

Probleme des Aufbaus und Umfangs einer Datenbank 'Historische Statistik'

Sieglerschmidt, Jörn

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Sieglerschmidt, J. (1988). Probleme des Aufbaus und Umfangs einer Datenbank 'Historische Statistik'. *Historical Social Research*, 13(1), 89-110. <https://doi.org/10.12759/hsr.13.1988.1.89-110>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:

<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see:

<https://creativecommons.org/licenses/by/4.0>

Probleme des Aufbaus und Umfangs einer Datenbank 'Historische Statistik'

*Jörn Sieglerschmidt **

Abstract: A data base for historical statistics, although in many respects and for many users useful and desirable, doesn't exist in Germany. For this fact account the many difficulties in modelling such a data base: from defining a computable historical fact, from problems of unification of data to simple questions of the formal representation of data. On the following pages a program for a possible data base design is developed.

Eine Datenbank 'Historische Statistik' will beanspruchen, möglichst erschöpfend historische Daten zu erfassen und zu deren Verknüpfung den Benutzer möglichst leicht zu handhabende Prozeduren zur Verfügung zu stellen. Das nächstliegende Problem ist die Bestimmung der Zielsetzung: ob es sich nämlich um eine quellennahe oder um eine Datenbank handeln soll, die auch und gerade für quellenfernere Daten zur Verfügung stehen soll. Im ersten Fall könnte ein pragmatischer Weg der Datennormierung bereits zum gewünschten Erfolg führen. Quellen werden als Dateien archiviert und durch einen ausführlichen - sicher als Datenbank konzipierten - Katalog erschlossen. Im anderen Fall - und dieser soll hier weiterverfolgt werden - wird der pragmatische Tugendpfad verlassen. Angestrebt wird eine umfassende Datenbank, die auf unterschiedlichen Aggregationsebenen sowohl quellennahe als auch quellenferne Daten enthalten soll. Quellennähe meint in einer vornehmlich auf quantifizierende Verfahren ausgerichteten Datenbank nicht den Versuch, uneingeschränkt quellentreu die Daten abzubilden. Gleichwohl sollen interpretative Schritte entweder im Rahmen der Datenbank kontrollierbar oder möglichst gering gehalten werden.

Da es sich um quantitative Daten handeln wird, ist erst zu klären, was unter quantitativen Daten verstanden werden soll. Die nicht enden kön-

* Address all Communications to: Jörn Sieglerschmidt, Landesmuseum für Technik und Arbeit, Am Ullrichsberg 16, D-6800 Mannheim 31.

nende Auseinandersetzung um quantitative Methoden ist hier freilich nicht aufzunehmen. Quantifizieren heißt - mit welcher aufwendigen statistischen Mitteln auch immer durchgeführt - abzählen. Ein solcher Vorgang findet allenthalben statt, auch wenn in der historischen Darstellung die genaue Zahl durch ein abschätzendes Mengenattribut ersetzt wird.

Gerade die Abzählbarkeit galt noch um 1900 als etwas, das gerade nicht das Wesen historischer Arbeit ausmacht. So unterschied z.B. Xénopol zwischen 'faits de répétition' und 'faits de succession'. Erstere sind zeitunabhängig und Gegenstand - ganz im Sinne von Windelband - nomologischer bzw. nomothetischer Wissenschaft. Letztere sind der eigentliche Gegenstand historischer Forschung: die den historischen Ablauf verändernde Tatsache, die über den dumpf sich wiederholenden Alltag sich erhebt, die große Tat. Xénopol lehnt daher auch die Suche Lamprechts nach historischen Gesetzen ab (Xénopol, Les faits de répétition). Solche Meinungen sind auch heute nicht außerhalb der Debatte, wenn z.B. nur solche historischen Aussagen als wertvoll bezeichnet werden, die sich gerade nicht auf quantifizierbare Daten stützen. Unbestritten wird aber heute zumeist Abzählbares als historisch bedeutsam akzeptiert, z.B. die Naturgeschichte durchaus nicht als zeitlos begriffen, sondern als wichtiger Teil der historischen Entwicklung. Ob ein Ereignis daher als individuell, als einzigartig oder als kollektiv, als sich wiederholend und in dieser Form abzählbar gilt, hängt vom Willen des Betrachters ab. Alles kann quantifiziert werden, folgt man dem Diktum Schumpeters: If you can quantify, quantify it, if you cannot quantify, quantify it just the same.

Eine Datenbank 'Historische Statistik' sollte Daten enthalten, die entweder durch die Quelle selbst oder den Bearbeiter in wohl definierte, und zwar möglichst kleine, abzählbare Einzelteile zerlegt worden ist. Darin besteht die Analyse des Quellentextes - bzw. der analytische Zustand des Textes - , die der (statistischen) Synthese vorauszugehen hat. Ausgeschlossen werden damit hermeneutische Verfahren, die das Gesamtverständnis eines Textes erstreben, also nicht einem methodologischen Atomismus, sondern Holismus verpflichtet sind. Auf die quantifizierende Inhaltsanalyse, die sicherlich auf der Grenze beider methodischen Verfahren liegt, kann hier nicht näher eingegangen werden. Ebenso müssen wichtige Bereiche der historischen Überlieferung wie z.B. die Behandlung von Sachobjekten (Archäologie, Kunst- und Technikgeschichte, Volkskunde) außerhalb der Betrachtung bleiben. Deutlich wird aus der eben getroffenen Unterscheidung, daß der Graben zwischen quantifizierenden, personen- sowie naturgeschichtliche Ereignisse umfassenden und hermeneutischen Verfahren unüberbrückbar bleibt.

Abzählbar sind eben nur auf kleinste mögliche Einheiten zurückgeführte Ereignisse. Die synthetischen Schritte - mehr oder weniger durch statistische Verfahren in ihrem Aussagewert abgesichert - beruhen auf Zurech-

nung (imputatio, Assoziationen sind beabsichtigt) von Handlungsmotiven, die dem verarbeiteten Material nicht entstammen. Der Zerlegung des Materials, das dadurch zunächst jeglichen Kontextes beraubt wird, folgt der Versuch, solchen Kontext nachträglich zu schaffen. Daß auf methodologischer und darstellerischer Ebene daher die Vermittlung quantifizierender und hermeneutischer Verfahren das Hauptziel bleiben muß, ist offensichtlich.

Eine Datenbank 'Historische Statistik' jedoch soll dazu dienen, analytisches Material sowie eine gewisse Anzahl synthetisierender Werkzeuge zur Verfügung zu stellen. Kennzeichen der dort zu findenden Daten ist die Abzählbarkeit, deren Voraussetzung Vergleichbarkeit ist. Daher kann Kennzeichen einer Datenbank 'Historische Statistik' allenfalls ihre Zuverlässigkeit, d.h. die Tatsache bzw. die Absicht gleiche Dinge gleich zu behandeln, nicht aber ihre Gültigkeit sein, d.h. die Absicht, auch individuelle Eigenarten abbilden zu wollen.

