu den Problemen bei der datentechnischen Bearbeitung historischen Quellenmaterials und deren Lösungsmöglichkeiten vgl. Manfred Thaller: Ungefähre Exaktheit. Theoretische Grundlagen und praktische Möglichkeiten einer Formulierung historischer Quellen als Produkte 'unscharfer' Systeme. In: Neue Ansätze in der Geschichtswissenschaft, hrsg. v. Herta Nagl-Docekal und Franz Wimmer. Wien 1984 (Conceptus-Studien 1), S. 77 - 100. – Manfred Thaller: The Winds of Change. Problems of a Databank

- Die Datenbank oder Teile daraus müssen in ein numerisches Codesystem übertragbar sein, um eine Datenbasis zu erhalten, die von den bekannten Programmpaketen¹¹ für statistische Analysen nach quantitativen Methoden bearbeitbar ist.
- Die Datenbank muß so organisiert sein, daß ihre Benützung auch auf kleineren Rechnersystemen wie etwa mit Festplatten ausgebauten PCs durchführbar ist.

Die Einlösung der eben aufgestellten Forderungen ist nur durch das wechselseitige Zusammenwirken zweier Komponenten möglich:

1. einem speziellen Design einer solchen Datenbank, das den konzeptionellen Vorgaben anzupassen ist, und
2. der Software – genauer: des Programmpakets –, die sowohl das Design verwirklichen als auch die anderen nicht geringen Anforderungen bewältigen kann.

Die PDB kann aus den oben angeführten Gründen keine Groß-Datenbank im herkömmlichen Sinne sein, die erhebliche Anforderungen an die Leistungsfähigkeit einer Rechenanlage an Rechenzeit, Speicherplatz und auch an Kosten stellt und demnach nur auf Großrechnern lauffähig wäre, sondern sie besteht zunächst aus einem Bündel einzelner Datenbanken, die jeweils den Inhalt eines bestimmten Urkundenbuchs oder eines Archivbestands repräsentieren. Diese als *Primärdatenbanken* bezeichneten Einheiten, die es in der Folge zu dokumentieren gilt (s.u.) stellen die Basis der PDB dar und sind nach dem Prinzip der quellenorientierten Datenverarbeitung¹² organisiert und erfaßt. Durch Verknüpfung mehrerer *Primärdatenbanken* können in der Folge *Sekundärdatenbanken* erzeugt werden, die ein völlig unterschiedliches Aussehen haben können und ihre Entstehung ebenso unterschiedlichen Erwägungen verdanken. Dabei sind im wesentlichen drei Möglichkeiten zu unterscheiden:

1. Durch die Verknüpfung von *Primärdatenbanken* gleicher Struktur im Sinne einer bloßen Aneinanderreihung entsteht die einfachste *Sekundärdatenbank*, ein kumulativer Index der Eintragungen. Ihr Informationsgehalt entspricht der Summe der Informationsgehalte der verbundenen *Primärdatenbanken*.
2. Wenn ein Forschungsprojekt aufgrund einer spezifischen Problemstellung zusätzliche Informationen einbringen will, können die *Primärdatenbanken* mit völlig anders strukturierten Datenbanken¹³ unter Zuhilfenahme einer gemeinsamen Information ebenfalls

Oriented System Using the Concept of Fuzzy Sets. Grenoble 1981 (Unpublished Paper presented at the 1981 Joint Conference of IFDO and IASSIST). – Vorschläge, die zu einer verbindlichen Lösung spezifischer Übertragungsprobleme führen könnten, finden sich bei Manfred Thaller: A Draft Proposal for a Standard for the Coding of Machine Readable Sources. In: *Historical Social Research* 40, 1986, S. 3 – 46; Manfred Thaller: Auf dem Weg zu einem Standard für maschinenlesbare Quellen. In: *Datennetze für die Historischen Wissenschaften? Probleme und Möglichkeiten bei Standardisierung und Transfer maschinenlesbarer Daten*, hrsg.v. Friedrich Hausmann et al. Graz 1987, S. 228 – 238.

¹¹ Hier seien vor allem SPSS^x (Statistical Package for the Social Sciences) und SAS (Statistical Analysis System) genannt. Vgl. dazu SAS User's Guide: Basics, Version 5 Edition. Cary 1985. – SPSS^x User's Guide. New York [etc.] 1983.

¹² Zum Begriff der quellenorientierten Datenverarbeitung und den dahinterstehenden Konzepten vgl. Thaller, Ungefähre Exaktheit, S. 82 f. sowie Thaller, Praktische Probleme, S. 180 f.

¹³ Ein Beispiel für eine derartige Anwendung bildet die Verknüpfung mit Volltext-Datenbanken, wie sie im

zu einer *Sekundärdatenbank* verschmolzen werden. Auf solche Weise entstehende *Sekundärdatenbanken* sind allerdings keine Bestandteile der PDB und werden in ihr auch nicht verwaltet. Diese Möglichkeit eröffnet jedoch den verschiedensten Forschungsprojekten eine sinnvolle problemorientierte Arbeit mit Beständen der PDB.

3. Weitere *Sekundärdatenbanken*, die sehr wohl Bestandteil der PDB sind, werden aus *Primärdatenbanken* entwickelte *Meta-Datenbanken* sein, die ihrerseits aus einem nach identifizierten Einzelpersonen geordneten prosopographischen Katalog bestehen. Hier sind ebenfalls zwei Schritte zu unterscheiden: Zum ersten muß die jeweilige Primärdatenbank "in sich selbst" verknüpft werden, um die Einzelpersonen zu identifizieren und die Belegstellen ihrer Nennung zu ordnen¹⁴, zum zweiten müssen auch die auf diese Weise entstandenen Meta-Datenbanken in gleicher Art zum endgültigen prosopographischen Katalog verschmolzen werden.

Ein weiterer Bestandteil der Datenbank wird eine Bibliothek von Hilfsmitteln sein, die sowohl die wichtigsten Algorithmen als auch jene Programmanweisungen enthält, aufgrund derer die Sekundärdatenbanken generiert wurden. Diese Bibliothek dient vor allem der Offenlegung der getroffenen Identifikationsentscheidungen und deren Nachvollziehbarkeit. Auf diese Weise steht es jedem Benutzer offen, einzelne Identifikation zu verändern.

Dieses Konzept einer *modularen Datenbankorganisation*¹⁵ bedingt eine ebenso modulare als auch auf die Bedürfnisse von Historikern zugeschnittene Softwarelösung, die der Fachwelt in Gestalt des Programmsystems CLIO zur Verfügung steht¹⁶. Bei dem von Manfred Thaller am Max-Planck-Institut für Geschichte in Göttingen entwickelten Programmpaket handelt es sich um ein datenbankorientiertes System der quellennahen Datenverarbeitung, geschaffen für die fachspezifischen Bedürfnisse der Historiker¹⁷. In

Rahmen der maschinellen Edition durchgeführt wird. Dies ermöglicht sowohl Selektion und Bearbeitung von Texten aufgrund bestimmter Informationen in den strukturierten Datenbankbereichen als auch die Auswahl von Information zu einzelnen Personen aufgrund thematisch indizierter oder nach Wortmustern aufgesuchter Textteile.

