Zeitschrift: Geschichte und Informatik = Histoire et informatique

Herausgeber: Verein Geschichte und Informatik

Band: 5-6 (1994-1995)

Artikel: Datenbanken = Datenleichen? Langfristige Zugangssicherung und Pflege maschinenlesbarer historischer Datensätze in der Schweiz

Autor: Pfister, Ulrich / Pfister, Christian / Sardet, Frédéric

DOI: https://doi.org/10.5169/seals-118841

Nutzungsbedingungen

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

Conditions d'utilisation

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

Terms of use

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

Download PDF: 08.10.2025

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

Datenbanken = Datenleichen? Langfristige Zugangssicherung und Pflege maschinenlesbarer historischer Datensätze in der Schweiz

Ulrich Pfister (Redaktion), Christian Pfister, Frédéric Sardet, Hannes Schüle und Lukas Vogel

Einleitung (Ulrich Pfister)

Der gegenwärtige Text geht aus einer Gesprächsrunde hervor, die der Schreibende an der Jahrestagung der Gesellschaft für Geschichte und Informatik in Lausanne (6. November 1992) veranstaltet hat. Zur Runde wurden drei Forschungsgruppen eingeladen, die langfristige Forschungsprojekte unterhalten, in deren Rahmen statistisches Datenmaterial unterschiedlichen Charakters erhoben und ausgewertet wird. Einerseits wurde davon ausgegangen, dass diese drei Gruppen bereits Erfahrung mit dem Problem langfristiger Pflege maschinenlesbarer Daten und bei seiner Bewältigung gesammelt haben. Andererseits ist in ihnen auch eine wichtige Zielgruppe einer institutsübergreifenden Lösung zu sehen. Den drei Gruppen wurde vorgängig ein Fragebogen vorgelegt, in dem Auskünfte zur Datenstruktur und Datenbankanwendung, zu den organisatorischen Problemen des Datenbankprojekts und zu den Aussenbeziehungen (Datenübernahme aus Fremdprojekten, projektexterne Kunden) erbeten wurden. Auch wurden sie eingeladen, Gedanken zu einer projekt- und institutsübergreifenden nationalen Lösung hinsichtlich der langfristigen Wartung maschinenlesbarer Daten in den Geschichtswissenschaften zu äussern. Der Aufbau der nachfolgenden drei Abschnitte folgt weitgehend dem Fragebogen, so dass Vergleiche leicht möglich sind. Diese kurzen Texte bieten einen knappen Überblick über die Ziele, Arbeitsweise und organisatorischen Probleme der unseres Wissens grössten Datenbankprojekte in der schweizerischen Geschichtswissenschaft. Zusätzlich zu den drei Forschungsgruppen wurde Hannes Schüle, wohl dasjenige Mitglied der Gesellschaft, das in den vergangenen Jahren den intensivsten Kontakt mit gleichgearteten Verbänden im Ausland gepflegt hat, gebeten, einen kurzen Überblick über Modelle nationaler Archive maschinenlesbarer historischer Daten vorzutragen. Die gegenwärtige Einleitung führt in die Problematik ein und hält einige Überlegungen der Diskussion fest.

Seit dem Beginn der modernen historischen Forschung spielt die Sekundäranalyse eine nicht zu unterschätzende Rolle in unserem methodischen Instrumentarium. Nie publizierte genealogische Zusammenstellungen, zu-

nächst zu projektspezifischen Zwecken angelegte Regestensammlungen und andere Zettelkästen, allesamt vielfach schon von respektablem Alter und in Bibliotheken oder Archiven öffentlich zugänglich, stellen wertvolle Stützen des Forschungsprozesses dar. Der Einsatz von Informatikmitteln lässt in diesem Zusammenhang Probleme entstehen, deren Implikationen und Lösungen noch kaum durchgedacht und durchdiskutiert worden sind. Einerseits erhöhen Informatikmittel unsere Kapazität, grosse Mengen an Primärquellen mit vertretbarem Aufwand zu erfassen und zu verarbeiten. Andererseits erschwert die (im Vergleich mit Tinte und Papier) komplexe Speichermethode und der rasche Wandel von Hardware-Plattformen und Anwendungen eine langfristige Konservierung dieses Materials. Dabei wäre bei der gestiegenen Kapazität zur Verarbeitung von Primärmaterial die Sekundäranalyse ein wichtiges Element im Bemühen, den Forschungsprozess durchschau- und überprüfbar zu erhalten.

In der Praxis führt dieses Dilemma dazu, dass die Lebensdauer maschinenlesbarer Daten kaum diejenige eines individuellen Projekts übersteigt. Nach dem Projektabschluss werden die Daten auf Band oder Diskette gespeichert, vielfach verlassen zudem die Projektmitarbeiter das Institut. Nachfolgende Assistenten wissen mangels ausreichender Dokumentation nicht mehr, wie sie dem Material sinnvolle Auskünfte entlocken können, und spätestens beim nächsten Umzug des Instituts oder bei der Emeritierung des Projektleiters werden die Daten zerstört. Im Vergleich mit den in Archiven lagernden handschriftlichen Nachlässen früherer Forscher eine ziemliche Katastrophe! Eine Situation zudem, die auch Institutionen der Forschungsförderung zu denken geben sollte, denn der Forderung, dass Forschungsergebnisse (zu denen als wertvolle Bestandteile auch im Rahmen eines Projekts erarbeitete Sekundär- oder Metaquellen gehören) der interessierten wissenschaftlichen Öffentlichkeit zur Verfügung zu stehen hätten, wird so nicht nachgelebt.