Die Frage, welcher Datenbanktyp gewählt werden soll, ist angesichts der in den letzten Jahren geführten Diskussionen weitgehend klar: eine relationale Datenbank hat gegenüber den konventionellen Datenbanktypen - hierarchische, netzwerkartige - nicht Vorteile, sondern umfaßt diese. Eine relational aufgebaute Datenbank läßt sich in einen der konventionellen Datenbanktypen verwandeln. Ohne auf Einzelheiten der Datenbankkonstruktion eingehen zu können und zu wollen, sind doch einige definitorische Klärungen vonnöten.

Relationale Datenbanken haben ihren Namen vom mathematischen Begriff 'Relation' geerbt (CODD 1970). Eine Relation - vorstellbar als Matrix oder Tabelle (flat file) - besteht aus Spalten (Attributen) und Zeilen (Tupeln). Der Wertebereich eines Attributs (= Gesamtheit zulässiger Werte einer Spalte) wird als Domäne bezeichnet. Es dürfen niemals zwei identische Tupel in derselben Relation vorkommen. Jede Relation besitzt eine minimale Kombination von Attributen, die als Schlüssel fungiert, d.h.: Sind die Werte des Schlüssels bekannt, so sind alle Nicht-Schlüsselwerte eindeutig bestimmt. Als Beispiel hierfür nehme man ein Postleitzahlenverzeichnis (Relation) mit den Attributen Postleitzahl und Ortsnamen. Dann spielt das Attribut Postleitzahl die Rolle des Schlüssels, da jeder Ortsname aus der Kenntnis der Postleitzahl eindeutig rekonstruierbar ist. Umgekehrt kann nicht geschlossen werden, da gleiche Ortsnamen auf unterschiedliche Postleitzahlen führen können, was nachfolgender Auszug aus einem DDR-Postleitzahlenverzeichnis verdeutlichen soll.

<i>Postleitzahl</i>	<i>Ortsname</i>
2201	Kemnitz
8701	Kemnitz
1711	Kemnitz
9901	Kemnitz
1501	Kemnitz
1921	Kemnitz
...

Obwohl die eben erwähnte Begriffsbildung (Relation, Tupel, Attribut, Schlüssel, Domäne) bei allen relationalen Datenbanksystemen einheitlich ist, gibt es mannigfache Unterschiede des jeweils verwendeten Modells. Besonders interessant sind dabei die Ansätze von Chen (CHEN 1976) und Codd (CODD 1979). Beide versuchen auf recht unterschiedliche Art, die sehr schwache semantische Konstruktionshilfe (Ausschluß identischer Tupel) des klassischen relationalen Modells zu erweitern. Während Chen das Schwergewicht seiner Betrachtungen auf den Unterschied zwischen Objektrelationen (entities) und aus Assoziationsrelationen (relationships) legt, versucht Codd das Problem unbekannter Informationen (null values) zu systematisieren. Besonders letzterer Ansatz könnte in einer Datenbank 'Historische Statistik' von großem Nutzen sein. Doch ungeachtet dessen, welches relationale Modell in einem Datenbanksystem verwirklicht ist, kann in allen Fällen gesagt werden, daß derartige Systeme eine uneingeschränkte Auskunftsbereitschaft garantieren, da sämtliche Relationen statisch wie dynamisch, sowohl permanent als auch temporär, physisch als auch logisch aufgebaut, verändert und gelöscht werden können.

In entsprechend konzipierten konventionellen Datenbanken werden die einzelnen Daten über festgelegte Pfade, über hierarchische oder Netzwerkstrukturen verknüpft, was sich bei neuen oder anders gerichteten Fragestellungen als hinderlich erweist. In einem relationalen System können solche Verknüpfungen auch nachträglich definiert werden, weswegen derartige Systeme auch als besonders flexibel gelten. Diese Flexibilität ist völlig unproblematisch, da Beziehungen zwischen Relationen lediglich durch entsprechende Beziehungsattribute zu gewährleisten sind.

Auszugehen ist bei der Konzeption einer Datenbank 'Historische Statistik' von den historischen Quellen, die von einzelnen Bearbeitern erfaßt und in eine mehr oder weniger an Rechner angepaßte Form umgesetzt worden sind. Es ist gerade bei starker Quellenbezogenheit zu betonen, daß die Datenbank 'Historische Statistik' keine Archivfunktion im Sinne des Konservierens hat. Das könnte eine gewünschte Nebenfolge sein für einen Teil der Daten. Zuallererst jedoch soll die Datenbank dem Informationsumschlag dienen. Dafür ist es notwendig, die anderswo erfaßten Da-

ten für die Zwecke einer nach einheitlichen Konventionen arbeitenden Datenbank zu verändern. Hierin wird eine der Hauptaufgaben der Datenbankverwalter bestehen. Darüberhinaus werden diese den gesamten internen Aufbau der Datenbank entwerfen einschließlich derjenigen Teile, die benutzernah sind wie z.B. Datenmasken zur Dateneingabe. Schließlich werden sich die Datenbankverwalter mit der Entwicklung und Beschreibung von Standards beschäftigen müssen. Die folgenden Überlegungen sollen für Teilbereiche erste Ansätze dazu bieten.

Für jede relationale Datenbank ist die entscheidende Frage, welches Attribut bzw. welche Attribute in einer Relation die Schlüsselattribute sein sollen. Die Bestimmung der (des) Schlüsselattribute(s) entscheidet u.a. darüber, wieviel Relationen gebildet werden, um gegebene Daten in einer Datenbank abbilden zu können. Sie gewährleistet daher, daß die Tupel in einer Relation eindeutig voneinander unterschieden werden können. Je umfangreicher und vielfältiger die Daten sind, desto schwieriger wird es, solche Schlüsselattribute zu bestimmen. Es handelt sich dabei um Probleme der sog. Normalisierung, d.h. der Zerlegung der Informationen, um zu einer kanonischen Datenstruktur zu gelangen (VETTER 1982, S. 93 ff.). Es ist hier allerdings nicht an eine Zerlegung der universellen Relation 'Historische Statistik' gedacht. Nebenbei bemerkt sei an dieser Stelle aber, daß in einem Datenbanksystem der oben beschriebene, auf methodologische Ebene angesiedelte analytische Prozeß methodisch gefordert wird, um Datenanomalien zu vermeiden.

Die ursprüngliche Idee, über Personen (Individuen) die Eindeutigkeit der Objekte (Tupel) zu gewährleisten, läßt sich bei der Aufnahme auch quellennaher Informationen nicht verwirklichen. So ist z.B. die Aufnahme von Kirchenbuchdaten nur dann möglich, wenn jede Person eindeutig identifiziert ist. Dafür bedarf es allerdings eines Bearbeitungsvorganges, der eine sog. Metaquelle schafft. Selbst dann sind nicht alle Personen eindeutig identifiziert. So bleibt als pragmatische Lösung lediglich die laufende Numerierung der Einträge, ohne daß z.B. ein Täufling als mit einem späteren Bräutigam identische Person dieselbe Personennummer erhält. Die Betrachtung dieses Beispiels zeigt, daß es für den quellennahen Teil der Datenbank nicht nur pragmatisch, sondern auch sinnvoll ist, vom vorgegebenen Datenaufbau auszugehen. Der Datenbankverwalter hat lediglich für eine datenbankgerechte Zurichtung der Daten zu sorgen.