¹⁴ Dieser Schritt steht eben in Bearbeitung – eine Publikation zu den theoretischen und methodischen Problemen und die Dokumentation der Meta-Datenbank ist in Vorbereitung. Zu diesen Bereichen im Design vgl. Ingo H. Kropač: The Prosopographical Data Bank on the History of the South-East Territories of the Old "Reich" up to 1250 (PDB). In: Data Bases in the Humanities and Social Sciences. Proceedings of the Conference on Data Bases in the Humanities and Social Sciences, Montgomery (Alabama) 1987. Osprey 1988 (im Druck).

¹⁵ Der Begriff *modular* bezieht sich hier sowohl auf die Möglichkeit, die Datenbank durch gezielte Operationen zu verändern als auch auf die eventuelle Beschränkung auf einzelne Teile der Datenbank, sei es aus thematischen Überlegungen oder wegen Beschränkungen durch den verwendeten Rechner.

¹⁶ Vgl. Manfred Thaller: CLIO. Einführung und Systemüberblick. Göttingen 1983 (Vorabdruck, die auch im folgenden zitierten Programm-Handbücher sind beim Autor auf Wunsch erhältlich). – Eine allgemein gehaltene jedoch nicht mehr ganz aktuelle Vorstellung des Systems findet sich bei Manfred Thaller: Automation on Parnassus. CLIO - A Databank Oriented System for Historians. In: Historical Social Research 15, 1980, S. 40 - 65.

¹⁷ Die derzeit verwendete Version von CLIO ist in PL/1 geschrieben und nur auf Rechnern der Marke SPERRY UNIVAC 1100 lauffähig. Die portable und somit auf verschiedenen Rechnern (vom PC bis zu Großrechnern) einsetzbare Version *κλειω* steht in Entwicklung, die Version 1.1.1 ist bereits erhältlich. Die Weitergabe des Programmsystems ist für nicht-kommerzielle Zwecke nach Abschluß eines Lizenzvertrags

naher Zukunft sollen die freigegebenen Teile der Datenbank und die zugehörige Software gemeinsam der interessierten Fachwelt zur Verfügung stehen.

2. DATENTECHNISCHE DURCHFÜHRUNG

2.1. Datenerfassung und Datenaufbereitung.

Entsprechend der Konzeption des Projektes wurde die Erfassung aller vorkommenden Personen in den Editionen der angesprochenen Regionen¹⁸ begonnen. Dabei werden aus den Volltexten der Urkunden die Angaben zu den vorkommenden Personen und Institutionen extrahiert. Die aufgenommenen Informationen lassen sich unterscheiden in unmittelbare - direkt in der Quelle vorkommende - und mittelbare - aus der Quelle erschlossene - Angaben. Der Name und die Beschreibung einer Person wären ein Beispiel für den ersten, die Identifikation einer Person ein Beispiel für den zweiten Fall. Nach der Aufarbeitung der Editionen ist die Bearbeitung der unedierten Quellen geplant.

Wie werden nun die Informationen zu den Personen erfaßt und aufbereitet? Eine Beschreibung dieser Vorgangsweise muß verbunden sein mit einer kurzen Vorstellung des im Rahmen dieses Projektes verwendeten Programmsystems CLIO, das sich nicht allein durch seine lateinische Kommandosprache von herkömmlichen Softwarepaketen unterscheidet¹⁹.

Entsprechend dem Design von CLIO werden die Informationen der Quelle logisch in einzelne Einheiten aufgespalten²⁰. Dabei muß versucht werden, das gesamte Datenmaterial in Einheiten ähnlicher logischer Struktur zu zerlegen, welche als *Dokumente* bezeichnet werden. Bei der Arbeit mit urkundlichem Material ist es naheliegend, eine Urkunde als solches anzusehen. Diese *Dokumente* setzen sich ihrerseits aus Teilen zusammen, die ebenfalls logisch abgeschlossen erscheinen, den *Informationsgruppen*. Sie sind in diesem Projekt die in einer Urkunde vorkommenden Personen und Institutionen. In den *Informationsgruppen*, dem *Dokument* logisch untergeordnet, kann in einer Urkunde eine variable Anzahl von Personen und Institutionen vorkommen; diese werden durch ihr Auftreten in einem Dokument dieser bestimmten Urkunde zugeordnet. Dieselbe Form der hierarchischen Abhängigkeit gibt es auch zwischen den *Informationsgruppen* und den *Elementarinformationen*. Sie enthalten etwa den Namen oder die Beschreibung einer Person und sind damit die eigentlichen Träger des Informationsgehaltes. Durch die Zugehörigkeit zur *Informationsgruppe Person* ist eindeutig festgelegt, daß der Name einer Person eben zu dieser Person gehört.

Die Abbildung von Informationen aus der Quelle in ein Datenbankkonzept ist oft mit einer Reduktion des Informationsgehaltes verbunden. Davon sind besonders solche Angaben betroffen, die im weitesten Sinn als Zusatz zu Informationen verstanden werden

mit dem Max-Planck-Institut für Geschichte in Göttingen kostenlos. - Vgl. Manfred Thaller: *κλειω*. Ein fachspezifisches Datenbanksystem für die Historischen Wissenschaften. Version 1.1.1, β -Test Version, Göttingen 1987.

¹⁸ Dasu gehören die Urkundeneditionen der einzelnen (Bundes)Länder oder Urkundenbücher der verschiedenen geistlichen und weltlichen Verbände.

¹⁹ Die vorliegende Beschreibung wie auch die Dokumentation beziehen sich auf den Prototyp CLIO, nicht auf die neue Version *κλειω*. Abgesehen von Einzelheiten des Kapitels 2 ist insbesondere die Dokumentation (Kapitel 4) auch für die Neuimplementation in allen Details gültig.

²⁰ Zur grundsätzlichen Aufbereitung des Informationsgehaltes innerhalb CLIOs vgl. Thaller, Einführung, S. 5 - 16.

können, oder als Kommentar des Historikers beim Studium einer Urkunde. Zur Beseitigung dieses Problems gibt es in CLIO sogenannte *Kommentarfelder* und *Originaltextfelder* (*Unterfelder*). Sie bilden eigentlich zwei äußerst flexible Anmerkungsapparate, die gezielt eingesetzt werden können. Zum Beispiel kann ein Investiturzeug durch ein Kommentarfeld in der Elementarinformation *Funktion* kenntlich gemacht werden.

Die *Informationsgruppen*, *Elementarinformationen* und *Unterfelder* werden durch *Datensignalzeichen* für das System kenntlich gemacht²¹. Diese teilen dem Programm mit, wo die Informationsgruppen, Elementarinformationen und Unterfelder beginnen beziehungsweise enden. Weil einer *Informationsgruppe* nur eine gleichlautende *Elementarinformation* zugewiesen werden kann, sieht CLIO die Möglichkeit vor, verschiedene Angaben als *Mehrfacheinträge* durch einen Strichpunkt getrennt aufzunehmen. Neben dieser hierarchischen Struktur besteht die Möglichkeit, *relationale Netzwerke*²² zu definieren, die Beziehungen zwischen *Informationsgruppen* und/oder *Dokumenten* beschreiben. In der PDB wurden die verwandtschaftlichen Beziehungen zwischen Personen auf diese Art dargestellt. Aus diesen Vorgaben entsteht eine spezifische *Datenstruktur*, die den Informationsgehalt der Quelle widerspiegelt (s.u., Punkt 3.).