Eine Lösung kann in zwei Richtungen vorangetrieben werden, die sich mit den Stichworten Datenbank und Datenarchiv bezeichnen lassen. Die beiden Richtungen können sich ergänzen und überschneiden. Die Datenbank löst das Problem der Datenkonservierung vornehmlich auf der Software-Seite, das Datenarchiv setzt im Bereich der Organisation ein. Die Funktionalität der beiden Lösungen kann sich stark unterscheiden.

Das gängige relationale Modell einer Datenbank verwaltet mehrere Tabellen, in denen die Zeilen möglicherweise unterschiedliche Entitäten (Personen, Gemeinden, Jahre usw.) und die Kolonnen Merkmale dieser Entitäten (Alter, Bruttosozialprodukt usw.) bezeichnen. Die Datenbankanwendung verwaltet einerseits die Eingabe neuer Daten, andererseits erlaubt sie die Datenextraktion in strukturierter Form über mehrere Tabellen hinweg.

Da die Datenbank eine stark formalisierte Datenstruktur voraussetzt, eignet sie sich vor allem zur Verwaltung von statistischen Datensätzen; alle drei unten vorgestellten Projekte verwalten statistisches Material. Die Funktion der Datenbank im Bereich der historischen Statistik ist vor allem akkumulativer Natur: Daten aus verschiedenen Projekten können zusammengeführt und im Hinblick auf neue Fragen ergänzt und ausgewertet werden. Sowohl BERNHIST wie auch FSWbase sind derartige Datenbanken. Über eine reine Archivierung von Daten hinaus werden diese unter einer einheitlichen Benützeroberfläche verfügbar und können so rasch weiterverwendet werden.

Vergleicht man kurz die nachfolgend präsentierten Projekte, so fällt auf, dass sie ähnliche, offene Entwicklungswerkzeuge benützen; alle haben (wenigstens in einer bestimmten Phase) Oracle eingesetzt, und alle verwendeten Werkzeuge haben ein SQL-Interface. Diese Arbeitsweise eröffnet für die Zukunft ein wichtiges Potential des Datentauschs. Die Entwicklung der Datenbankanwendung ist allerdings unterschiedlich weit gediehen. Während mit BERNHIST ein ausgereiftes Produkt vorliegt und die Historische Datenbank Genf wenigstens projektintern im Alltag lauffähig ist, befindet sich FSWbase noch in der Projektierungsphase. Die personelle Kontinuität ist nur bei BERNHIST wenigstens mittelfristig gesichert; im Gegensatz zu einem Datenarchiv sind diese Datenbanken noch weitgehend an den Lebenszyklus von Einzelprojekten gebunden. Auch im Bereich der Funktionalität bestehen beträchtliche Unterschiede: Während die Historische Datenbank Genf stark projektorientiert scheint, versteht sich FSWbase primär als institutsinternes Datenarchiv. BERNHIST schliesslich ist klar multifunktional ausgerichtet, wird doch neben der Unterstützung der Institutsforschung auch die Schule als Kundin anvisiert. Darüber hinaus ist die Datenbank 1991 mit Erfolg an der Ausstellung BEA einer weiteren Öffentlichkeit vorgestellt worden. Das bewusste Ausnützen einfacher, «intuitiver» Benützeroberflächen durch dieses Projekt zur Vermittlung eines neuartigen Zugangs zur Geschichte auch für ein nicht-wissenschaftliches Publikum könnte mit der relativen Stabilität seiner finanziellen Basis in einem gewissen Zusammenhang stehen. Eine Zusammenarbeit mit weiteren Kantonen wird angestrebt.

Ein Datenarchiv bietet in der Regel keine Software-gestützte Datenverwaltung an und gewährt damit keine Handhabe für eine unmittelbare Weiterverwendung von Daten. Während die Datenbank wohl ein unerlässliches Instrument zur Verwaltung umfangreicher Datensätze im alltäglichen Forschungsprozess darstellt, bietet das Datenarchiv eine Infrastruktur zur langfristigen Lagerung und Pflege unterschiedlichster, thematisch nicht notwendigerweise verbundener Datensätze an. Über die in den hier vorgestellten

Projekten im Vordergrund stehenden statistischen Daten hinaus rücken damit insbesondere auch textuelle Datensätze ins Blickfeld. Neben reinen Textkorpora (Editionen, Regesten) ist insbesondere auf Prosopographien oder Kollektivbiographien hinzuweisen, zu deren Verwaltung weniger Datenbankwerkzeuge als Abfragesysteme verwendet werden (information retrieval systems). Beispiele sind etwa die unter Prof. Schmugge in Zürich erarbeiteten Prosopographien geistlicher Stifte in der Schweiz.

Hauptziel eines Datenarchivs ist in der Regel die langfristige Lagerung von elektronischen Datenbeständen (Lagerarchiv). Die wichtigsten Funktionen sind die Pflege der Datenbestände (Erhaltung ihrer Lesbarkeit), die Dokumentation und die Weitergabe von Daten an ForscherInnen zwecks Reanalyse. Insbesondere die Dokumentation kann Ausgangspunkt für eine weitergehende Beratungstätigkeit eines Datenarchivs darstellen. Die Unterstützung von individuellen ForscherInnen und Forschungsgruppen bei der Dokumentation der von ihnen deponierten Datensätze ist zunächst selbst eine wichtige und zeitaufwendige Arbeit. Die damit verbundene Produktion von Normen hinsichtlich Datendarstellung und Datendokumentation ist jedoch darüber hinaus von Nutzen sowohl für den Aufbau von neuen Projekten als auch im Hinblick auf eine standardisierte Kommunikationskultur, die eine Reanalyse von Daten erheblich erleichtert.