An eine Datenbank 'Historische Statistik' werden zwei Anforderungen gestellt: Sie soll zum einen quellennahe Daten enthalten, zum anderen Daten darüber sich erhebender Aggregationsstufen. Diese Stufen der Aggregation werden im Rahmen der Datenmodellierung durch Bildung von geographischen Räumen immer größerer Ausdehnung gebildet. Vom Individuum als Raumpunkt über das Haus, das Dorf oder die Straße, das Stadtviertel, das (Ober)amt bzw. die Stadt bis zu noch größeren räumli-

chen Einheiten werden Individuen zusammengefaßt. Auf den höheren Aggregationsebenen sind also Daten zu finden, die sich auf Lebensverhältnisse, d.h. nicht auf Individuen beziehen: Preise, Löhne, Flächennutzung, Bodenbeschaffenheit, Klima u.a.m. Weitere Formen der Aggregation z.B. durch die Vereinigung zahlreicher unterschiedlicher Quellen (sachliche Aggregation) oder die Bildung von Epochen (zeitliche Aggregation) sind vom Benutzer zu definieren über die verfügbaren Datenmanipulationssprachen. Auf diese Weise könnten im übrigen auch Räume in anderer als der durch die Datenbank vorgegebenen Weise aggregiert werden.

Der dreidimensionale Raum sowie als vierte die zeitliche Dimension werden Attribute fast aller Relationen einer Datenbank 'Historische Statistik' sein, wenn sie auch nicht in gleicher Häufigkeit zu den Schlüsselattributen zählen werden. Die sachlichen Dimensionen machen dann die Datenbank einer multidimensionalen Matrix gleich, die als universelle Relation (VINEK/RENNERT/TJOA 1982, S. 68ff.) zu bezeichnen ist. Dieser multidimensionale Raum muß in einfachen Relationen abgebildet, die universelle Relation 'Historische Statistik' in Teilmengen projiziert werden, ohne einerseits zu viel Speicherplatz zu verschenken, ohne andererseits zuviel Redundanz zu erzeugen. Je mehr Relationen gebildet werden, desto häufiger müssen verknüpfende Attribute oder dieselben Attribute in mehrere Relationen aufgenommen werden. Allerdings erscheint das noch sinnvoller als die Bildung von Relationen mit einer Fülle von Attributen, die häufig einen Leerwert (null value) annehmen müssen. Das wäre in einer Datenbank für historisches Material laufend der Fall. Der Historiker ist an die Lückenhaftigkeit seiner Daten gewöhnt, während für Datenbankkonstrukteure darin erhebliche Probleme liegen (DATE 1986, 2, S. 210ff.; VINEK/RENNERT/TJOA 1982, S. 209 - 233). Grundsätzlich sollte daher darauf gesehen werden, daß Relationen mit möglichst wenigen Attributen gebildet werden. Dabei kann eine Ausweitung der Domänen nicht nur in Kauf genommen werden, sondern sogar erwünscht sein. Mit solchen Maßgaben sind wesentliche Schritte in Richtung auf die zweite Normalform getan (VINEK/RENNERT/TJOA 1982, S. 81f.).

Bevor einzelne sachliche Bereiche näher betrachtet werden, ist noch etwas zur Vereinheitlichung der Schreibweisen, der Maße, Gewichte und Währungen zu sagen. Seit der Zeit um 1800 beginnt eine zunehmende Ordnungswut: ganze Lebensbereiche werden zum Zwecke der Vereinheitlichung und Vergleichung erst einmal gleich zugerichtet, in Reihe und Glied gebracht. Eine Datenbank 'Historische Statistik' muß für die Zeit davor ebenso verfahren, auch wenn gerade in diesem Bereich zahlreiche methodische Fallen sich auftun.

Werden längerfristige Serien bearbeitet, so hat sich bei den Währungen die Umrechnung auf Gold- und Silberäquivalente durchgesetzt, obwohl die Probleme des wechselnden Wertes der Edelmetalle wie des ebenfalls

wechselnden Verhältnisses von Gold- und Silberwert nicht als gelöst gelten können. Ebenso problematisch bleibt die Währungsumrechnung für das 20. Jahrhundert dort, wo man sich geldpolitisch vom Goldstandard gelöst hat. Grundsätzlich schwierig wird allenthalben die Frage zu lösen sein, wie die Kaufkraft des jeweiligen Geldes sich räumlich und zeitlich ändert. Erst Lebenshaltungskostenindices vermögen auf der Ebene des einzelnen Haushaltes eine gewisse überzeitliche und überregionale Vergleichbarkeit zu gewährleisten.

Was Maße und Gewichte angeht, so gibt es nach 1800 - zumindest für den Bereich des Deutschen Reiches - kaum noch ernsthafte Probleme zu bewältigen, da zu Beginn des 19. Jahrhunderts in vielen Territorien neue Maßsysteme eingeführt werden, die der örtlichen Vielfalt der Maße ein Ende setzen. Allerdings gibt es eine lange Übergangszeit - fast bis 1871 - , in der frühere Hohlmaße noch nicht auf Gewichtsmasse umgestellt worden sind. Wie vor 1800 ist in diesen Fällen über die Litergewichte eine methodisch problematische Umrechnung auf heutige Gewichtseinheiten notwendig. Nicht in der Schärfe, aber nicht minder schwierig ist Sicherheit bei den Flüssig-, Längen- und Flächenmaßen zu erhalten. Methodische Risiken dieser Art könnten ohne viel Aufhebens in Kauf genommen werden, wenn einmal berechnete, durch die Mühen methodischer Skrupel gegangene Werte nicht die unangenehme Eigenschaft hätten, sich solchen Skrupeln gegenüber zu verselbständigen, ohne einen Bezug zu den Mühen der Berechnung zu hinterlassen. Wer kennt nicht den zur weitausholenden Interpretation führenden Zauber einmal berechneter, in Tabellenform gebrachter Daten? Das mehrfach gebrochene Abbild, der Ausschnitt von Wirklichkeit wird die Wirklichkeit. Obwohl methodische Risiken dieser Art besonders für eine langfristig angelegte Datenbank unübersehbar sind, bleibt die Benutzbarkeit der Datenbank abhängig von der Vereinheitlichung der Maße und Gewichte.

Das Problem wird etwas entschärft, wenn für die Datenerfassung die Aufnahme der Originalmaße vorgesehen wird. Die Umrechnung würde über eigens angelegte Umrechnungsrelationen erfolgen. Das böte den Vorteil, bei neuen Erkenntnissen über den Umrechnungsfaktor lediglich die entsprechenden Werte in der Relation ändern zu müssen, nicht aber den gesamten Satz der quellennahen Daten. Im folgenden wird jedoch von bereits umgerechneten Werten ausgegangen.