Für sämtliche Auswertungen zieht CLIO nicht die Rohdaten heran, sondern arbeitet mit einer sogenannten *Systemdatei*, was zur Folge hat, daß eine solche aus den Rohdaten nach bestimmten Regeln angelegt werden muß. Eine *Systemdatei* ist vergleichbar mit einem wohlgeordneten Archiv²³, in dem alle Informationen gezielt zugänglich sind. Bei der Errichtung derselben werden die in der vorstrukturierten Art erfaßten Daten eingelesen, und dem System mitgeteilt, welche Form der Strukturierung für die Aufbereitung der Informationen verwendet wurde. Es wird auch angegeben, an welcher logischen Stelle die jeweiligen Informationen stehen, daß zum Beispiel jede Information, die als erste bei einer Person erscheint, als deren Identifikation definiert ist. Das verwendete Programmsystem ermöglicht auch die Parallelität von strukturierter Information und Volltexten. Dabei wird unter Volltextaufnahme sowohl die Eingabe einer ganzen Urkunde, als auch die Einbindung eines Regests in die Datenbank verstanden. Strukturierte Informationen sind Zuweisungen des Informationsgehaltes einer Quelle an ein abstraktes Kategorienschema, die *Datenstruktur*.

Hier ist nun auch der Ort für die Diskussion der verschiedenen Plausibilitätsroutinen, die die Fehlerlosigkeit der Daten in der Datenbank gewährleisten. Dabei kommt den unterschiedlichen *Datentypen* CLIO's große Bedeutung zu. Beim Einlesen der Rohdaten wird vom System unterschieden zwischen *Elementarinformationen*, die Texte enthalten (Datentyp *LINGUA*), Kalenderangaben (Datentyp *TEMPORA*), numerischen Angaben (Datentyp *NUMERUS*), Kürzeln (Datentyp *CONDICIO*) und Angaben über relationale Verknüpfungen (Datentyp *RELATIO*)²⁴. Um die Bedeutung dieser Datentypen entspre-

²¹ Im Rahmen der PDB finden die Standard-Datensignalzeichen Verwendung, vgl. dazu Thaller, Einführung, S. 5.

²² Vgl. Manfred Thaller: CLIO. Ein datenbankorientiertes Programmsystem für Historiker. Teil I: Göttingen 1982 (Vorabdruck), S. 89 - 102.

²³ Zur Systemfileerrichtung vgl. Thaller, Einführung, S. 17 - 32.

²⁴ Für die neue Version von CLIO werden derzeit Verhandlungen über die Implementation eines Datentyps SITUS geführt, der die Verwaltung von geographischen Angaben (etwa durch die Umsetzung in

chend darstellen zu können, ist ein kurzer Rückblick zur Datenerfassung notwendig. Kalenderdaten können in der Form eingegeben werden, in der sie in der Quelle erscheinen. Sie werden systemintern in Tage seit dem 1.1.1 umgerechnet, wobei die Einführung des Gregorianischen Kalenders berücksichtigt wird. Die Bewahrung der Schreibung der Quelle ist auch bei der Arbeit mit Maß- und Währungsangaben möglich. Hier wird vom Benutzer eine Umrechnungstabelle definiert, die die eingegebenen Werte in eine absolute Größe verwandelt. Weiters wird die Eingabe wesentlich erleichtert, wenn bestimmte, eindeutige Sachverhalte in abgekürzter Form aufgenommen werden können. Im Rahmen der PDB wird die *Funktion* einer Person in einer Urkunde mit einem solchen *Condicio*-Feld aufgenommen. Für die Plausibilitätstests sind die unterschiedlichen Datentypen insofern von Vorteil, als sie bestimmte Eingabeformen verlangen. Bei einer geschickten Schachtelung unterschiedlicher Datentypen werden Eingabefehler, wie etwa das unbeabsichtigte Fehlen einer Information sofort angezeigt. Außerdem kann die Zuverlässigkeit der Eingabe mit Hilfe des Protokolls überprüft werden, das bei der Errichtung der Systemdatei - dem sogenannten *Prüf*lauf - angelegt wird. Es enthält nämlich eine Häufigkeitsauszählung der einzelnen Informationen, die ebenfalls auf Eingabefehler aufmerksam machen kann. Nach der allfälligen Korrektur der Rohdaten wird ein neuer *Prüf*lauf gemacht, so lange, bis keine formalen Fehler mehr diagnostiziert werden. Eine intensive inhaltliche Überprüfung ist mit den unten besprochenen Formen des *Data Retrievals* möglich.

2.2. Auswertung und Transformation der Daten.

CLIO erlaubt vier grundsätzlich verschiedene Bereiche der Evaluation und Transformation des Datenmaterials, die hier allerdings lediglich angedeutet werden können.

1. *Data Retrieval*:

Darunter ist prinzipiell jede Form der Zugänglichmachung der in der Datenbank enthaltenen Informationen zu verstehen²⁵. Es reicht von der gezielten Suche nach bestimmten logischen Einheiten - etwa alle Personen, die in einem vorgegebenen Zeitraum in Schenkungsurkunden an ein bestimmtes Kloster als Zeugen aufgetreten sind - bis hin zur Erstellung komplexer Indices - Verzeichnis aller vorkommender Personen mit ihrem Namen, ihrer Beschreibung, ihrer Funktion, etc. - und Erstellung von Häufigkeitszählungen. Diese dem Standard aller Datenbanksysteme entsprechenden Möglichkeiten werden für den Historiker interessanter, wenn er weiß, daß bei jeder *Data Retrieval*-Routine Normalisierungen, Kategorisierungen und Unschärfeparameter verwendet werden können. Das heißt, daß bei der Suche nach einer Person nicht nach deren Namen in der eingegebenen Form gesucht wird, sondern daß dieser Name nach bestimmten Regeln für die Suche in eine Normalform gebracht wird. Diese Reduktion auf eine lautliche Grundform wird in CLIO durch *Algorithmen* durchgeführt. Sie bieten dem Benutzer einen Rahmen an, innerhalb dessen spezifische Vereinbarungen für einen Quellenbestand getroffen werden können. Im Rahmen dieses Projektes wurden bislang vor allem *Soundex-Algorithmen*²⁶ verwendet. Diese setzen den Namen

Landkarten) ermöglichen wird.

²⁵ Vgl. Thaller, Einführung, S. 36 - 61, eine ausführliche Beschreibung dieser Komponenten findet sich bei Manfred Thaller: CLIO. Ein datenbankorientiertes Programmsystem für Historiker. Teil II: Göttingen 1983 (Vorabdruck).

²⁶ Vgl. Thaller, Einführung, S. 90 - 94 und Thaller, CLIO I, S. 201 - 210. - Zur Verwendung eines

um in ein vierstelliges Codesystem, wobei jeder Code einer Lautgruppe entspricht, die der Bearbeiter vorzugeben hat. Entsprechend der Grundkonzeption von CLIO sind diese Direktiven jederzeit änderbar und stellen somit ein wichtiges und flexibles Instrument zur Personenidentifikation dar. Die Anlage eines Kategorienschemas wird von sogenannten *Thesauren* durchgeführt. Dabei werden Begriffe entweder kodiert, das heißt, daß *episcopus* den Code 1000 erhalten könnte, oder in ein hierarchisches System überführt, was bedeutet, daß *episcopus* in diesem Fall eine Untergruppe zu *Geistliche Würdenträger* wäre. Beide Formen der Kategorisierung können bei der Suche, Sortierung oder Zählung verwendet werden. Damit werden nicht nur Schreibungsvarianten ausgeschaltet, sondern auch die gezielte Suche nach bestimmten Personengruppen erleichtert. Während des *Data Retrieval* ist es oft unmöglich bestimmte Suchkriterien so genau zu formulieren, daß ohne die Verwendung von *Unschärfeoperatoren* ein befriedigendes Ergebnis erzielt werden könnte. So ist es viel effizienter nach Personen zu suchen, die *ungefähr* um 1100 vorgekommen sind, als sämtliche möglichen Datumsangaben in einem Suchprofil anzugeben.