In der Diskussionsrunde von Lausanne wurde vielfach nicht klar zwischen Datenbank und Datenarchiv unterschieden. Tatsächlich ist es bis zu einem gewissen Grad offen, inwieweit unsere Bemühungen in die Richtung des Aufbaus einer interaktiven historischen Statistik der Schweiz gehen sollten (ein massiv ausgeweitetes StatInf also) oder in die Richtung eines Lagerarchivs (in der Art von SIDOS). Gerade im Bereich aggregierter Daten (auf Gemeinde- bzw. Bezirksebene) dürfte heute für den Aufbau einer interaktiven historischen Statistik mittels einer Datenbankanwendung ein grosses Potential bestehen. Die Bedürfnisse einer weiteren Forschungsgemeinde, deren maschinenlesbare Daten sich nur zum kleineren Teil auf aggregierte statistische Daten beziehen, können jedoch damit nicht abgedeckt werden.

Hinsichtlich der Trägerschaft eines Lagerarchivs für maschinenlesbare historische Datensätze wurden in der Diskussion unterschiedliche Vorstellungen geäussert. Eine auch anderswo öfters zu vernehmende Ansicht lautet, dass Bibliotheken, statistische Ämter, Rechenzentren und Archive als dezentrale Träger von Datenarchiven fungieren könnten bzw. zu fungieren hätten. Dieser Ansicht ist aus mehreren Gründen zu widersprechen. Rechenzentren sind nur auf der Hardware-Seite für die Datenpflege verantwortlich (insbesondere hinsichtlich der Migration auf neue Hardware-Plattformen); im Dokumentationsbereich können sie keine Unterstützung anbieten.

Archive und andere Amtsstellen haben in der Regel einen gesetzlichen Auftrag, der die Archivierung von Sekundärdaten aus Forschungsarbeiten nicht umfasst. Aus diesem Grund hat an einer anderen Tagung der Gesellschaft Geschichte und Informatik ein prominentes Archiv die Funktion des Lagerarchivs für maschinenlesbare historische Daten ausdrücklich abgelehnt. Es sei auch darauf hingewiesen, dass die genannten dezentralen Institutionen in der Regel nicht über das erforderliche Know-how zur langfristigen Datenpflege und -dokumentation verfügen.

Auch wenn man mehr oder weniger zwangsläufig auf eine ausdifferenzierte, zentrale (nationale) Trägerschaft verwiesen wird, sind mehrere Optionen offen. Dies zeigt nicht zuletzt der Beitrag von Hannes Schüle (S. 119ff.). Abgesehen von zentralen statistischen Datenbanken (Schweden, USA) bestehen im Bereich des Datenarchivs die Möglichkeiten des Anschlusses an ein zentrales Datenarchiv (wie in Dänemark) oder der Aufbau eines eigenen historischen Datenarchivs (Leiden). In der Schweiz werden wir uns an der vorhandenen, von der Schweizerischen Akademie der Geistes- und Sozialwissenschaften (SAGW) gestützten Infrastruktur zu orientieren haben. Im Vordergrund steht dabei der neugeschaffene SIDOS, sekundär ist an die Datenbank Schweizerischer Kulturgüter (DSK) zu denken. Der Rahmen für die Weiterarbeit ist damit abgesteckt: Er besteht einerseits in der Kontaktnahme mit den genannten Institutionen, insbesondere mit dem SIDOS, andererseits in der Sensibilisierung der SAGW für unsere Bedürfnisse.

Der Vorstand des Vereins Geschichte und Informatik beabsichtigt, die Thematik in diesem Sinn weiter zu verfolgen. Der Kontakt mit ForscherInnen und Forschungsgruppen, die grössere Datenbestände aufbauen oder verwalten, ist deshalb sehr erwünscht. Wir begrüssen eine Kontaktaufnahme über das Sekretariat der Gesellschaft.

Das Historisch-Geographische Informationssystem BERNHIST (Christian Pfister, Hannes Schüle)

1. Daten und Projekt. *Wir halten uns an die Numerierung des Fragebogens, stellen aber Punkt 1.2. (Zielsetzung) dem Punkt 1.1. (Datenbasis) voran.*

1.2. Zielsetzung. Erkenntnisleitende historische Fragestellungen in weiterem Sinne lassen sich nie auf ein einzelnes Subsystem der Gesellschaft und das theoretisch-methodische Instrumentarium der betreffenden Teildisziplin einschränken. Sie zielen stets auf die Gesellschaft als ganze und müssen daher **fächerübergreifend** angegangen werden.

Durch die relationale Datenbank BERNHIST sollen – zunächst am Beispiel des Kantons Bern als Pilotprojekt – neue Voraussetzungen zur *«systematischen Aneignung historischer Prozesse»* (Claudia Honegger) geschaffen werden. Im Vordergrund steht in einem ersten Schritt der gesellschaftliche *«Modernisierungsprozess»* von 1750 bis 1914, insbesondere hinsichtlich der Wechselwirkung seiner demographischen, ökonomischen, sozialen und ökologischen Komponenten, ausserdem die damit verbundene Veränderung räumlicher Disparitäten **(«Zentren und Peripherien»)**.