Ebenfalls schwierig zu entscheiden ist, ob Eigennamen und Benennungen z.B. für Berufe in der Schreibweise vereinheitlicht werden. Für eine Datenbank zur historischen Statistik mit starken Anteilen nichtnumerischer, quellennaher Daten scheint nun die Aufnahme sämtlicher Informationen in der originalen Schreibweise selbstverständlich. Soweit zuverlässige Algorithmen zur Vereinheitlichung verfügbar sind, wäre gegen ein solches Verfahren nichts einzuwenden. Es ist im übrigen zu bedenken, daß

sich die Zahl der Untersuchungen, die sich der Schreibweise annehmen, außerordentlich klein ist. Daher wird für die Zwecke der hier zu entwerfenden Datenbank von einer durchgängigen Vereinheitlichung ausgegangen: es handelt sich schließlich um historische Statistik. Konkurrierende Nutzenwendungen - z.B. historische Namensforschung - müssen hinter diesem Ziel zurückstehen. Denkbar wäre aber auch, daß über Quellenbanken oder auch im Rahmen der hier vorgeschlagenen Datenbank über Wörterbücher und andere Hilfsmittel solche Informationen wenigstens zugänglich bleiben.

Nun könnte die Vereinheitlichung derart vorangetrieben werden, daß über unterschiedliche Wörterbücher (dictionaries) und eine entsprechende Schreibmaske die Eingabe von Daten so gesteuert wird, daß in die Datenbank nur eindeutige numerische Werte eingetragen werden. Eine solche Vorgehensweise ist in der Vergangenheit vielfach gewählt worden, um Platz zu sparen, weil ein achtzigspaltiges Datenformat dazu zwang. Unerheblich ist hier, ob es sich um mnemotechnische alphanumerische oder numerische Codes gehandelt hat. Heute sind solche Beschränkungen in vieler Hinsicht unnötig. Wichtig sind lediglich zwei Kriterien: Erfassung und Umarbeitung der Daten. Für eine Datenbank ist zusätzlich der präsumtive Benutzer entscheidend: die Lesbarkeit der Information muß gewährleistet sein, ohne daß dickleibige Benutzerhandbücher gewälzt werden. Vereinheitlichung der oben beschriebenen Art kommt also nur in begrenztem Maße in Frage dort, wo die Zahl der möglichen Fälle, wo die Domäne außerordentlich klein ist. Angaben bei Personen wie z.B. Geschlecht, Zivilstand, Verwandtschaftsverhältnis, Religion sollten in (alphanumerischen Codes erfaßt werden. Auch wenn der Schreibaufwand größer ist, bleiben die übrigen Angaben uncodiert, werden aber vereinheitlicht: so sollten Ortsnamen grundsätzlich in der heute üblichen Schreibweise aufgenommen werden, Berufsbezeichnungen ebenso. Für Personennamen läßt sich das nicht in gleicher Weise zur Norm erheben. In der französisch-kanadischen Forschung sind offensichtlich außerordentlich leistungsfähige Algorithmen zur Identifizierung unterschiedlich geschriebener Namen entwickelt worden. Ähnliches ist für andere Sprachen vermutlich auch möglich. Die Übernahme der Quellenschreibweise wäre hier aus mancherlei Gründen bei der Datenerfassung vorzuziehen, auch wenn der immer wieder beschworene Namensforscher ein weiteres Mal ausbleibt. Vornamen dagegen können vereinheitlicht werden, da die meisten Namen auch heute noch üblich sind.

Im folgenden werde ich nun kein vollständiges Datenmodell für eine Datenbank 'Historische Statistik' entwerfen, sondern lediglich an einigen Beispielen die wesentlichen Verfahren zu erläutern versuchen. Das liegt auch insofern nahe, als der Aufbau einer solchen Datenbank sich an dem angebotenen Material auszurichten hat. Jeglicher zuvorkommender Über-

eifer wäre daher unangebracht. In Erinnerung zu rufen ist die Feststellung, daß historische Statistik analytisch verfährt: die aufgenommene Quelle wird also in den meisten Fällen für die Zwecke der Datenbank verändert, Informationen, die in der Quelle kompakt zusammengefaßt sind, werden auseinandergerissen und - wo notwendig - auf unterschiedliche Relationen verteilt. Um als Beispiel Urbare zu nennen: während für das Quellenabschreiben wie für die Auswertung sich die Aufnahme der Gesamtinformation in einem Satz anbietet (Besitzer, Grundstück, Größe des Grundstücks, Steuerwert usw.), ist für eine Datenbank ein solches Verfahren zu platzaufwendig. Es ist sinnvoll, die Information zu teilen: in die Relationen PERSON, PARZELLE und ABGABE. Denkbar wären auf der Grundlage der Urbarinformationen weitere Relationen wie z.B. BERUF und andere (SCHRENK 1987). Es ist außerdem wichtig nochmals zu betonen, daß eine Datenbank der vorgeschlagenen Art keine Quellentreue gewährleistet, daher auch in keinem Fall die Funktion einer Quellenbank übernehmen kann, wenn das zuweilen wohl auch eine beabsichtigte Nebenfolge sein wird. Gleichwohl sind für mehr editorisch ausgerichtete Datenbanken andere, auch andere institutionelle Lösungen erforderlich.

Eindeutig zu identifizieren sind zunächst die Orte. Als Orte sollen hier sämtliche möglichen Ortsbezeichnungen verstanden werden, ob nun Einzelpunkte oder mehr oder weniger große Flächen damit bezeichnet werden. Jeder Ort muß über seine Lagebezeichnung eindeutig lokalisierbar sein. Geographen haben das notwendige datentechnische Rüstzeug, um solche Probleme lösen zu können. Dazu gehört auch die Frage, wie die Information zur geographischen Lage beschaffen sein muß, um kartographisch umsetzbar zu sein. Das gilt besonders für Flächen, also Gemeinde-, Amts- und Territorialgrenzen. Schließlich sollte neben dem (rechtlichen) Status des Ortes auch seine Zugehörigkeit zu größeren Gebietseinheiten vermerkt werden. Es versteht sich fast von selbst für einen Historiker, daß der Fundort der Information genannt wird. Auf die Quellen und ihre eindeutige Bezeichnung ist gleich noch einzugehen. Hier genügt es die Quellennummer zu kennen. Es ist im übrigen zu überlegen, ob die Ortsnamen, die eindeutig mit einer Ortsnummer gekennzeichnet sind, in alphanumerischer Form in einem Wörterbuch katalogisiert werden, das darüberhinaus auch Schreibvarianten enthalten könnte. Die Flächengröße wird in Quadratmetern angegeben.

ORT

<u>PN</u>	QN	NAME	FLAECHE	BREITE	LÄNGE	ZUGEH
01	01	Rom	3765431	N43.07	15.07	Ital
02	01	Kiev	2345678	N50.45	30.29	USSR
.

Getrennt von dieser Ortsliste, die eine eindeutige Zuordnung von Ortsnummern und -namen erlaubt, sollte eine Ortsrelation mit historischer Dimension gebildet werden. Über die Attribute ZEIT und DAT werden die historisch belegten Fälle des Vorkommens zeitlich zugeordnet. Hier wird davon ausgegangen, daß ein Ort immer einem bestimmten Datum zugewiesen werden kann, daß also ein Ort in einer Quelle nicht über sehr lange Zeiträume laufend erwähnt wird. In diesem Fall müßte der Ort entweder entsprechend der Anzahl der den Zeitraum bildenden Jahre als Tupel in der Relation auftauchen, oder es müßte eine gesonderte Relation vorgesehen werden.