2. *Record Linkage*:

Dieses Modul dient zur logischen Verknüpfung zweier oder mehrerer gleich oder verschieden strukturierter Datenbanken aufgrund von in ihnen enthaltenen gemeinsamen Informationen²⁷. In den herkömmlichen Anwendungsbereichen dieses Verfahrens²⁸ wird die Suche nach jenen Personen, die als Identifikatoren für die Verknüpfung dienen, systematisiert. Das bedeutet, daß zwei Datenbanken, die verschiedene Quellenbestände repräsentieren, nach identischen Personen durchsucht und die Informationen zu denselben in einer neuen Datenbank akkumuliert werden. Für die PDB ist dieses Modul in zweifacher Hinsicht notwendig. Einerseits, um die Personennennungen in den *Primärdatenbanken* zu Informationen über die Personen zu überführen, und andererseits mehrere solcher *Sekundärdatenbanken* zu einem prosopographischen

ähnlichen Ansatzes in anderen Projekten vgl. Theodore Hershberg: *The Philadelphia Social History Project: A Methodological History*. Stanford 1973, bes. S. 196 - 197. - Ian Winchester, *The Linkage of Historical Records by Man and Computer: Techniques and Problems*. In: *Journal of Interdisciplinary History* 1, 1970, S. 115 - 119. - M. Boot, P. Lourens, J. Lucassen: *A Linguistic Preprocessor for Record Linkage in Socio-Economic Historical Research*. In: *Computers and the Humanities* 17, 1983, bes. S. 46 - 48. - J. Dennis Willigan und Katherine A. Lynch: *Sources and Methods of Historical Demography*. New York 1982 (*Studies in social discontinuity*), S. 178 - 181. - Theodore Hershberg, Alan Burstein, Robert Dockhorn: *Verkettung von Daten. Record Linkage am Beispiel des Philadelphia Social History Project*. In: *Moderne Stadtgeschichte*, hrsg. v. Wilhelm Heinz Schröder. Stuttgart 1979 (*Historisch-sozialwissenschaftliche Forschungen* 8), S. 43 - 46.

²⁷ Vgl. Thaller, *Einführung*, S. 82 - 120.

²⁸ Ein mittlerweile klassisch zu nennender Anwendungsbereich dieser Vorgangsweise ist die Familienrekonstitution, bei der zu jedem Heiratseintrag die Informationen über die reproduktive Tätigkeit des Paares und die entsprechenden Sterbedaten hinzugefügt werden. Zur Vorgangsweise vgl. Arthur E. Imhof: *Einführung in die Historische Demographie*. München 1977 (*Beck'sche Elementarbücher*), S. 101 - 112. - Michel Fleury und Louis Henry: *Nouveau Manuel de Dépouillement ed d'Exploitation de l'État civil ancien*. Paris 1985, S. 115 - 172. - Willigan, Lynch, *Sources and Methods*, S. 161 - 182. - Ian Winchester: *The Linkage of Historical Records by Man and Computer: Techniques and Problems*. In: *Journal of Interdisciplinary History* 1, 1970, S. 107 - 124.

Gesamtkatalog zu verbinden. Die *Algorithmen* zum Ausgleich der Schreibungsvarianz sind hier natürlich von entscheidender Bedeutung.

3. Textanalyse:

Damit können aus Volltexten die verschiedenen Wortformen- und Wortklassenlisten, Fundstellenverzeichnisse von Worten und Auszählungen von Worthäufigkeiten erstellt werden.²⁹ Mit einem gesonderten Modul zur interaktiven Textbearbeitung (IAC)³⁰ können zudem gemeinsam vorkommende Termini gesucht werden, wie zum Beispiel eine Person mit einer bestimmten Beschreibung³¹. Im Rahmen des Forschungsprojektes³² ist dieser Auswertungsbereich für die Bearbeitung der Editionsprojekte notwendig.

4. Maschinelle Kodierung³³:

CLIO stellt Routinen zur Verfügung, mit Hilfe derer Teile einer Datenbank, die ja zum Teil den originalen Wortlaut der Quelle enthalten, in numerische Dateien überführt werden können. Diese ermöglichen die Bearbeitung durch Statistikprogramme (s.o.) im Sinne quantifizierender Verfahren. Die Kodierung der Begriffe erfolgt maschinell, entsprechend den Vorgaben, die der Benutzer in Kategorienschemata - *Thesauri* - festlegt.

3. DATENSTRUKTUR DER PRIMÄRDATENBANKEN

Das im folgenden angeführte Datenmodell besteht aus 16 Feldern, die insgesamt 4 Informationsgruppen zugeordnet sind, wobei die Informationsgruppen lediglich in 2 logischen Ebenen geschachtelt erscheinen:

```
U(rkunde):  Zitat/Datierung/Ort/Überlieferung
|-- P(erson):  Id(entifikation)/(Vor)Name/Beschreibung/Funktion/F(unktions)Zahl
|              /Status/Beziehung
|-- I(nstitution):  Funktion/Typ/Name/Ort
|-- D(atierung):  Text
```

Nach den Möglichkeiten des Programmsystems CLIO kann jede Informationsgruppe pro Dokument nahezu beliebig oft vorkommen, jedes Feld pro Informationsgruppe nur ein einziges Mal. Das bedeutet, daß pro Urkunde über tausend Personen vollständig beschrieben werden können. Die derzeit verwendete Programmversion gestattet die gemeinsame Verarbeitung von 200.000 Dokumenten pro Primärdatenbank (eine Systemdatei).

4. DOKUMENTATION DER PRIMÄRDATENBANKEN

Zu Beginn seien einige Punkte erwähnt, die den semantischen Bezug aller Felder betreffen und später nicht mehr aufgeführt werden:

²⁹ Vgl. Thaller, Einführung, S. 121 - 132 .

³⁰ Vgl. Elke Gildenstein: Interactive Concordance System. Dokumentation. Göttingen 1985.

³¹ In der zukünftigen Version von CLIO werden diese interaktiven Möglichkeiten auch in Verbindung mit dem Retrieval (s.o.) für strukturierte Daten zur Verfügung stehen.

³² Siehe Anm. 4.

³³ Vgl. Thaller, Einführung, S. 62 - 81 .