1.1. Inhalt der Daten

1.1.1. Konzept. Zur Erreichung dieses Zieles bedarf es eines Informationssystems, welches:

- die für die intersubjektive Überprüfung nötige **Transparenz** bietet
- von Beginn weg Schlüsseldaten für mehrere Subsysteme der Gesellschaft enthält und für thematische Erweiterungen offen ist **(Trans-Disziplinarität)**
- auf einem überzeitlich gültigen räumlich-territorialen Grundraster beruht **(räumliche Vergleichbarkeit)**
- auch nominelle serielle Information, beispielsweise in Form von gemeindebezogenen Personenlisten, verwalten kann
- die Ergebnisse von laufenden Forschungen integrieren, adäquat dokumentieren und damit für weitere Fragestellungen fruchtbar machen kann **(Datenrecyling)**.

1.1.2. Inhalt. Serielle Daten aus protostatistischen oder statistischen Erhebungen, die den Kanton oder ein grösseres Teilgebiet flächendeckend erfassen.

Potentiell sollen alle Themenbereiche einbezogen werden können, welche Subsysteme der Gesellschaft und ihr Verhältnis zur natürlichen Umwelt betreffen (Konzept des Bundesamtes für Statistik 1985).

1.3. Umfang des Datenbestandes. Die Datenbank umfasst insgesamt drei ME (Mega-Entitäten)[1], maschinenlesbare, mit Schlüssel für Raum, Zeit und Term greifbare Einzelinformationen. Davon sind zur Zeit ca. 20% direkt abrufbar. Nach gut einjährigem Unterbruch wurde die Arbeit am Gefäss der Datenbank im November 1992 wieder aufgenommen. Die vollständige Übernahme der Daten ist Ende 1994 abgeschlossen worden.

1.4. Beginn und Ende. Zum besseren Verständnis muss hier zwischen dem Inhalt des Historisch-Geographischen Informationssystems (Pfister/Schüle, Datenedition) auf der einen Seite und dem 1990–1991 entwickelten Gefäss

1 ME: Eine Million Informationseinheiten: Zahlen, Fakten, Daten. Zum Begriff der Entität in diesem Zusammenhang vergleiche Schüle/Siffert 1991 und Pfister/Schüle 1992.

mit Abfrageoberfläche (Imfeld et al. 1995, Design und Applikationsentwicklung) unterschieden werden.

Die drei Schritte auf dem Wege zu BERNHIST

1. Schritt: 1983–1984: Erhebung **flächendeckender demographischer Aggregatdaten** (Taufen/Geburten, Sterbefälle) auf der Ebene der Kirchgemeinden für den Zeitraum 1700–1980 und Umsetzung in Zeitreihen (im Auftrag des Amts für Unterrichtsforschung).
2. Schritt: 1986–1989: **Nationalfondsprojekt «Bevölkerung, Wirtschaft und Umwelt im Kanton Bern 1700–1980»:** Aufnahme und quellenkritische Prüfung ergänzender Daten aus den Surroundings (Land- und Forstwirtschaft, Gewerbe, Industrie, Umwelt, Energie, Soziales) in Form **sequentieller Files** (Software SPSS^{X}) als Kern einer «Datenbank» Historische Statistik. Umsetzung als Karten, Tabellen und Zeitreihen auf verschiedenen räumlichen Ebenen (Kanton, Landesteile, Amtsbezirke, Gemeinden). Berechnung von Metaquellen (Landnutzung, Agrarproduktion, Ernährung, «Tragfähigkeit»).
3. Schritt: 1989–1991: **Projekt BERNHIST:** Restrukturierung des Datenmaterials in einer **relationalen Datenbank** mit einer benutzerfreundlichen Oberfläche in Form einer raschen PC-Version (auf der Basis von Oracle) und einer Mainframe-Version (auf der Basis von RdB). Präsentation der PC-Version an Ausstellungen BEA und OHA im Rahmen des 800jährigen Jubiläums des Kantons (BE800).

Anschliessend an eine Finanzierungslücke in den Jahren 1991/92 folgt der

4. Schritt: 1993–1996: **Die Publikation BERNHIST** auf drei Ebenen:

a) Ausbau **zur öffentlichen Datenbank** (vgl. dazu Imfeld et al. 1995)
b) Publikation einer **wissenschaftlichen Monographie** (Pfister, 1995)
c) Publikation eines gut bebilderten, mit farbigen Karten ausgestatteten, knapp und plakativ kommentierten **Historisch-Statistischen Atlasses** zu Handen eines gebildeten Laienpublikums (gestützt auf obige wissenschaftliche Publikation).

2. Datenstruktur und Datenbankanwendung

2.1. Datenformat. Die Daten lagen bis 1990 in Form von SPSS^{X}-Files auf einem IBM-Grossrechner vor. Die neue Datenbank ist mit dem Produkt RdB (von DEC) auf einem Alpha-Computer der Risk-Generation realisiert worden. Jede Entität ist mit den Schlüsseln Raum, Zeit, Term (BERNHIST-Würfel) abgelegt, zwei hierarchische Thesauri kontrollieren die Term- und die Raum-Achse.