TOPOS

<u>QN</u>	<u>PN</u>	DAT	STATUS	ZUGEH
02	03	1786.00.00	Dorf	Neuenbürg
03	03	1912.01.01	Stadt	Konstanz
04	03	1923.07.13	Landkreis	Baden
.

Jede in die Datenbank eingearbeitete Information bedarf, da sie in den meisten Fällen aus ihrem Quellenzusammenhang gerissen ist, einer genauen Identifizierung, um die Zusammengehörigkeit der Daten rekonstruieren zu können. Die Relation QUELLE wird nun lediglich ein Katalog sein können, der die Informationen, die über methodische Probleme, über die Qualität der Quelle ausführlich Auskunft geben, für den Benutzer aufschließt. Zu den Attributen dieser Relation gehören die Quellenummer, der Quellentyp, Laufzeit oder Entstehungsdatum. Letzteres wird in dieser Relation lediglich als einfache Jahreszahl vermerkt, auch wenn Monat und Tag angegeben sind. Denkbar wäre allerdings wie bei der Relation TOPOS eine Teilung der Relation in Quellen mit Laufzeiten und Quellen mit genauen Datenangaben. Die Quellenummer könnte dann auch als implizite Datierung in anderen Relationen benutzt werden. Dagegen spricht allerdings eine große Zahl von Quellen, wie z.B. viele Büchererien - Amtsbücher, Marktbücher, Kirchenbücher usw. - , die neben klaren Laufzeiten genaue Datierungen für die Einzeleinträge haben. In diesem Fall wird es daher sinnvoller sein, die Relation nicht zu teilen. Entstehungsort und Ortsbetreff sollten in der Relation QUELLE dagegen besser getrennt werden; sie entsprechen den archivalischen Ordnungsprin-

zipien der Provenienz und der Pertinenz, dem Orts- müßte lediglich der Sachbetreff hinzugefügt werden. Schließlich ist zu überlegen, ob eine sachliche Zuordnung der Quelle möglich erscheint angesichts der erheblichen taxonomischen Probleme. Allerdings könnte auf vorhandene Sachordnungsschemata zurückgegriffen werden. Wünschenswert ist eine solche Information, da selbst gleich benannte Quellen landschaftlich außerordentlich unterschiedliche Sachbereiche umgreifen können. Außerdem ist eine erste Orientierung über den der Datenbank zugrundeliegenden Quellenbestand und dessen sachliche Erstreckung möglich.

QUELLE

<u>2N</u>	<u>ZEIT</u>	<u>QUELLART</u>	<u>PROV</u>	<u>PERT</u>	<u>SACHE</u>	<u>BESCH</u>
01	1786	Inventar	Stockach	Heudorf	Preise	001
01	1786	Inventar	Stockach	Heudorf	Erbrecht	001
02	1653-1876	Sterbebuch	Orsingen	Orsingen	Demograph.	002
03	1684	Stat.anim.	Aach	Aach	Familie	003
04	1710-1806	Amtsbuch	Bodman	Bodman	Immobilien	004
04	1710-1806	Amtsbuch	Bodman	Bodman	Devianz	004
04	1710-1806	Amtsbuch	Bodman	Wahlwies	Devianz	004
*	*	*	*	*	*	*

Um nähere Informationen über die Zuverlässigkeit der Quellen zu bekommen, muß der Datenbankbenutzer auf die detaillierten Quellenbeschreibungen (BESCH) verwiesen werden, soweit solche notwendig und vorhanden sind. Daß mit diesen Quellenbeschreibungen ein methodisches Kernstück etwas abseits der mehr, meist weniger harten, von der Datenbank verfügbar gemachten Daten zugänglich ist, muß in Kauf genommen werden. Hier kann auf die Syntax solcher Quellenbeschreibungen nicht eingegangen werden. Ausgeklammert werden müssen auch die offensichtlichen taxonomischen Probleme bei der Bestimmung von Zuverlässigkeit. Kernstück ist dieser Teil der Datenbank für den Historiker allemal, da mit seiner Hilfe eine gründliche Quellenkritik möglich bleibt.

Ort und Quelle bilden das Netz, das über allen anderen Relationen liegt. In allen Relationen wird darüberhinaus als drittes Attribut das Datum bzw. ein Zeitraum enthalten sein, selbst wenn in der Relation QUELLE ein Zeitpunkt oder -räum bereits genannt wird.

Nun kann im folgenden kein vollständiger Wegweiser durch die übrigen Relationen erwartet werden. Dazu gibt es zuviele Gebiete, die den Sachverstand des Spezialisten erfordern, um den Weg von der Quelle zum Datenbankmodell sicher und für spätere Benutzer nachvollziehbar zu machen. Es soll hier nur an beliebig herausgegriffenen Beispielen verdeutlicht werden, wie die Datenbank 'Historische Statistik' grundsätzlich aussehen könnte.

Methodisch grundlegend wie Orts- und Quellennachweise sind die Angaben, die sich aus der oben geforderten Vereinheitlichung von Maß, Gewicht und Währung ergeben. Es wird davon ausgegangen, daß das Edelmetalläquivalent bei Währungen den Wert bestimmt. Erstrebenswert wäre die Erstellung von Lebenshaltungskostenindices. Über viele Jahrhunderte zuverlässige und vergleichbare Indices scheinen freilich nach dem heutigen Wissensstand nicht zu errechnen zu sein. Diesem Problem wird daher im Rahmen des ökonomischen Teils der Datenbank langfristig entsprechende Aufmerksamkeit zu widmen sein. Im folgenden wird die im Dezimalsystem übliche Maß- oder Gewichtseinheit im Attribut UM genannt, während der umgerechnete Wert im Attribut UMRECH zu finden ist.

WAEHRUNG

<u>2N</u>	<u>ORT</u>	<u>DAT</u>	<u>NAME</u>	AUR	ARG
01	Basel	1490.00.00	Rappen	5.3	7.4
02	Frankfurt	1673.03.12	Gulden rh.	4.5	6.3
.

MASS

<u>QN</u>	<u>DAT</u>	<u>ORT</u>	<u>NAME</u>	UM	UMRECH
01	1689.00.00	Radolfzell	Malter rauh	kg	121.3
01	1689.00.00	Radolfzell	Malter glatt	kg	83.2
01	1689.00.00	Radolfzell	Viertel rauh	kg	7.6
01	1689.00.00	Radolfzell	Imi rauh	kg	1.9
.

GEWICHT

<u>QN</u>	<u>DAT</u>	<u>ORT</u>	<u>NAME</u>	UM	UMRECH
01	1753.00.00	Konstanz	Pfund leicht	g	465
01	1753.00.00	Konstanz	Pfund schwer	g	575
01	1753.00.00	Konstanz	Lot leicht	g	14
.