- Erschließungen von Informationen werden immer nur im Rahmen eines Dokuments durchgeführt. Bezüge, die über eine Urkunde hinausgehen, können erst durch die Erstellung der Meta-Datenbank hergestellt werden.
- Personen aus Vorurkunden und inserierten Urkunden (im Druckbild einer Edition durch den Petitsatz leicht zu erkennen) werden als Genannte (Kategorie *G* des Feldes *Funktion*) klassifiziert.
- Aus Depertita werden nur Personen aufgenommen, die in der "verlorenen" (etwa inserierten) Urkunde erwähnt sind: der Aussteller der Bestätigungsurkunde o.ä. werden nicht aufgenommen.
- Alle erschlossenen Informationen werden mit einem ! am Feldende gekennzeichnet. Dies hat den Vorteil, daß jeder Bearbeiter sofort erkennen kann, welche Informationen unmittelbar der Quelle entstammen und welche nicht.
- Alle fraglichen oder in irgendeinem Zusammenhang unsicheren Informationen sind ausnahmslos am Feldende mit einem ? ausgezeichnet.

Ein weiteres Problem ergibt sich bei der Darstellung von Sonderzeichen und Diakritika. Hier eine Auflistung der derzeit gültigen Regelungen³⁴:

- Akzentkonvention durch die Verbindung des Indikators "\", des entsprechenden Akzentzeichens und dem zu akzentuierenden Buchstabens ($\acute{a} \Rightarrow \backslash 'a$, $\ddot{u} \Rightarrow \backslash "u$)
- Konvention für die in der Standard-ASCII-Tabelle nicht definierten Buchstaben und Sonderzeichen gemäß der $\text{\TeX}$ -Eingabekonvention, z.B. $\beta \Rightarrow \backslash \{ss\}$
- Cauda-Konvention durch die Verbindung des Indikators "\", der Kennung für die Cauda (c) und den betroffenen Buchstaben ($\xi \Rightarrow \backslash c\{e\}$, $\zeta \Rightarrow \backslash c c$), wobei zu beachten ist, daß entweder zwischen der Kennung und dem Buchstaben ein Leerzeichen zu stehen hat oder der Buchstabe zwischen einem Paar geschwungener Klammern erscheinen muß.
- Konvention für hochgestellten Buchstabenbestand durch die Verbindung des Indikators "\", der Kennung "~" und des hochzustellenden Buchstabenbestandes in einem Paar geschwungener Klammern ($a^e \Rightarrow a\sim\{e\}$)
- Konvention für tiefgestellten Buchstabenbestand durch die Verbindung des Indikators "\" mit der Kennung "-" und des tiefzustellenden Buchstabenbestandes in einem Paar geschwungener Klammern ($A.1 \Rightarrow A\backslash\{-1\}$)
- Konvention für Ligaturen durch die Verbindung des Indikators "\" und den Einzelbuchstaben der Ligatur ($\ae \Rightarrow \backslash ae$)

4.1. Zitat.

- Datentechnische Beschreibung:
CLIO-Datentyp *LINGUA*, Abhängigkeit hierarchisch, Informationsgruppe *U(rkunde)*, Ebene 1, Position 1, Feldlänge max. 12 Stellen.

³⁴ Die angeführte Beschreibung entspricht weitgehend den Eingabekonventionen des Programmsystems $\text{\TeX}$, das im Rahmen des Projekts HIFADOK als Postprozessor eingesetzt wird. Auch die Eingabekonventionen der portablen Version CLIO/C wird sich an diesen Standards orientieren. Neben den hier angeführten Regeln sind eine große Zahl nationaler Alphabete auf analoge Weise darstellbar. Dazu vgl. Donald E. Knuth: *The $\text{\TeX}$ book*. 6. Aufl. Reading [etc.] 1986, bes. S. 37 - 41, 51 - 54.

- Semantische Beschreibung:

Zitat beschreibt

- entweder den Fundort der Urkunde in einer Edition mit dem Sigel des Urkundenbuches mit Bandangabe (4 Stellen), einem "*", der Urkundennummer (3 Stellen), einem weiteren "*" und der Seitenangabe (3 Stellen)
- oder den Lagerort mit dem Sigel des Archivs, einem "*" und der Archivsignatur.

4.2. Datierung.

- Datentechnische Beschreibung:

CLIO-Datentyp *TEMPORA*, Abhängigkeit hierarchisch, Informationsgruppe *D(okument)*, Ebene 1, Position 2, Feldlänge variabel, Mehrfacheintragungen möglich.

- Semantische Beschreibung:

Datierung beinhaltet das Ausstellungsdatum der Urkunde in diplomatischer Notation (Jahr.Monat.Tag), wobei bei fehlenden Teilangaben "0" zu schreiben ist und unklare Angaben im Terminus postquem/Terminus antequem - Verfahren einzugeben sind. Sollten mehrere Daten zu einer Urkunde aufscheinen, so sind sie in der Reihenfolge ihres Auftretens als Mehrfacheintragungen erfasst.

4.3. Ort.

- Datentechnische Beschreibung:

CLIO-Datentyp *LINGUA*, Abhängigkeit hierarchisch, Informationsgruppe *U(rkunde)*, Ebene 1, Position 3, Feldlänge variabel, Mehrfacheintragungen möglich.

- Semantische Beschreibung:

Ort gibt den aufgelösten Ausstellungsort der Urkunde wieder, wobei nur der Ortsname angeführt wird, nicht aber nähere Bezeichnungen (z.B. in der Annenkirche). Bei mehreren Nennungen von Ausstellungsorten mit Bezug auf eine Urkunde - ev. bei mehreren verschiedenen Fassungen - sind die Orte in der Reihenfolge ihres Auftretens erfasst.

4.4. Überlieferung.

- Datentechnische Beschreibung:

CLIO-Datentyp *CONDICIO*, Abhängigkeit hierarchisch, Informationsgruppe *U(rkunde)*, Ebene 1, Position 4, Feldlänge variabel, Mehrfacheintragungen im Rahmen der legalen Zeichen möglich. Legale Zeichen: F, O, K, U, D.

- Semantische Beschreibung:

In Überlieferung können unter Subsumierung der in den Editionen verwendeten Begriffe nur folgende eindeutige Kategorien abgebildet werden:

F = Fälschung (darf nur zusammen mit anderen legalen Zeichen vorkommen!)

O = Original

K = Kopialüberlieferung: Abschrift, Druck, Insert

U = Unvollständige Urkunde: Fragment, Auszug, Konzept

D = Depeditum

4.5. Id(entifikation).

- Datentechnische Beschreibung:

CLIO-Datentyp *LINGUA* und *RELATIO PASSIVA*, Abhängigkeit hierarchisch und relational, Informationsgruppe *P(erson)*, Ebene 2, Position 1, Feldlänge 12, keine Mehrfacheintragen, keine Unterfelder.

Bei Systemdateien, die lediglich für das Nominative Record Linkage angelegt wurden, kommt es zu einer Abweichung: Da im genannten Verfahren relationale Netzwerke auf direktem Weg nicht übertragen werden können, ist als CLIO-Datentyp ausschließlich *LINGUA* definiert. Alle anderen Parameter bleiben gleich.

- Semantische Beschreibung:

Id(entifikation) weist allen auftretenden Personen im einzelnen Dokument und in allen für die Gesamtdatenbank angelegten Datensätzen eine eindeutige Identifikation zu. Diese Kennung besteht aus dem Code für das Urkundenbuch inkl. Bandangabe (4 Stellen), dem Zeichen "*", der Urkundennummer (3 Stellen), dem Zeichen "*" und der durchlaufenden Numeration von Personen in der einzelnen Urkunde. Sie dient gleichzeitig als passive Beziehung des relationalen Netzwerks *VERWANDT*, außer bei jenen Systemdateien, die nur dem Nominative Record Linkage dienen.

