

EuroClimHist als Werkzeug des Historikers¹

Datenbanken wurden bereits im Zeitalter der Großrechenanlagen, als zu deren Bedienung noch speziell ausgebildetes Personal erforderlich war, von einigen Historikern eingesetzt. Mit dem Aufkommen des PC und relativ billiger Software wurde die Zahl quantitativ arbeitender Sozialwissenschaftler größer. Heute ist der Einsatz von Textverarbeitungsprogrammen und einfachen Tabellenkalkulationen weit verbreitet, und auch die Zahl der Historiker, die Datenbanken einsetzen, wächst stetig.

Der vorliegende Artikel zeigt Anwendungsmöglichkeiten von Datenbanken für Historiker und geht näher auf EuroClimHist² ein, eine an der Universität Bern entwickelte Datenbank zur Speicherung von Klimadaten, die sich hervorragend eignet, den Nutzen solcher Instrumente für Historiker aufzuzeigen, ohne deren Komplexität zu verschweigen. Bedingt durch den interdisziplinären Charakter ergeben sich aber eine Reihe von Problemen: Schon die Darstellung solcher Projekte gestaltet sich wegen der Fachsprache der Informatiker schwierig, doch erst diese Fachsprache ermöglicht präzise Aussagen. In einem ersten Schritt soll versucht werden, EuroClimHist in der Forschungslandschaft zu positionieren. Dazu dient als Beispiel, und um die Unterschiede herauszuarbeiten, eine Datenbank, die anders aufgebaut ist als EuroClimHist. Danach erläutere ich kurz das Zustandekommen, die Entwicklung und die Struktur von EuroClimHist und zeige den heutigen Stand der Arbeiten.

Vom Quellentext zur Datenbank

Mit Hilfe eines selber entwickelten Makropakets, ASKSAM³, und mit MS-Access arbeiten Hagnauer und Bartlome⁴ die Berner Staatsfinanzen anhand von Jahresrechnungen aus verschiedenen Ämtern mit jeweils Tausenden von Buchungen auf.⁵ Diese geben Aufschluss über die Besoldung der Beamten, über die Auslagen der öffentlichen Hand für die Instandhaltung von Gebäuden, Straßen und Brücken, sowie über die Höhe der Zolleinnahmen. Darüber hinaus lassen sich aus den zusätzlichen Angaben in diesen Jahresrechnungen Hinweise auf den Speisezettel oder auf Art und Verlauf von Kriminalität herauslesen. Weil sowohl qualitative als auch quantitative Elemente in diesen Jahresrechnungen enthalten sind, werden in einem ersten

Schritt die zur Verfügung stehenden Quellen vollständig transkribiert, was den Erhalt aller Informationen, die eine Quelle aus inhaltlicher Sicht liefern kann, sicherstellt.⁶ Nach dieser sehr aufwendigen Arbeit werden die Texte inhaltlich bearbeitet. Dies umfasst die Suche nach der gewünschten Information, zum Beispiel einer Ortsangabe oder einer Buchung, und die Zuweisung eines Kodes über die direkt in Word eingebettete Programmoberfläche, um später von einem weiteren Programm diese Informationen »sammeln« lassen zu können. Neben der Zuweisung eines Kodes zu einer Einzelinformation wie einem Datum oder einem Buchungswert können auch ganze Textabschnitte einem oder mehreren Stichworten zugeordnet werden. Ebenso ist es möglich, eigene Kommentare direkt im Text anzubringen und durch einen speziellen Kode vom ursprünglichen Text abzugrenzen. Zusätzlich zu dieser Ad-hoc-Markierung und Kommentierung einzelner Textinformationen ist auch das systematische Klassifizieren (das Versehen mit Schlagworten) von Textabschnitten möglich, sofern die gewünschten Klassifikationssysteme als Listen, die hierarchisch sein dürfen, gespeichert sind. Der Benutzer kann Schritt für Schritt durch den Text gehen, die Listen nach dem präzisesten Schlagwort durchsuchen und den Abschnitt durch einen Mausklick diesem Schlagwort zuordnen. Durch die gut gestaltete Oberfläche, die dem geübten Benutzer zusätzlich noch die Verwendung von Tastaturbefehlen erlaubt, ist dieser Arbeitsschritt auch für den technisch weniger versierten Anwender kein Problem. Die Möglichkeit, eigene Listeneinträge und Untereinträge zu setzen und neue Kategorien zu definieren, lassen dem Benutzer völlige Freiheit bei der Wahl seiner Fragestellung in bezug auf die vorhandenen Texte und damit seiner Schlagworte. Damit können die erfassten Texte aus den unterschiedlichsten Perspektiven heraus mit Schlagworten versehen werden, was den Aufwand der eingangs erwähnten langwierigen Transkribierung rechtfertigt.