Einlese- und Austauschformate führen einen Header mit allen Termen und deren Supratermen sowie den Raumnamen und den «Supra»-Raum-

namen zur Einordnung in die entsprechenden Thesauri und beschreiben maschinenlesbar den folgenden Inhalt des Files[2].

Realisiert ist auch das Fussnoten-Konzept, welches zu jeder Entität verschiedene, klassierte Anmerkungen mitführen und ebenso maschinengenerierte Fussnoten für aggregierte Werte selbst produzieren kann[3].

2.2. Datenverwaltung. Eine interaktive Auswahlmaske erlaubt, spezifische Termini und Raumeinheiten aus den jeweiligen Thesauri auszuwählen und die Auswahl zeitlich auf eine beliebige Subperiode einzuschränken.

2.3. Eigenentwicklungen. Die gesamte Übernahme, die Oberfläche und die Weitergabe an SAS oder Harvard-Grafics ist in der Programmierungssprache geschrieben worden.

3. Wechselfälle von Manpower, Hard- und Software

3.1. Verantwortlichkeit, Datenbankadministration. Innerhalb des BERNHIST-Teams sind die Posten des Datenbank-Administrators (DBA) und des Aussenverantwortlichen fest vergeben. Für einzelne Entwicklungsschritte werden jeweils eigenverantwortliche Arbeitsgruppen zusammengestellt.

3.2a. Wechsel von Personen. Der grosse Wechsel im Frühjahr 1990 ging mit dem Beginn des Aufbaus der RdB-Datenbank vom Einzelverantwortlichen (Schüle) zum Team einher. Der Wechsel vom Team 1990/91 zum Team 1993/94 war insofern problematisch, als eine für Frühjahr 1992 geplante fliessende Übergabe wegen Verzögerung der Projektbewilligung nicht abgewickelt werden konnte. Durch die inzwischen erfolgten und begonnenen Studienabschlüsse war die Kontinuität in Frage gestellt.

3.2b. Wechsel der Hard- und Software. Der Wechsel von SPSS^{X} auf IBM zur Datenbank auf VAX verlief relativ problemlos. Die Probeme liegen nicht im EDV-Bereich, sondern im Bedarf nach stringenten Thesauri für eine Datenbank. Die Migration von VAX auf Alpha hat nur Vorteile gebracht.

Hingegen musste die ursprünglich auf Oracle auf der Microvax des Bernischen Mittelschul-Informationssystems (MIS) begonnene Entwicklungsarbeit wegen der unerwarteten Infragestellung der Zukunft des MIS durch die Verwaltung notfallmässig auf RdB umgestellt werden.

2 Abweichend von Thaller (1990, 1992) wird bei BERNHIST der gesamte maschinenlesbare File-Beschrieb nicht in einem Kontroll-File, sondern als Header geführt.

3 Etwa in der Form: *«12% der Ursprungsdaten sind unsicher, 7% sind interpoliert».*

3.3. Wechsel im Entwicklungs-Tool. Siehe Punkt 3.2b.

4. Aussenbeziehungen

4.1. Übernahme von Daten. Übernommen werden verschiedene aktuelle Basiszahlen via Magnetband oder e-mail vom Bundesamt für Statistik (etwa Ergebnisse neuer Volkszählungen) und von der bernischen Staatsverwaltung (etwa Wahl- und Abstimmungsresultate).

4.2. Weitergabe von Daten. Es besteht die Möglichkeit des öffentlichen Zugriffs mit Terminal in der Schulwarte, in der Stadtbibliothek Bern, über das – inzwischen gerettete – MIS und über Internet. In grossem Ausmasse wurden Daten als Excel-Tabellen über das MIS verbreitet. Vereinzelt werden Daten als Printout oder auf Diskette auf Anfrage weitergegeben.

4.3. Benutzerregelung. Zur Zeit haben nur Personen auf die gesamte Datenbank Zugriff, welche einschlägige Arbeiten in der Forschungsstelle für Regional- und Umweltgeschichte der Universität Bern verfassen oder an liierten Projekten mitarbeiten. Der Öffentlichkeit (siehe Punkt 4.2.) werden bis zur Fertigstellung der wissenschaftlichen Auswertung aus naheliegenden Gründen nur Teile der Daten freigegeben.

4.4. Copyright. Das **gesamte BERNHIST** wird **bezeichnet** als «Historisch-Geographisches Informationssystem BERNHIST realisiert durch die Forschungsstelle für Regional- und Umweltgeschichte am Historischen Institut der Universität Bern (Leitung: Prof. C. Pfister)».

Datenbank-Gefäss und -Oberfläche sind als Auftrag erstellt worden und werden wie folgt **zitiert**: «Rüfenacht, Thomas, Ruetsch, Stefan, Datenbankdesign und Applikationsentwicklung des Historisch-Geographischen Informationssystems BERNHIST, i.V.». Die wissenschaftliche «Vermarktung» liegt grundsätzlich in den Händen der Autoren.

Für den **Inhalt der Datenbank** wird die Autorschaft grundsätzlich gleich gehandhabt wie bei einem Sammelband bzw. einer Edition. Pfister und Schüle zeichnen als Herausgeber, die «ProduzentInnen» und AuswerterInnen einzelner Datenbestände als EinzelautorInnen. Daten werden wie folgt **zitiert**: «Pfister, Christian, Schüle, Hannes, Datenedition Historisch-Geographisches Informationssystem BERNHIST, Bern 1990ff.»