4.6. (Vor)Name.

- Datentechnische Beschreibung:

CLIO-Datentyp *LINGUA*, Abhängigkeit hierarchisch, Informationsgruppe *P(erson)*, Ebene 2, Position 2, Feldlänge variabel, keine Mehrfacheintragen, keine Unterfelder.

- Semantische Beschreibung:

(Vor)name enthält den eigentlichen Namen aller in den Urkunden genannten Personen. Er wird in Originalschreibweise der Quelle erfaßt, die diversen Kasus werden jedoch in den Nominativ gesetzt. Alle im Namensmaterial vorkommenden Sonderzeichen werden gemäß der unter Punkt 3 vorkommenden Vorschriften behandelt.

4.7. Beschreibung.

- Datentechnische Beschreibung:

CLIO-Datentyp *LINGUA*, Abhängigkeit hierarchisch, Informationsgruppe *P(erson)*, Ebene 2, Position 3, Feldlänge variabel, Mehrfacheintragen möglich, Kommentarfeld möglich.

- Semantische Beschreibung:

In dieses Feld finden alle Beschreibungen Eingang, die zu einer Person in unmittelbarem semantischen Bezug stehen, wie z.B. Zuname, Titel, Stand, Amt, Beruf etc., oder auch Beschreibungen, die mehrere solche Bezüge zugleich ausdrücken. Eine Ausnahme bilden jedoch die verwandtschaftlichen Beziehungen, diese werden im Feld *Beziehung* dargestellt, um als relationales Netzwerk verwaltet zu werden. Jene Informationen, die nur in mittelbarem semantischen Bezug zu einer Person stehen, wie etwa die Information, daß eine Person X zu einem Grundstück Y im Verhältnis Z steht, bleibt für die PDB unberücksichtigt.

Um den Erfordernissen des Record Linkage zu entsprechen, wird in diesem Feld die bestmögliche Erschließung von Information angestrebt, wobei jedoch zu beachten ist, daß jede auf diese Weise in die Datenbank eingehende Beschreibung mit einem ! gekennzeichnet ist (s.u.). Bei der Erfassung dieses Feldes gilt für die Erschließung von Namen folgendes Regelwerk:

x et filius eius y de s wird aufgelöst als: x de s; y de s
Adelochus et filius eius Gundacher de Grez $\Rightarrow$ *Adelochus/de Grez ... Gundacher/de Grez ...*
 x de s et filius eius y wird aufgelöst als: x de s; y de s!
Adelochus de Grez et filius eius Gundacher $\Rightarrow$ *Adelochus/de Grez ... Gundacher/de Grez! ...*
 x (et) y fratres de s wird aufgelöst als: x de s; y de s
Adelochus (et) Gundacher fratres de Grez $\Rightarrow$ *Adelochus/de Grez ... Gundacher/de Grez ...*
 x de s et frater eius y wird aufgelöst als: x de s; y de s!
Adelochus de Grez et frater eius Gundacher $\Rightarrow$ *Adelochus/de Grez ... Gundacher de Grez! ...*
 x de s et y wird aufgelöst als: x de s; y
Adelochus de Grez et Gundacher $\Rightarrow$ *Adelochus/de Grez ... Gundacher ...*
 x et frater eius y de s wird aufgelöst als: x de s!; y de s
Adelochus et frater eius Gundacher de Grez $\Rightarrow$ *Adelochus/de Grez! ... Gundacher/de Grez ...*
 x et y de s wird aufgelöst als: x de s!; y de s
Adelochus et Gundacher de Grez $\Rightarrow$ *Adelochus/de Grez! ... Gundacher/de Grez ...*
 w x y de s wird aufgelöst als: w de s; x de s; y de s
Chunradus Adelochus Gundacher de Grez $\Rightarrow$ *Chunradus/de Grez ... Adelochus/de Grez ...*
Gundacher/de Grez

Dieses Regelwerk dokumentiert bis dato aufgetretene Fälle und ist in der Folge beliebig erweiterbar.

4.8. Funktion.

- Datentechnische Beschreibung:

CLIO-Datentyp *CONDICIO*, Abhängigkeit hierarchisch, Informationsgruppe *P(erson)*, Ebene 2, Position 4, Feldlänge variabel (max. 3), Mehrfacheintragungen im Rahmen der legalen Zeichen möglich, Kommentarfeld zur Bildung der Kategorien unbedingt erforderlich. Legale Zeichen: A, E, F, X, G, Z, S, U, V.

- Semantische Beschreibung:

Hier wird der semantische Bezug einer Person im diplomatische Sinne festgelegt, d.h. der Person wird jene Funktion zugewiesen, die sie im Hinblick auf das Dokument eo ipso einnimmt. Folgende eindeutige Kategorien werden abgebildet:

A = Aussteller (Unterschrift wird nicht extra vermerkt)

E = Empfänger

X = Adressat

G = "Genannter" (auch "Beteiligte"!)

Z = Zeuge (siehe unten)

S = Siegler

F = Falscher ... (bei hineingefälschten Personen)

U = Unterfertiger (dazu gehören alle jene, die Kanzleifunktionen ausüben, wie Notare, Schreiber, Kanzler etc.)

V = Verursacher der Rechtshandlung, wenn sie nicht mit dem Aussteller ident sind (bei Traditionsnotizen etc.).

Ein besonderes Problem ergibt sich bei der Kategorie Z, den Zeugen. Darin wird jede Art von Zeugen subsumiert, auch Investiturzeugen. Weiters ist es notwendig, die verschiedenen Zeugenreihen, die in einer Urkunde vorkommen können, auseinanderzuhalten, wofür insbesondere das Kommentarfeld Verwendung findet.

- Z steht bei Zeugenreihen, welche die Ausstellung der Urkunde bezeugen.
- Z#x steht bei x-Zeugenreihen, die Rechtshandlungen bezeugen, welche in der Urkunde geschildert werden, wobei x für die Position der Zeugenreihe innerhalb der Urkunde steht.
- Z#Ax, bzw. Z#Bz steht bei Zeugenreihen, die in verschiedenen Originalen oder Abschriften vorkommen, wobei A, B etc. den Überlieferungsstrom bezeichnet, in dem die jeweilige Zeugenreihe vorkommt. Diese Form wird nur dann gewählt, wenn die Zeugenreihen zwischen den einzelnen Überlieferungsarten differieren.

Es werden nur Zeugen aufgenommen, die eindeutig als Individuen erkannt werden können, d.h., daß Angaben wie "et alii" etc. hier ignoriert werden. Wenn in der Urkunde steht "N.N. et filius suus" gilt der filius als eindeutig identifizierbare, wenn auch hier namenlose Person und wird aufgenommen; steht jedoch "N.N. et filii" sind die Söhne in der Zahl nicht verifizierbar und werden nicht aufgenommen, ihre Existenz ist jedoch im Kommentarfeld des Feldes *Beziehung* unter N.N. vermerkt.

Personen, die in inserierten Urkunden vorkommen, werden stets als *G(enannte)* der vorliegenden Urkunde aufgenommen; solche, die eine Urkunde durch ihre Unterschrift bestätigen (... *subscripti*), gelten als Unterfertigte.