In einer zweiten Phase werden die mit zusätzlichen Informationen angereicherten Texte in den Informationsmanager ASKSAM übertragen. Die Texte behalten dabei ihren originalen Aufbau, werden aber nach Abschnitten oder individuellen Trennmarken auf eine Art von elektronischen Karten übertragen, ähnlich dem Vorgehen, das man wählen würde, wenn man die Arbeit von Hand und mit Karteikarten erledigen müsste. Im Gegensatz zur echten Karteikarte kann mit ASKSAM aber weit einfacher nach Worten oder Wortfolgen gesucht werden, die dann speziell gruppiert oder sortiert ausgegeben werden. Als weitere Stärke kommt die Flexibilität eines solchen Informationsmanagers hinzu, da anders als bei einer klassischen Datenbank sowohl strukturierte als auch unstrukturierte Informationen gespeichert werden können.

In der dritten Phase werden die Daten aus ASKSAM in eine relationale Datenbank exportiert, die viele mächtige Funktionen im Bereich Datendarstellung, Selektion, Sortierung, Berichtsgenerierung und logische Verknüpfung mit anderen Datenbeständen bereitstellt.

Sieht man also einmal vom Initialaufwand ab, sind die Vorteile doch offensichtlich. Die Quellen – im Prinzip Textquellen jeder Art – werden einmal vollständig

elektronisch erfasst und stehen dann als Basis zur weiteren Bearbeitung bereit, und zwar einem wesentlich breiteren Publikum, als das bei einem Zettelkastensystem der Fall wäre. Ob sich der Historiker nun ausschließlich quantitativ betätigt oder wie hier sowohl qualitative als auch quantitative Fragen an die Quellen stellt, spielt keine Rolle. Der Text kann unbeschränkt oft mit unterschiedlichen Fragestellungen bearbeitet werden, das Produkt des ursprünglichen Aufwands kann also intensiv genutzt werden. Sind die Texte einmal mit den Kodierungen versehen, ist ein wesentlicher Teil der heiklen Arbeit bereits getan: Fehler durch Unachtsamkeit beim Auszählen von irgendwelchen Beständen, das Weglassen von wichtigen Informationen, weil man eine Textstelle überlesen hat, oder falsche Umrechnungen von einer Währung in die andere gehören der Vergangenheit an, geht man von einem korrekt programmierten System und einer sorgfältigen Verschlagwortung aus. Die Verwendung einer Datenbank, die es erlaubt, gezielt nach Informationen zu suchen, die komplexen Bedingungen entsprechen, erlaubt praktisch ohne zeitlichen Mehraufwand eine Differenzierung der Fragestellung. In unserem Beispiel könnten Durchschnittswerte für einzelne Jahrgangsabrechnungen, alle Angaben zu einer bestimmten Person, eine Zusammenstellung aller Angaben, die Informationen zu Speisen enthalten oder die Gegenüberstellung der Jahreseinnahmen aller Ämter Ausgangspunkt für eine Datenbankabfrage sein.

EuroClimHist

EuroClimHist verfolgt etwas andere Ziele. Während das erste Projekt dem Historiker Hilfe bei alltäglichen Arbeiten bietet und universell einsetzbar ist, wurde EuroClimHist 1992 als Datenbasis für einen relativ engen Fragenkomplex entworfen.⁷ Gespeichert werden in dieser Datenbank eine Vielzahl von Daten über Wetterereignisse, Überschwemmungen, die Anzahl von Regentagen in einer Periode, aber auch Angaben zur Qualität von Ernten, zum Datum des Erntebeginns in einer Region und vieles mehr.⁸ Die Anzahl der erfassbaren Parameter liegt im Moment bei ungefähr 700 Angaben, die in einem sogenannten Kodebuch gespeichert sind. Werden Daten für diese Datenbank von Historikern aufgenommen, werden die Quellen nach Angaben, die im Kodebuch enthalten sind, durchforstet und nach gewissen Regeln transkribiert. Diese Transkribierung erfolgt nicht wie im ersten Beispiel vollständig, sondern nur partiell, weil die Transkription quasi identischer Beobachtungen von Witterungsphänomenen wie »sonnig« oder »Regen« wenig sinnvoll erscheint, was allerdings eine weitere direkte Nutzung der Quelle aus anderer Perspektive verhindert. Als Fließtext übernommen werden nur Textpassagen, die vom Üblichen abweichende Phänomene wie das Überfrieren eines Gewässers oder eine schwere Überschwemmung umschreiben. Darüber hinaus muss jeweils auch ein Datum oder eine Periode sowie eine Ortsangabe mit aufgenommen werden. Mit zur Zeit einigen hunderttausend Datensätzen, die Beobachtungen aus dem Zeitraum