Im Anschluß an die oben aufgeführten Relationen liegt es nahe, sich der im engeren Sinne ökonomischen Problembereiche mit einigen Beispielen anzunehmen. Diese werden ausschließlich aus der Zeit vor 1800 stammen. Für die Zeit der amtlichen Statistik ist seit langem klar, daß es dringend einer Datenbank bedarf, die besonders häufig benutztes Material so zugänglich macht, daß nicht jeder Bearbeiter von neuem das entsprechende Datenverarbeitungspulver entdecken muß (JARAUSSCH 1986). Es wird sich dabei herausstellen, daß einer Datenbank 'Historische Statistik' Vorreiteraufgaben zuwachsen können: große Teile der amtlichen

Statistik werden von den Historikern gar nicht oder wenig genutzt, weil der Verarbeitungsaufwand dem präsumtiven Ertrag nicht entspricht. Im übrigen wird an dieser Stelle die Vielfältigkeit der Datenbank deutlich, da es neben den Textanteilen nahezu gleichwertig ausschließlich statistisch orientierte Teile geben wird.

Preise und Löhne sind ein klassisches Kapitel wirtschaftsgeschichtlicher Forschung. Vornehmliche Quellen dafür sind Rechnungs-, Markt- oder auch Zollbücher, die über den Warenumschatz am Markt Auskunft geben. Offen muß die Frage bleiben, ob in diesen Relationen die bereits umgerechneten Werte für Maße, Gewichte und Währungen oder die Quellenangaben aufgenommen werden. Ist letzteres der Fall, so sollten die Quellenwerte wenigstens in Dezimalwerte umgerechnet sein. Diese könnten dann über die entsprechend angelegten Relationen in heutige Maße umgerechnet werden.

PREIS

QN	<u>DAT</u>	ORT	<u>WARE</u>	<u>MENGE</u>	AUR	ARG
01	1801.03.12	Überlingen	Roggen	dz	6.5	9.3
02	1787.11.27	Konstanz	Zwillich	m	0.8	1.2
.

MARKT

QN	<u>DAT</u>	ORT	<u>VERKNAME</u>	VERKORT	<u>KAUFNAME</u>	KAUFORT	WARE	UM	MENGE
01	1724.05.12	Engen	Martin J	Orsingen	Bern er	H Engen	Kernen	dz	23.7
01	1787.12.24	Engen	Martin J	Orsingen	Buch er	B Arbon	Hafer	dz	11.3
.

In dieser Relation wird die Maßeinheit explizit angegeben, ein Verfahren, daß einem impliziten Maß sicherlich vorzuziehen ist. Gerade an dieser Relation wird deutlich, daß eine weitere Zerlegung sicherlich sinnvoll wäre.

LOHN

QN	<u>DAT</u>	ORT	<u>ARBEIT</u>	MASS	AUR	ARG
01	1694.03.12	Engen	Schlachter	001	0.1	0.2
02	1713.05.08	Orsingen	Holzhauer	002	0.1	0.2
03	1725.00.00	Volkertsh.	Papierer	003	6.5	9.9
.

Das Attribut MASS dient der Bestimmung der Arbeitsleistung, die für den genannten Lohn erbracht werden mußte. So könnte z.B. 001 heißen, daß ein Rind geschlachtet worden ist, 002 wäre gleichbedeutend mit einem Klafter Holz, 003 mit einem Jahreslohn. Die Aufnahme eines Getreide-

äquivalents für den Lohn erübrigt sich, da ein solcher Wert über die Relation PREIS errechnet werden könnte, Übereinstimmung der Daten vorausgesetzt.

Die Zahl der ökonomisch wichtigen Daten scheint unerschöpflich. Aus dem gewerblichen und dem agrarischen Bereich sollen zumindest einige Beispiele vorgeführt werden. Es fehlen dabei besonders die vielen, ab dem 19. Jahrhundert vorhandenen Daten aus betriebs- und volkswirtschaftlichen (Gesamt)Rechnungen: Umsatzentwicklung bestimmter Güter, Export, Import, Sozialprodukt usw..

Die gewerbliche Entwicklung kann zunächst einmal aus den am Markt getätigten Umsätzen (MARKT) abgelesen werden. Es ist darüber hinaus wünschenswert, über die Größe und Größenentwicklung gewerblicher Betriebe etwas zu erfahren. Das Problem besteht in der unterschiedlichen Art der Quantifizierung: Einmal werden Webstühle, das andere Mal die Zahl der Lehrlinge, ein drittes Mal die Beschäftigtenzahl überhaupt gezählt. Es ist daher notwendig, zusätzlich zur numerischen Größenangabe ein Attribut TYP einzuführen, das die Art der gemessenen Größe bezeichnet. Es mag allerdings sinnvoller sein, Personen und Sachen entweder innerhalb der Relation durch entsprechende Attribute zu trennen oder gar zwei Relationen zu bilden.

GEWERBE

<u>OK</u>	<u>DAT</u>	<u>ORT</u>	<u>BETRIEB</u>	<u>GROESSE</u>	<u>TYP</u>
01	1657.00.00	Engen	Weber An	3	Webstuhl
02	1891.00.00	Singen	Maggi	257	Arbeiter
03	1791.00.00	Augsburg	Wirker Jo	7	Gesellen
03	1791.00.00	Augsburg	Spuler Be	1	Meister
.

Strenggenommen könnten auch landwirtschaftliche Betriebe in der oben skizzierten Relation Platz finden. Es scheint jedoch sinnvoller, für diesen Bereich eine eigene Relation zu bilden, um neben der Betriebsgrößenstruktur auch Probleme der Nutzung hier unterzubringen. Im Anschluß an urbariale Quellen wäre eine Analyse der Bodennutzung auch über die Relation PARZELLE möglich (vgl. SCHRENK 1987, S. 112ff.), doch gibt es zahlreiche agrarhistorische Quellen höherer Aggregation, die das eben vorgeschlagene Verfahren rechtfertigen, die aber eine Relation PARZELLE nicht ausschließen oder ersetzen.

HOF

<u>QN</u>	<u>DAT</u>	<u>ORT</u>	<u>BESITZER</u>	<u>NUTZUNG</u>	<u>GROESSE</u>	<u>RECHT</u>
01	1618.00.00	Orsingen	Fritschi Jo	Acker	2173	Lehen
01	1618.00.00	Orsingen	Fritschi Jo	Wiese	707	Lehen
01	1618.00.00	Orsingen	Fritschi Jo	Reben	11	Eigen
.

Von nicht weniger Interesse ist das, was ein Betrieb produziert. Auch hier ist es schwierig, die unterschiedlichen Wirtschaftszweigen angehörigen Betriebe in einer Relation unterzubringen.

PRODUKTION

<u>QN</u>	<u>DAT</u>	<u>ORT</u>	<u>BETRIEB</u>	<u>WARE</u>	<u>MASS</u>	<u>MENGE</u>
01	1630.00.00	Langenst.	Oekonomie	Dinkel	kg	754
01	1630.00.00	Langenst.	Oekonomie	Roggen	kg	237
02	1754.00.00	Volkerst.	Hammers ehm.	Stabeis.	z	132
03	1793.00.00	Augsburg	Weber Johann	Wolltuch	m	584
.

Für die branchenspezifische Einordnung des Betriebes könnte eine weitmaschige Taxonomie ausreichen, wie sie in zahlreichen wirtschaftshistorischen Arbeiten zu finden ist. Das gilt gleichermaßen für die oben angeführte Relation GEWERBE.