4.9. F(unktions)Zahl.

- Datentechnische Beschreibung:

CLIO-Datentyp *NUMERUS*, Abhängigkeit hierarchisch, Informationsgruppe *P(erson)*, Ebene 2, Position 5, Feldlänge variabel, Mehrfacheintragen und Unterfelder möglich.

- Semantische Beschreibung:

F(unktions)Zahl gibt die Position der einzelnen Person in der Liste ihrer Funktion wieder. Bei mehreren und unterschiedlichen Überlieferungsströmen ist es nötig, die Positionsangaben in Mehrfachnennungen zu verspeichern. Wenn dies der Fall ist, wird in einem Kommentarfeld der jeweilige Überlieferungsstrom angegeben.

4.10. Status.

- Datentechnische Beschreibung:

CLIO-Datentyp *CONDICIO*, Abhängigkeit hierarchisch, Informationsgruppe *P(erson)*, Ebene 2, Position 6, Feldlänge 1, Mehrfacheintragen nicht erlaubt, keine Unterfelder.

- Semantische Beschreibung:

Der Status einer Person erlaubt nur drei eindeutige Kategorien:

- L = lebendig
- T = tot

U = unklar (ev. zu erschließen).

Dabei muß davon ausgegangen werden, daß Aussteller, Zeugen und Unterfertigte einer Urkunde immer der Kategorie L zuzuordnen sind. Auch bei Fälschungen gilt dieser Grundsatz, da jene ohnehin durch das Feld *Überlieferung* gekennzeichnet sind. Die Kategorie U dient als Zwischenstadium und sollte im Verlauf des Nominative Record Linkage den jeweils anderen zuzuordnen sein.

Wenn Personen in einer Urkunde auftreten, die entweder einem datierten Insert oder einer vom Ausstellungsdatum differierenden, datierten Rechtshandlung zuzuweisen sind, so wird ein zusätzliches Unterfeld eingeführt, welches die für das Auftreten der Person jeweils relevante Datierung enthält.

4.11. Beziehung.

- Datentechnische Beschreibung:

CLIO-Datentyp *LINGUA* und *RELATIO ACTIVA*, Abhängigkeit hierarchisch und relational, Informationsgruppe *P(erson)*, Ebene 2, Position 7, Feldlänge variabel, Mehrfacheintragen möglich, keine Unterfelder.

Alles über die Transformation von Netzwerken beim Feld *Id(entifikation)* Gesagtes gilt analog auch hier.

- Semantische Beschreibung:

Beziehung stellt die aktive Komponente des relationalen Netzwerks *VERWANDT* dar. Die Verwandtschaftsbeziehungen werden wie folgt gekennzeichnet:

- Verwandtschaftsbezeichnungen aus der Quelle sind im Originaltext zur übernehmen.
- Erschlossene Verwandtschaftsbezeichnungen werden nur dann im Wortlaut der Quellenstelle wiedergegeben, wenn sie eindeutig sind:
x et frater eius y wird aufgelöst als (fiktives Beispiel):
P\$BUB2*429*001/x/////BUB2*429*002 frater
P\$BUB2*429*002/y/////BUB2*429*001 frater
- Erschlossene Verwandtschaftsbeziehungen mit nicht so eindeutigem Bedeutungsgehalt werden auf der Originalschiene im Originalwortlaut aufgenommen, die erschlossene Schiene dagegen mit der vermuteten deutschen Übersetzung. Ist eine zumindestens ungefähre Übersetzung nicht möglich, wird die erschlossene Schiene mit dem Terminus *verwandt* bezeichnet. Wenn ein und dieselbe Person innerhalb einer Urkunde in verschiedenen Funktionen auftritt und demnach auch mehrfach erfaßt wird, wird sie in diesem Feld jeweils mit dem Terminus *ident* bezeichnet.
- Bei allen nur irgendwie erschlossenen Beziehungen wird hinter den Terminus das Ausrufezeichen gesetzt.

4.12. Funktion.

- Datentechnische Beschreibung:

CLIO-Datentyp *CONDICIO*, Abhängigkeit hierarchisch, Informationsgruppe *I(nstitution)*, Ebene 2, Position 1, Feldlänge variabel (max. 3), Unterfelder möglich.

- Semantische Beschreibung:
Dieses Feld entspricht dem gleichen Namens in der Informationsgruppe *P(erson)*, siehe 4.8.

4.13. Typ.

- Datentechnische Beschreibung:
CLIO-Datentyp *LINGUA*, Abhängigkeit hierarchisch, Informationsgruppe *I(nstitution)*, Ebene 2, Position 2, Feldlänge variabel, Unterfelder möglich.
- Semantische Beschreibung:
Hier ist der Typus einer in der Urkunde als Handlungsträger auftretenden oder auch nur genannten Institution bezeichnet – etwa Kloster, Erzstift, Bistum oder auch Altar etc., wobei die Originalschreibweise der Quelle (*fratres*, *capitulum* etc.) beibehalten wird.

4.14. Name.

- Datentechnische Beschreibung:
CLIO-Datentyp *LINGUA*, Abhängigkeit hierarchisch, Informationsgruppe *I(nstitution)*, Ebene 2, Position 3, Feldlänge variabel, Unterfelder möglich.
- Semantische Beschreibung:
Dieses Feld enthält den aufgelösten Namen der jeweiligen Institution – etwa (Kloster) St. Lambrecht, (Bistum) Seckau, aber auch (Altar) St. Anna etc.

4.15. Ort.

- Datentechnische Beschreibung:
CLIO-Datentyp *LINGUA*, Abhängigkeit hierarchisch, Informationsgruppe *I(nstitution)*, Ebene 2, Position 4, Feldlänge variabel, Unterfelder möglich.
- Semantische Beschreibung:
Bis zur Entwicklung eines geeigneten Datentyps beinhaltet *Ort* nur den aufgelösten Namen jenes Orts in amtlicher Normalschreibweise, an dem die jeweilige Institution ansässig war.

4.16. Text.

- Datentechnische Beschreibung:
CLIO-Datentyp *LINGUA*, Abhängigkeit hierarchisch, Informationsgruppe *D(atierung)*, Ebene 2, Position 1, keine Mehrfacheintragen oder Unterfelder möglich.
- Semantische Beschreibung:
Dieses Textfeld enthält den genauen Wortlaut der Datierungszeile einer Urkunde zur Kontrolle der Datumsauflösung.

Kontaktadresse:

Dr. Ingo H. Kropač

Forschungsinstitut für Historische Grundwissenschaften der Karl-Franzens-Universität Graz

A-8010 Graz, Körblergasse 20, Tel.: (0316) 32-4-49

ANHANG:
BEISPIEL FÜR EIN DOKUMENT EINER PRIMÄRDATENBANK

1. Text der Urkunde*:

293.

1175 October 27.

Bischof Roman II. schenkt der Custodie des Kapitels zwei Huben in Dolz (s. Weitensfeld).

...