5. Schweizerisches Datenarchiv

Grundsätzlich begrüssen wir die Schaffung eines schweizerischen Datenarchivs, welches nicht nur Daten archivieren, sondern vor allem der Forschung zugänglich machen sollte. Allerdings muss sichergestellt werden,

dass die Rechte der AutorInnen und HerausgeberInnen von Datenbeständen angemessen gewahrt sind.

La Base de Données Historiques Genevoises (Frédéric Sardet)

Avertissement: Ce rapport, élaboré à la demande de l'association «Histoire & Informatique», n'engage que son auteur sur toutes les prises de position qui ne sont pas assimilables à l'exposé de faits objectifs (essentiellement le point 5). Il ne saurait donc être retenu comme l'opinion du responsable du projet: le professeur Alfred Perrenoud.

1. Le projet:

1.1. Contenu et objet des données: La Base de Données Historiques Genevoises (B.D.H.G.) a été constituée par une équipe dirigée par le professeur Alfred Perrenoud (Université de Genève, Département d'Histoire Economique et Sociale) à partir du relevé exhaustif des registres de sépultures urbains genevois entre 1685 et 1844, grâce au financement du FNRS.

1.2. But de l'analyse: Cette base de données doit permettre une analyse de la mortalité urbaine genevoise entre 1685 et 1844. Cette étude doit permettre de mieux cerner les conditions du recul de la mortalité entre les XVIIe et XIXe siècles. En outre, les informations géographiques, socio-juridiques et socio-professionnelles doivent permettre une extension de la recherche à d'autres facteurs intervenant dans l'étude de la mortalité: inégalité sociale devant la mort en particulier.

1.3. Taille de la base: La base contient 113 000 actes de décès représentant 255 400 mentions nominatives (chiffres ronds). En outre, elle offre un relevé exhaustif des causes de décès, de l'âge au décès, du statut juridique de l'individu dans la cité (Citoyen, Bourgeois, Habitant, Natif, Etranger, Réfugié protestant), des origines géographiques, des lieux de décès ainsi que des professions. La taille de la base est de l'environ de 25 MgBytes.

1.4. Chronologie: Le projet est «né» en août 1986 (rapport préliminaire de Frédéric Sardet). Il fut mis en chantier au cours de l'année 1988 (F. Sardet, D. Hiler avec l'accord de principe d'Alfred Perrenoud). Une demande de subside fut déposée en 1989. Elle donna lieu au financement par le FNRS d'une année pilote en 1989–1990. Une demande pour poursuivre le projet fut refusée pour l'année 1990–1991.

Il convient de rappeler que ce relevé devait alors servir à la construction d'un fichier de population (sépultures, baptèmes, mariages) qui, après jumelage, reconstitution automatique des familles et constitution de généalogies, pourrait permettre l'ouverture de nouveaux chantiers en histoire écono-

mique et sociale d'Ancien Régime. Ces nouveaux chantiers seraient à nos yeux toutes les recherches dites de «micro-histoire» et relevant de la «network analysis» ou étude des réseaux, dans un cadre économique et social. Il est évident qu'un tel projet, réalisable, ne sera pas réalisé dans l'immédiat, faute de crédits (le FNRS a décidé de ne pas financer la constitution de bases de données. Il est difficile de savoir s'il s'agit d'une résolution définitive ou d'une décision touchant uniquement notre projet).

Afin de pouvoir exploiter pour une recherche historique les relevés déjà effectués, une nouvelle demande au FNRS fut déposée et acceptée pour l'année 1991–1992 qui constitue donc la borne finale à l'élaboration de cette base. Entre 1986 et 1992, il n'y aura donc eu effectivement que deux années académiques financées pour l'élaboration de la base.

2. Structure des données et applications:

2.1. Enregistrement des données: La base de données est actuellement disponible en deux versions relationnelles, une version Dbase IV (PC/DOS) et une version Oracle (PC/DOS et Sun). Une édition imprimée de la source avec index assure sa diffusion publique à la communauté des historiens. Une utilisation informatisée est toujours possible soit sur PC, soit sur gros système, en bénéficiant du réseau informatique de l'Université de Genève. Elle requiert des connaissances informatiques assez poussées (langage de manipulation). Des applications pour l'interrogation publique n'ont pu être développées.

3. Changements de personnel:

La base a été intégralement conçue et administrée par Frédéric Sardet. Les rotations de personnel de saisie ou de correction n'ont pas posé de problème dans la mesure où l'administration de la base est restée sous une même responsabilité durant toute la période. Il faut noter que la position temporaire de Frédéric Sardet au sein de l'institution académique pourra poser des problèmes de suivi et de gestion des données.

4. Relations avec l'extérieur du projet:

4.1. Données externes: Toutes les données utilisées dans la base émanent des relevés effectués en son cadre. Seules les données d'analyse, comme les chiffres de population genevois, sont tirées de la littérature. Toutefois, la base est ouverte et pourrait accueillir de nouveaux types de données sans aucun problème.

4.2. Transmission des données: La transmission électronique des données n'est pas prévue.

4.3. Règles de consultation: Des protocoles d'utilisation étaient prévus au début du projet, dans le cadre d'un projet public. Les contraintes financières ont transformé le projet en pur projet de recherche – mis à part l'édition imprimée des données. Il suffit donc de contacter le responsable (Alfred Perrenoud) pour obtenir une autorisation de consultation.