Neben der gewerblichen ist die landwirtschaftliche Produktion besonders zu berücksichtigen. Hier geht es um die Ertragsberechnungen, die am Verhältnis Aussaat und Ernte bzw. an der Flächenproduktivität durchgeführt werden. Die entsprechenden Relationen sind getrennt. Hier wie in der folgenden Relation sollte das Maß explizit genannt werden.

HEKTAR

<u>QN</u>	<u>DAT</u>	<u>ORT</u>	<u>WARE</u>	<u>ERTRAG</u>
01	1727.00.00	Langenstein	Winterg.	156
01	1727.00.00	Langenstein	Sommerg.	74
.

ERTRAG

<u>QN</u>	<u>DAT</u>	<u>ORT</u>	<u>WARE</u>	<u>ERTRAG</u>
01	1727.00.00	Langenstein	Dinkel	73
01	1727.00.00	Langenstein	Gerste	117
01	1783.00.00	Langenstein	Dinkel	83
.

Hier wie in anderen Relationen erscheinen Ziffern als ganze Zahlen, deren Maßeinheit zuweilen implizit gesetzt ist. Eine Verbindung von Lohn und Lohnempfänger bietet sich dort, wo entsprechende Listen verfügbar sind. Es kann über das Attribut ARBEIT der jeweilige Lohn abgefragt und mit dem Namen des Lohnempfängers verbunden werden.

ARBEIT

QN	DAT	ORT	BETRIEB	<u>NAME</u>	<u>ARBEIT</u>
01	1758	Langenst.	Oekonomie	Martin AG	Magd
01	1758	Langenst.	Oekonomie	Fritschi Jo	Pferdekn.
02	1713	Volkertsh.	Hammerschm.	Schmid Ja	Schmied
03	1725	Volkertsh.	Papiermühle	Auer Js	Papierer
*	*	*	*	*	*

In diesen Relationen werden sich auch sämtliche in unterschiedlichen Quellen (Kirchenbücher, Testamente z.B.) zu findenden Berufsangaben wiederfinden. Ungelöst ist hier das Problem der Erwähnung zweier Berufe für die gleiche Person (z.B. Bauer und Schneider). Zugleich bietet das Attribut NAME gemeinsam mit den Attributen ZEIT und DAT die Möglichkeit, Verbindungen zu sozialhistorischen Quellen herzustellen. Es wäre im übrigen zu überlegen, ob das Attribut ARBEIT durch eine branchenspezifische und/oder hierarchische Kennzeichnung ergänzt werden sollte. Während die erstere in der Relation PRODUKTION bereits angesprochen worden ist, setzt letztere eine Taxonomie sozialer Ordnung voraus, für die es zahlreiche Vorschläge gibt (vgl. z.B. SCHULTZ 1986, ergänzend auch WEYRAUCH 1980). Über diese Taxonomie ergeben sich ebenfalls Bezüge zum sozialhistorischen Teil der Datenbank. Hierbei sollte jedoch keinerlei Fakteninterpretation zum Zeitpunkt der Datenerfassung dem Prinzip der Uniformität geopfert werden. Es wäre dann sinnvoller, eigene Relationen zu bilden, die eine vereinheitlichende Interpretation übernehmen müßten.

Der sozialhistorische Teil der Datenbank enthält Informationen zur Bevölkerung, ihrem Vermögen, den daraus zu zahlenden Steuern, den Ausbildungsverhältnissen, dem abweichenden Verhalten u.v.m.. Auch hier sollen einige Beispiele lediglich die Vorgehensweise bei der Modellierung der Daten veranschaulichen.

Angaben zum Vermögen bzw. zur Besteuerung sind in vielfältiger Form vorhanden. Bereits die oben genannten Relationen PARZELLE und HOF geben über ländliches Vermögen Auskunft, allerdings nur das Liegende. Abgesehen davon, daß es sich in diesen Fällen häufig um Lebensbesitz handelt, bietet auch die Einordnung der Abgaben bzw. Steuern Probleme. Obrigkeitlicherseits erhobene Steuern sind von Abgaben, also Renten klar zu trennen, auch wenn die Grenzen zuweilen fließend sind. Das Vermögen bestimmter (natürlicher oder juristischer) Personen läßt sich am be-

sten erheben aus Quellen der Besteuerung oder der Vermögensteilung. Ihr Vorteil ist, daß das Gesamtvermögen meist in Geldwert angeschlagenen wird.

STEUER

<u>QN</u>	<u>DAT</u>	<u>ORT</u>	<u>NAME</u>	<u>VERMOEG</u>	<u>BETRAG</u>
01	1759.00.00	Konstanz	Muntprat Jo	38974	390
01	1759.00.00	Konstanz	Martin Ja	6789	68
02	1698.00.00	Engen	Auer Js		23
.

VERMOEG

<u>QN</u>	<u>DAT</u>	<u>ORT</u>	<u>NAME</u>	<u>WARE</u>	<u>MENGE</u>	<u>A LR</u>	<u>ARG</u>
01	1758.04.13	Orsingen	Fritsch Jo	Acker	745	278	356
01	1758.04.13	Orsingen	Fritsch Jo	Rinder	10	123	167
02	1787.11.07	Engen	Britsch An	Fahrhabe		96	128
.

ERBL

<u>QN</u>	<u>DAT</u>	<u>NAME</u>	<u>VERMOEG</u>
01	1786.01.01	Martin Jo	3456
02	1786.02.15	Fritschi Jo	5678
.	.	.	.

ERBE

<u>QN</u>	<u>DAT</u>	<u>NAME</u>	<u>ERBTEIL</u>	<u>FAMBEZ</u>
01	1786.01.01	Martin Ja	1728	Sohn
01	1786.01.01	Martin Ma	1728	Tochter
02	1786.02.15	Martin Ag	5678	Frau
.

Zahlreiche Angaben aus Vermögensteilungen oder Besteuerungsakten werden auch in der Relation PREIS zu finden sein. Der Einfachheits halber wird für die Zahlenwerte des Attributes VERMOEG eine einheitlich bestimmbare Größe angenommen, die z.B. der Goldwert sein könnte.

Bevölkerungshistorische Tatbestände für die Zwecke einer Datenbank sind an anderer Stelle aufgelistet worden (SIEGLERSCHMIDT 1986). Hier geht es um die technische Umsetzung. Für die Kirchenbucheintragen bedarf es einer ganzen Reihe von Relationen.

TAUFE

<u>QN</u>	<u>DAT</u>	ORT	<u>NAME</u>	GS	UM	VATER	MUTTER
01	1631.03.18	Orsingen	Auer Jo	01	01	Jakob	Eliner Ma
02	1635.10.13	Engen	Spann Ma	01	01	Aloys	Britsch Ro
03	1636.01.02	Heudorf	Martin Ja	01	02		Martin An
.

Die Umstände der Geburt: ob ehelich, unehelich, notgetauft, totgebo-
ren, Zwilling etc. werden in einem eigenen Attribut UM erfaßt, das Ge-
schlecht im Attribut GS.