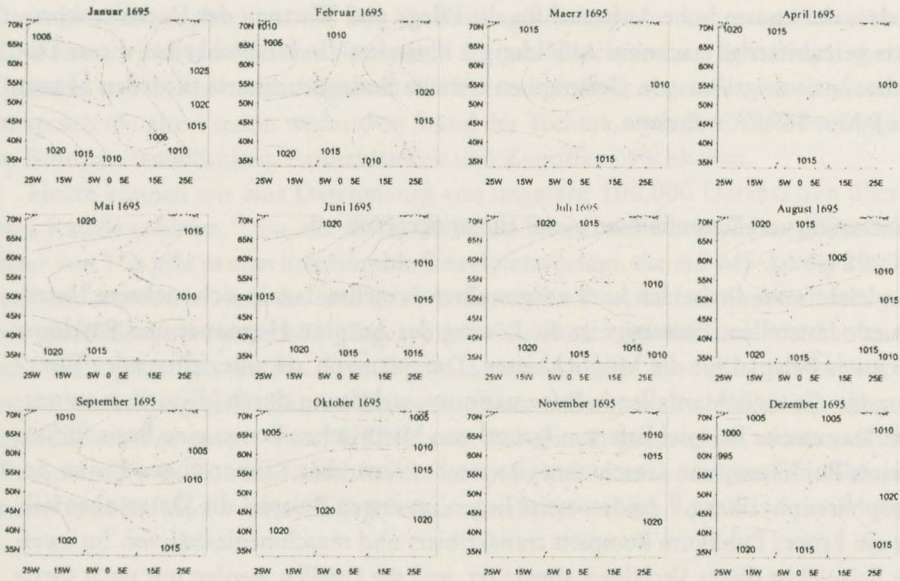


Abbildung 1: Karten des rekonstruierten mittleren monatlichen Bodendrucks in Europa im Jahr 1695

zwischen 1675 und 1715 enthalten, lassen sich detaillierte Statistiken zu verschiedensten Fragestellungen erstellen. Mögliche Auswertungen könnten sich auf die Anzahl von Regentagen pro Monat über die Dauer der gesamten Periode beziehen, auf einen Vergleich zwischen zwei Regionen in Europa bezüglich der Weinernten in qualitativer Hinsicht, oder auf die Anzahl der großen Überschwemmungen in dieser Zeit. Das erlaubt aus geschichtswissenschaftlicher Sicht die Bearbeitung einer breiten Palette an quantitativen und qualitativen Fragestellungen in den Bereichen Umweltgeschichte und Sozial- und Wirtschaftsgeschichte und eröffnet Perspektiven für überregionale Fragestellungen und Zusammenhänge.

Doch damit sind die Möglichkeiten des Instruments noch lange nicht ausgeschöpft. Durch komplizierte Transformationen und Umrechnungen von Wetterbeobachtungen sind heute die Klimatologen der Universität Bern unter der Leitung von Prof. Heinz Wanner in der Lage, Wetterkarten der Vergangenheit zu zeichnen.⁹ Das vorliegende Datenbanksystem erlaubt also die Speicherung von historischen Datenbeständen, die nach strengen Regeln aufgenommen werden und für ein breites Publikum von Interesse sind. Da Schnittstellen zu Grafikprogrammen, Statistikpaketen und einem GIS-Werkzeug¹⁰ vorhanden sind, ist die Weiterverarbeitung der Daten möglich. Die Ergebnisse können als Kurvendiagramme, statistische Angaben oder Karten, beispielsweise des mittleren monatlichen Luftdrucks auf Meereshöhe, präsentiert werden.

Um die Verwendung durch weitere Benutzer zu ermöglichen, sind die Daten weltweit einfach abrufbar. Dies wird durch eine Internetschnittstelle gewährleistet,

so dass der enorm hohe Aufwand für die Pflege und Wartung des Datenbankprojekts gerechtfertigt erscheint. Abbildung 1 illustriert die Möglichkeiten dieser Datenbank; sie zeigt eine von Geographen erstellte Bodendruckkarte für jeden Monat des Jahres 1695¹¹ in Europa.