In nomine sancte et individue trinitatis. Ego Romanus dei gratia sancte Gurcensis ecclesie episcopus. Universitati tam presentium quam futurorum Christi fidelium noticie transmittimus, quemadmodum nos ob reverentiam beate dei genitricis Marie et memoriam domine Hemme fundatricis ecclesie nostre pie memorie et nostra antecessorumque ac successorum perpetua salute ad petitionem etiam dilecti nostri Dietrici prepositi, Weslini decani, Hainrici custodie, ceterorum fratrum donavimus ad eorum videlicet custodiam duos mansus in Dolz cum terminis suis quos a nobis in beneficio Dietmarus antecessorum nostrorum officialis habuerat. Huius rei testes sunt: Purchardus, Engelscalcus, Engelbertus, Eglolfus, frater eius Sifridus, W[al]therus, Rainhalmus, Dietricus, Walbrunus, Hainricus officialis, Wolfgangus et alii quam plures.

Acta sunt hec VI kal. novemb., anno ab incarnatione domini M^oC^oLXXV, indictione VIII; feliciter amen.

S.

2. Dokument des Rohdatensatzes:

U\$MDC1*293*223/1175.10.27//O

P\$MDC1*293*001/Romanus/dei gratia sancte Gurcensis ecclesi{c{e}} episcopus/AS/1/L

P\$MDC1*293*002/Hemma/fundatrix ecclesie nostre/G/1/T

P\$MDC1*293*003/Dietricus/prepositus#von Gurk/G/2/L

P\$MDC1*293*004/Weslinus/decanus#von Gurk/G/3/L

P\$MDC1*293*005/Hainricus/custos#von Gurk/G/4/L

IQE/custodia/Gurk/Gurk

P\$MDC1*293*006/Dietmarus/antecessorum nostrorum officialis/G/5/U

P\$MDC1*293*007/Purchardus//Z/1/L

P\$MDC1*293*008/Engelscalcus//Z/2/L

P\$MDC1*293*009/Engelbertus//Z/3/L

P\$MDC1*293*010/Eglolfus//Z/4/L/MDC1*293*011 frater

P\$MDC1*293*011/Sifridus//Z/5/L/MDC1*293*010 frater

P\$MDC1*293*012/Waltherus//Z/6/L

P\$MDC1*293*013/Rainhalmus//Z/7/L

P\$MDC1*293*014/Dietricus//Z/8/L

P\$MDC1*293*015/Walbrunus//Z/9/L

P\$MDC1*293*016/Hainricus/officialis/Z/10/L

P\$MDC1*293*017/Wolfgangus//Z/11/L

D\$Acta ... VI kal. novemb., anno ... MCLXIV, indictione VIII;

* Die Gurker Geschichtsquellen 864 - 1232, hrsg. v. August v. Jaksch. Klagenfurt 1896 (Monumenta Historica Ducatus Carinthiae 1), S. 223.

MEDIUM AEVUM QUOTIDIANUM

newsletter 10

Krems 1987

Herausgeber: Medium Aevum Quotidianum. Gesellschaft zur Erforschung der materiellen Kultur des Mittelalters. Körnermarkt 13, A-3500 Krems, Österreich. – Für den Inhalt verantwortlich: Univ.Prof. Dr. Harry Kühnel. – Druck: HTU-Wirtschaftsbetrieb Ges.m.b.H., Karls gasse 16, 1040 Wien.

INHALTSVERZEICHNIS / CONTENTS

Vorwort	4
Manfred Thaller, The Daily Life of the Middle Ages, Editions of Sources and Data Processing	6
Ingo H. Kropač und Peter Becker, Die Prosopographische Datenbank zur Geschichte der südöstlichen Reichsgebiete bis 1250 (PDB): Konzepte und Kurzdokumentation ...	30
Brigitte Rath, Auf der Suche nach Alltagen – der 'Frauenalltag' im Mittelalter oder "Leicht hatten es die Frauen nicht"	49
Rezensionen und Berichte	57

Vorwort

Das vorliegende Heft 10 von *Medium Aevum Quotidianum*-Newsletter widmet sich einem Aspekt, der in der internationalen historischen Diskussion seit geraumer Zeit einen entscheidenden Schwerpunkt darstellt: der Anwendung computerunterstützter Methoden.

Es braucht nicht gesondert betont zu werden, daß in der Alltagsgeschichte, und nicht allein in jener des Mittelalters, diesbezüglich ein bedeutender Nachholbedarf zu konstatieren ist. Doch auch in unserem Fachbereich mehren sich Einsicht und Zeichen, daß EDV mehr bieten kann als Textverarbeitungsprogramme zur schnellen und billigen Produktion von Publikationen. Ein entscheidender Vorstoß in diese Richtung ergibt sich beileibe nicht nur für den deutschsprachigen Raum auf Grund der Initiative des Max-Planck-Instituts für Geschichte in Göttingen und dessen EDV-Spezialisten Manfred Thaller. Wir sind deshalb besonders froh, gerade ihn für einen sehr grundsätzlichen Beitrag in *Medium Aevum Quotidianum*-Newsletter gewonnen zu haben. Auf Grund der internationalen Relevanz des Aufsatzes haben wir uns entschlossen, diesen in englischer Sprache wiederzugeben.

Der zweite umfassendere Beitrag des Heftes, verfaßt von Ingo Kropač und Peter Becker (beide Graz), widmet sich einem Anwendungsbeispiel, dessen inhaltlicher Schwerpunkt zwar auf dem Gebiet der mittelalterlichen Prosopographie zu suchen ist, welches jedoch nicht nur wegen der angewandten Methoden auch für die Alltagsgeschichte des Mittelalters von Wichtigkeit erscheint.

Beide Beiträge sollen den Auftakt darstellen für eine in unserer Reihe fortzusetzende Diskussion zur Anwendung computerunterstützter Methoden.

Der bereits seit längerem angekündigte Band mit den Ergebnissen der von *Medium Aevum Quotidianum* im Oktober 1985 mitveranstalteten Tagung "Migration und horizontale Mobilität vom Mittelalter bis zum Ende des Ancien Régime" befindet sich im Druck (Campus-Verlag, Frankfurt am Main – New York) und wird noch heuer (für unsere Mitglieder als MAQ-Newsletter 11/12) zur Auslieferung gelangen.

Ein besonderer Hinweis muß auf den bereits im vorigen Heft erwähnten Kongreß gegeben werden, den *Medium Aevum Quotidianum* in Zusammenarbeit mit dem Institut für mittelalterliche Realienkunde der Österreichischen Akademie der Wissenschaften vom 27. bis 30. September 1988 in Krems veranstalten wird. Er befaßt sich mit dem Thema "Mensch und Objekt im Spätmittelalter und in der frühen Neuzeit. Leben – Alltag – Kultur" und läßt im Rahmen der Diskussion von international anerkanntesten Fachleuten vor allem im theoretischen und methodischen Bereich neue Ansätze und Wege erwarten. Voreinladungen an unsere Mitglieder wurden bereits vor einiger Zeit versandt.

Wie gewohnt wird sich einer der im Jahre 1988 erscheinenden Newsletter den Zusammenfassungen der am genannten Kongreß gebotenen Referate widmen. Darüber hinaus wird, wie bereits zuvor angekündigt, Teil II der Auswahl-Bibliographie zur materiellen

Kultur des Mittelalters erscheinen. Im Planungsstadium befinden sich einige Hefte, die sich mit der Erforschung von Alltag und materieller Kultur des Mittelalters in einzelnen Ländern auseinandersetzen werden und einen Schwerpunkt insbesondere auf Forschungs- und Literaturberichte legen sollen. Das erste Heft dieser Reihe wird vermutlich Dänemark gewidmet sein.

Gerhard Jaritz, Schriftleiter