4.4. Copyright: Les modules développés pour saisir, corriger et manipuler les données ont été écrits par Frédéric Sardet. Leur caractère purement «utilitaire» n'a pas justifié l'élaboration de protocoles concernant le copyright. La question d'un copyright sur le produit «fini» et tel qu'il a été conçu est en revanche à discuter, même si le financement public et académique ne doit pas conduire à des applications marchandes du produit.

5. Archives centrales:

La question d'un archivage des bases de données élaborées à des fins scientifiques pose de sérieux problèmes déontologiques, qui nécessitent un certain nombre de garanties, qui touchent avant tout:

- la transparence du produit: sources utilisées (cotes archives), modélisation des sources (exhaustivité/échantillon), conditionnement des informations (données brutes/classification préalable)
- la reconnaissance du travail fourni d'un point de vue informatique, dès lors qu'un autre veut l'exploiter
- la priorité de l'exploitation scientifique d'un produit par celui qui l'a construit.

Le besoin essentiel pour les historiens face aux bases de données concerne leur maintenance, souvent précaire du fait de la rotation du personnel et de l'inégale maîtrise de l'informatique par les historiens (cf. enquête auprès des membres d'«Histoire & Informatique»). Une «archive centrale» devrait donc être un lieu qui non seulement stocke l'information, mais dispose d'un personnel compétent, capable de faire vivre les bases, i.e.:

- créer des produits consultables
- générer des sous-produits en fonction des demandes
- savoir distribuer électroniquement les données
- assurer l'exportation et l'importation d'informations entre divers systèmes et entre divers formats.

Un tel projet entre en concurrence avec les services statistiques fédéraux et le SIDOS. Les universitaires et les archives, pour des raisons plus politiques que scientifiques, peuvent voir dans la naissance d'un tel service une

concurrence directe à leur production. Les compétences nécessaires à un tel service se transformeront facilement et rapidement sans doute, en lieu d'analyse et de recherche. Il s'agit donc d'être prudent et précis sur le rôle dévolu à un tel service.

Das Projekt FSWbase (Lukas Vogel)

1. Daten und Projekt:

Im Lauf der Jahre wurden an der Forschungsstelle für Schweizerische Sozial- und Wirtschaftsgeschichte der Universität Zürich im Rahmen verschiedener Forschungsarbeiten (Seminarien, Lizentiate, Dissertationen, Nationalfonds-Projekte der Professoren Fritzsche und Siegenthaler) zahlreiche Datensätze erhoben. Ein spezielles Projekt «FSWbase» soll alle an der Forschungsstelle elektronisch vorliegenden Daten standardisieren und in abrufbare Form bringen.

1.1. Inhalt der Daten: Die nachstehende Liste dient einer groben Orientierung und ist nicht vollständig. Erhoben wurden vor allen Daten der eidgenössischen Volkszählung (Zeitraum 1850–1910). Vereinzelt liegen auch Datensätze für spätere Zeiträume vor (1930, 1960, 1980). Untersuchungseinheit sind alle Bezirke der Schweiz oder ausgewählte Gemeinden (Sample zwischen 10 [Städte] und 415 Gemeinden [Zentren]). Die erhobenen Variablen decken hauptsächlich folgende Gebiete ab:

- demographische Strukturdaten (Wohnbevölkerung, Geschlecht, Geburts-, Arbeitsort, Konfession usw.)
- Daten zur Wirtschaftsstruktur (Beschäftigte in einzelnen Wirtschaftszweigen und Berufen)
- Viehzählungen
- Daten zur Verkehrsstruktur: schweizerische Fahrpläne zwischen 1870 und 1910 (Verkehrsträger: Schnell-, Personen-, Güterzüge, Postkutschen, Dampfschiffe), Postpassagiere, Warenströme in Tonnen usw.
- Daten zur Kommunikationsstruktur (Telefonabonnenten, Anzahl geführte Gespräche, Aufgabe und Empfang von abonnierten Zeitungen, Kartenschlüsse usw.)
- Neben diesen jeweils zu mehreren Jahren erhobenen Querschnittdaten liegen mehrere tausend Zeitreihen für den Zeitraum zwischen 1850 und 1990 vor. Es ist geplant, diese Längsschnittdaten in Kürze in Buchform zu veröffentlichen.

1.2. Zweck im Rahmen der weiteren Projektarbeit (Auswertungsziele): Die Daten dienen der Beschreibung von strukturellen Entwicklungen in der Schweiz. Für bestimmte Daten (z.B. Fahrplandateien) wurden umfangreiche Programmsets zur Analyse entwickelt.

1.3. Umfang der Daten: Die Grösse der Primärdateien liegt zwischen ca. 20 und 30 Megabyte. Die Grösse der Sekundärdateien (Resultate aus Analysen) dürfte sich zwischen 40 und 60 Megabyte bewegen. Diese Daten werden allerdings nach Abschluss der jeweiligen Projekte grösstenteils gelöscht. Einzelne Dateien (Streckendateien der Eisenbahnen) erreichen eine Grösse von bis zu 17 Megabyte.

1.4. Chronologie: Mit dem Aufbau einer speziellen Datenbank wurde Anfang 1992 mit sehr kleinen Ressourcen begonnen. Sofern das Projekt mittelfristig nicht finanziert werden kann, wird es Ende 1993 abgeschlossen.