Typisierung von Datenbanken in der Historiographie

Vergleicht man die beiden kurz vorgestellten Projekte, lassen sich mehrere Unterschiede feststellen. Einerseits ist die Lösung der Autoren Hagnauer und Bartlome ein gutes Beispiel für die Möglichkeiten, Datenbanken als unterstützendes Werkzeug für die täglich anfallende Informationsverarbeitung durch Historiker zu nutzen. Das zweite Beispiel liefert aufgrund von historischen Daten eine Basis für ein breites Publikum, von Geschichtswissenschaftlern über Klimatologen bis zu Atmosphärenphysikern.¹² Andererseits liegen im ersten Beispiel die Daten ebenfalls für ein breites Publikum komplett transkribiert und maschinenlesbar vor. Im zweiten Beispiel ist dieses Verfahren abgekürzt, und die Quellen werden nur nach spezifischen Informationen durchforstet, die extrahiert und mit einer kurzen Textpassage für eine Überprüfung gesichert werden. Damit finden sich die Daten direkt von der Quelle in die Datenbank geschrieben, ohne den ›Umweg‹ über die Textverarbeitung und den Informationsmanager zu nehmen. Die Vor- und Nachteile sind offensichtlich. Möchte ein Forscher auf Quellenbestände zurückgreifen, die durch Mitarbeiter von EuroClimHist bearbeitet worden sind, um sie nach anderen Inhalten als den bereits im Kodebuch gespeicherten zu durchsuchen, ist ein weiterer Gang ins Archiv unvermeidlich, während die Daten im ersten Beispiel im schlimmsten Fall ohne vorhandene Verschlagwortung, aber vollständig transkribiert weitergegeben werden können. In der Literatur für Historiker werden dementsprechend quellenorientiertes und modellorientiertes¹³ Vorgehen unterschieden; der Informatiker kennt die beiden Typen als Netzwerk- beziehungsweise relationales Modell. Wie oben gezeigt wurde, muss sich derjenige, der die modellorientierte Variante bevorzugt, früh entscheiden, welche Daten wie gespeichert werden sollen, während beim quellenorientierten Ansatz diese Entscheidung jedem Arbeitsschritt angepasst werden kann. Das bedeutet bezüglich des modellorientierten Ansatzes auch, dass sich der Forscher früh über die Quellenlage, die Fragestellung und seine Ziele klar sein muss.

EuroClimHist heute

EuroClimHist war früher klar dem quellenorientierten Typus zuzurechnen. Heute ist es eine neue Datenbank im Stadium des Prototyps, die dem relationalen Typus zuzurechnen und damit dem Gros der in der Geschichtswissenschaft verwendeten

Datenbanken vergleichbar ist. Die Umstellung erfolgte aus mehreren Gründen: Zum einen wurden die für das alte System nötigen Maschinen aus lizenzrechtlichen Gründen von den Informatikdiensten der Universität Bern abgestellt, zum anderen entsprach die alte Version weder dem Stand der Technik noch den Anforderungen an Benutzerfreundlichkeit, Ausbaubarkeit und Zugriffsmöglichkeiten.

Heute können wir eine Datenmenge von ungefähr 100.000 Datensätzen über zwei Kanäle anbieten.¹⁴ Für alle abfragbar ist die Datenbank über Internet,¹⁵ für Besitzer von PCs gibt es eine komfortable Einzelplatzversion, die mit MS-Access 2000 realisiert worden ist. Auf der einfach zu bedienenden Oberfläche können eine Periode spezifiziert, ein bestimmter geographischer Raum ausgewählt und eine thematische Wahl getroffen werden. Das Ergebnis einer Abfrage wird als Liste in einer neuen Maske dargestellt. Für die Weiterverarbeitung kann es anschließend in unterschiedlichen Formaten (Text, HTML oder MS-Excel) gespeichert werden. Auf Wunsch sollen in Zukunft thematische Europakarten in verschiedenen Grafikformaten lieferbar sein, sodass die Ergebnisse visualisiert werden können.