2. Datenstruktur und Datenbankverwaltung:

2.1. Datenformate: Die Daten liegen als ASCII- oder Excel-Tabellen vor. Ein Grossteil der Volkszählungsdaten ist ausführlich dokumentiert und liegt als Codebook vor.

2.2. Anwendung zur Datenverwaltung: Diese Ausführungen beziehen sich auf das geplante Projekt FSWbase. Der Aufbau der Datenbank ist in drei Stufen vorgesehen, in denen eine sukzessive Ausweitung der Funktionalität erfolgt:

1. Stufe:
- Die Daten werden je nach Struktur (Längs-, Querschnitt, Erhebungsjahr, Untersuchungseinheit) in einzelnen Tabellen abgelegt
- Abruf einzelner Tabellen nach Untersuchungseinheit oder Jahrgang
- Durchsicht der ausgewählten Daten
- Anzeige des Kommentars
- Erstellen deskriptiver Statistiken
- Sortieren nach wählbaren Variablen
- Export der Tabellen als ASCII- oder Excel-Dateien
- Ausdruck gewählter Tabellen mit Kommentar als Standard Reports.

2. Stufe:
Abruf mehrerer Tabellen nach Themen und Variablen möglich. Eingabe gewünschter Relationen (automatisches Aufzeigen möglicher Relationen).

3. Stufe:
- Hilfsmittel zur Auswertung der Tabellen
- Bereitstellen von Standardanalysen (Entwicklung, Wachstumsraten usw.)
- Möglichkeit der graphischen (Linien, Balkendiagramme usw.) und kartographischen Darstellung von Datenreihen.

2.3 Eigenentwicklungen: Die Datenbank FSWbase wird als Oracle-Datenbank vorliegen und mit SQL verwaltet werden. Als Benutzerschnittstelle

dient Hypercard/Oracle Card. Die Datenbank wird also einerseits mit SQL bzw. Hyper*SQL entwickelt, die Benutzerschnittstelle andererseits mit HyperTalk und OracleTalk. Im Endausbau wird FSWbase auf einem zentralen Datenbank-Server (Betriebssystem A/UX) zur Verfügung stehen und via Oracle Card von Macintosh oder PCs (Windows) abgefragt werden können.

3. Wechselfälle von Manpower, Hardware und Software:

3.1. Verantwortung für Datenbankaufbau, -pflege und -auswertung: Das Projekt wird momentan als Nebenprodukt des Nationalfonds-Projekts Verkehrs- und Siedlungsentwicklung in der Schweiz sowie als Teil einer ⅓-Assistenz weitergeführt. Die verantwortliche Person wird den Datenbankaufbau in diesem Rahmen bis Ende 1993/Anfang 1994 weiterführen. Eine Weiterführung des Projekts wird mit institutsinternen Stellenprozenten nicht möglich sein.

3.2. Wechsel von verantwortlichen Personen und Hardware-Plattformen: Bislang keine.

3.3. Wechsel bei Entwicklungsinstrumenten: Als Entwicklungsinstrument wurde von Anfang an eine SQL-Datenbank favorisiert. An der Wahl von Oracle/Oracle Card wird sich vorläufig nichts ändern, da die Portierung von Oracle auf fast beliebige Hardware-Plattformen gewährleistet ist und eine Portierung von Oracle Card zwischen Finder (Mac) und Windows (MS-DOS) problemlos möglich ist. Von einer direkten Entwicklung der Applikation mit Oracle für A/UX wurde aufgrund der Komplexität von A/UX sowie der miserablen Dokumentation von Oracle Abstand genommen.

4. Aussenbeziehungen:

4.1. Datenübernahme: Die vorliegenden Daten wurden mit verschiedenster Hard- und Software erfasst. Sie müssen grundsätzlich als kommentierte ASCII-Datei vorliegen. Da der Erhebungszweck bzw. die Analyse der Daten sich von Projekt zu Projekt massiv verändern kann, sind die Tabellen punkto Struktur unterschiedlich angelegt. Muster für die Erstellung von Tabellen existieren zwar, die Durchsetzung eines Standards (Normierung der ID-Nummern und Tabellenstrukur) war bis anhin aber nur teilweise möglich. Diese Tabellen werden mit dem Importmodul der Oracle-Datenbank eingelesen und je nachdem modifiziert.

4.2. Weitergabe von Daten an auswärtige Benützer: Eine Regelung für auswärtige Benützer existiert momentan nicht. Langfristig ist aber die Möglichkeit einer Abfrage von FSWbase via Terminal vorgesehen. Es werden nur

Daten aus abgeschlossenen Projekten und mit Einverständnis der Autoren freigegeben.

4.3. Copyright: Copyright-Regelungen bestehen nicht. Der Bereich scheint kaum problematisch zu sein (siehe auch 4.2.).

5. Bemerkungen:

Die Realisierung einer zentralen Datenbank wird an der Forschungsstelle als problematisch betrachtet. Dies aufgrund der Schwierigkeiten, die sich durch die Standardisierung der unterschiedlich strukturierten Tabellen ergeben würden. Als sinnvoll wird aber die Einrichtung einer zentralen «Meta-Datenbank» erachtet, in der Kataloge der vorhandenen Daten, Aufbewahrungsort, Format, Ausleihformalitäten usw. verzeichnet sind. Ebenso wünschenswert wäre eine zentrale Stelle zur Langzeitarchivierung elektronisch erfasster Quellen.