Fazit

Eine Diskussion, welcher der beiden Methoden unter welchen Umständen der Vorzug zu geben ist, würde den Rahmen dieses Beitrags sprengen. Aus welchen Gründen auch immer man sich für den Einsatz einer Datenbank entscheidet, und egal, ob man den quellenorientierten oder den modellorientierten Typ wählt, sicher ist, dass man sich mit den Methoden der Informatik gründlich auseinandersetzen hat. Ich habe meine Erfahrungen fast ausschließlich mit dem relationalen Modell gemacht und neben eigenen Projekten eine größere Anzahl beratend begleitet. Erfahrungsgemäß kommen sich traditionelle Arbeitsweisen des Historikers und Anforderungen, die ein Softwareprojekt an die Beteiligten stellt, in die Quere. Ebenso oft werden die Komplexität heutiger Softwaresysteme und die ihnen zugrunde liegenden Konzepte unterschätzt. Fehlendes Fachwissen und eine damit verbundene unzulängliche Planung und Realisierung lassen solche Projekte in vielen Fällen wenn nicht scheitern, so doch aufwendiger und teurer werden als geplant. Daher stehen Historikern, die Datenbanken für ihre Projekte einsetzen wollen, grundsätzlich zwei Wege offen: Der eine bedeutet Auseinandersetzung mit und Einarbeitung in die Bereiche Programmierung, *Software Engineering* und Datenbanktechnologie mit Hilfe der entsprechenden Informatik-Fachliteratur, die andere Möglichkeit besteht darin, solche interdisziplinäre Fragestellungen mit Kollegen aus der Informatik anzugehen und zu lösen.

Anmerkungen

- ¹ Dieser Artikel ist im Rahmen des vom Schweizerischen Nationalfonds geförderten Projekts FLOODRISK (11-52786.97) entstanden.
- ² Vgl. den Artikel von Christian Pfister in diesem Heft.
- ³ Vgl. <http://www.asksam.com/>; u. Ines Grund, Von der Urkunde zum Regest – EDV-Anwendung im Editionsprojekt »Regesten Kaiser Friedrichs III. (1440-1493)« an der Akademie der Wissenschaften und der Literatur, in: Arbeitsgemeinschaft Geschichte und EDV, Hg., Geschichte und EDV: Probleme und Fortschritte – Probleme mit dem Fortschritt, Bochum 1997, 69-80. In dieser Anwendung wird ebenfalls mit ASKSAM gearbeitet.
- ⁴ Vgl. Stephan Hagnauer u. Niklaus Bartlome, Vom Quellentext zur Datenbank – ein Konzept zur integrierten Verarbeitung quantitativer und qualitativer Daten, in: Peter Bär, Gerold Ritter, u. Christophe Koller, Hg., Geschichte und Informatik 9 (1998), 141-158.
- ⁵ Vgl. Andreas Kellerhals, Finanzgeschichte mit dBase, in: Hannes Schüle, Hg., Geschichte und Informatik 1 (1990), 50-55.
- ⁶ Vgl. zum folgenden Hagnauer u. a., Quellentext, wie Anm. 4, 141 f.
- ⁷ Vgl. Hannes Schüle, Outlines for a European Databank for Historical Climate Data – The "EURO-CLIMHIST" Project, in: Hans Jørgen Marker u. Kirsten Pagh, Hg., Yesterday. Proceedings from the 6th International Conference Association of History and Computing, Odense 1991, 252-257; u. Christian Pfister u. a., Ein Programm- und Methodenpaket zur Rekonstruktion von Klimaverhältnissen seit dem Hochmittelalter, in: Manfred Thaller, Hg., Datenbanken und Datenverwaltungssysteme als Werkzeuge Historischer Forschung, St. Katharinen 1986, 75-92, hier 75 f.
- ⁸ Vgl. Christian Pfister, Witternachhersage. 500 Jahre Klimavariationen und Naturkatastrophen, Bern 1999, 15-19.
- ⁹ Vgl. Jürg Luterbacher u. a., Reconstruction of Monthly Mean Sea Level Pressure over Europe for the Late Maunder Minimum Period (1675-1715) based on Canonical Correlation Analysis, in: International Journal of Climatology 20 (2000), in Druck.
- ¹⁰ GIS bedeutet Geographisches Informationssystem und erlaubt die Darstellung verschiedenster Daten auf einer Karte.
- ¹¹ Vgl. Luterbacher, Reconstruction, wie Anm. 9; ich danke Jürg Luterbacher für die Bereitstellung der Karte.
- ¹² Während eines Aufenthaltes 1999 in Boulder, Colorado, konnte ich mit einem Atmosphärenphysiker sprechen, der für seine Tornadoforschung ebenfalls auf Dokumentendaten zurückgreift und ein System entwickelt, das EuroClimHist sehr ähnlich ist.
- ¹³ Charles Harvey u. Jon Press, Databases in Historical Research, London 1996, 11.
- ¹⁴ Für die Zukunft ist mit zwischen einer halben und einer Million Datensätzen zu rechnen.
- ¹⁵ Vgl. <http://histserver.unibe.ch/euroclimhist/default.html>; unter dieser Adresse sind heute einige tausend Datensätze verfügbar; die Datenbank befindet sich aber noch in einem frühen Prototypstadium und wird wahrscheinlich ab Mitte Mai einer breiteren Öffentlichkeit zugänglich gemacht.