ZEUGE

<u>QN</u>	<u>DAT</u>	<u>NAME</u>	ART
01	1631.03.18	AuerSb	T
02	1635.10.13	Britsch Ka	T
03	1646.11.08	Meier Fd	H
.	.	.	.

In dieser Relation sollen sämtliche Zeugenangaben aus allen Kirchen-
büchern erfaßt werden. Der Bezug ergibt sich zum einen aus der Quellen-
nummer und dem Datum, zum anderen wird die Art der Zeugenschaft
gekennzeichnet. Ob für jede Kirchenbucheintragung eine eigene laufende
Nummer notwendig ist, um bei Bezügen die Eindeutigkeit zu gewährlei-
sten, wäre zu erörtern.

An dieser Stelle soll auch das Problem der zusätzlich zum Namen auf-
tauchenden Angaben zur Person aufgenommen werden. Beruf, Herkunft,
Aufenthaltort, Stand, Amt u.v.m. sind Daten, die nur ab und zu, manch-
mal auch gehäuft in Quellen auftauchen. Im letzteren Fall wäre an eine
Erweiterung einer Relation um ein Attribut zu denken. Im ersten Fall
werden eigene Relationen gebildet. Von diesem wird, da er häufiger an-
zutreffen ist, als Norm ausgegangen. Es ist zu betonen, daß solche Perso-
nenstandsdaten nicht nur aus Kirchenbüchern, sondern aus vielen ande-
ren Quellen zu erheben sind. In den bisherigen Relationen sind Attribute
dieser Art daher auch bisher nicht genannt worden. Lediglich in der Re-
lation MARKT gehören die Herkunftsorte von Verkäufer und Käufer in
Relation, um das Markteinzugsgebiet berechnen zu können. Auf Beispiele
von Relationen der hier in Rede stehenden Art wird verzichtet. Die Si-
cherstellung der Bezüge wird das wesentliche Problem bilden.

Bezüglich der Ehebücher soll lediglich eine Frage angesprochen wer-
den. In den Ehebucheinträgen sind bereits ab dem 17. nicht selten die
Eltern der Brautleute verzeichnet. Da diese in der Relation BRAEUTI-
GAM und BRAUT keinen Platz finden, wären die Relationen VATER und
MUTTER sinnvoll. In diesem Fall wären die gleichnamigen Attribute aus
der Relation TAUFE herauszulösen. Auch Angaben zu den Eltern im Ster-

bebuch wären dann an der richtigen Stelle untergebracht. Angaben zum Geburts- oder Todesdatum in den Registern könnten im übrigen in den jeweiligen Relationen TAUFGE oder TOD Platz finden. Inwiefern eine weitgehend automatisch erfolgende Familienrekonstitution möglich ist, muß hier offen bleiben. Das hängt auch sehr stark vom Material ab.

Auch bei anderen bevölkerungshistorischen Quellen ist die Information in Teile zu zerlegen. Was Seelenbeschriebe oder Haushaltslisten angeht, wäre die Mitführung einer laufenden Nummer für jeden Haushalt ein Verfahren, um die von der Quelle bereits hergestellten Bezüge in sämtlichen Relationen reproduzieren zu können.

HAUS

QN	DAT	ORT	<u>STRASSE</u>	BESITZER	WERT
01	1787.00.00	Konstanz	Münzg. 13	Muntprat J	2345
01	1787.00.00	Konstanz	Katzg. 1	Muntprat J	3456
.

HAUSHALT

QN	DAT	ORT	STRASSE	ID	<u>NAME</u>	<u>FUNKT</u>
01	1787.00.00	Konstanz	Münzg. 13	01	Weber Ja	Vater
01	1787.00.00	Konstanz	Münzg. 13	01	Martin Ag	Mutter
01	1787.00.00	Konstanz	Münzg. 13	02	Muntprat J	Vater
.

Ein letztes Beispiel sei aufgegriffen, da es zunächst dazu verführt, die Tabellenform der Quelle in einer Relation eine Entsprechung finden zu lassen (vgl. Abbildung).

Die Relation ALTER hätte aber so auszusehen:

ALTER

QN	DAT	ORT	GRUPPE	<u>GS</u>	<u>STAND</u>	ZAHL	ARBEIT	ZAHL
01	1777.00.00	Hüfingen	01-10	01	01	84		
01	1777.00.00	Hüfingen	01-10	02	01	100		
01	1777.00.00	Hüfingen	01 10	01	03	1		
01	1777.00.00	Hüfingen	11-19	01	01	89	Stud.	2
.

Auf den letzten Seiten sind vor allem solche Beispiele angeführt worden, die quellennahe Daten betreffen. Relationen dieser Art sind schwierig zu konzipieren, da eine ganze Reihe möglicher Bezüge im Auge behalten werden muß. Daten auf höherer Aggregationsebene stellen den Bearbeiter vor weit weniger Probleme. Auf diese Art sind allerdings die Grundzüge der Datenmodellierung deutlich geworden, die anfangs genannt worden sind.

Eine Datenbank 'Historische Statistik*' soll keine größengleiche Abbildung der Quellen bieten, sondern enthält die Daten, welche zerlegbar und abzählbar sind, d.h. keine auf das Gesamtverständnis der Quelle gründende Interpretation erfordern. Die Zerlegung ist auch deshalb notwendig, um die Information in unterschiedlichen Relationen unterbringen zu können: Es galt nämlich, die Zahl der Attribute klein zu halten zugunsten einer Erweiterung der Domänen, d.h. der Menge der Werte, die einem Attribut zugewiesen werden. Bei dem Entwurf der oben abgebildeten Relationen ist der Modellierungsvorgang allerdings nicht bis zu einer kanonischen Datenstruktur vorgedrungen (VETTER 1982, S. 93ff.; VINEK/RENNERT/TJOA 1982, S. 77ff.)- Hier müßte die weitere Arbeit zum einen ansetzen. Zum anderen sind keine Überlegungen zum Einsatz von Wörterbüchern und von Schreibmasken für die Dateneingabe angestellt worden. Auch sind Erörterungen zu Auswertungsroutinen unterblieben. Schließlich ist der so wichtige Bereich der möglichen Verknüpfung der Daten überhaupt nicht angesprochen. Der historisch Kundige wird Anhaltspunkte genug haben. Eine fernere Datenmodellierung wird an diesem Punkt entscheidende Arbeit leisten müssen.

Das Erstellen einer Datenbank 'Historische Statistik' erfordert also einen erheblichen Aufwand nicht nur bei der Konzeption wie auch der Modellierung der Daten, sondern auch einen großen laufenden Aufwand für die Ergänzung bzw. Korrektur der Daten und die Verbesserung der Programmroutinen. Ein solcher Aufwand bedarf einer ständigen Institution, von der dann auch gewisse Impulse für eine Vereinheitlichung bei der Erhebung und Speicherung der Daten ausgehen könnten. Freilich soll damit nicht einem blinden Ordnungsfetischismus das Wort geredet werden, sondern lediglich dem Versuch, durch die wuchernde Anarchie der Daten einige gangbare Wege zu bahnen. Die produktive Anarchie muß dabei erhalten bleiben.

REFERENCES:

- CHAUN'U, P. (1964), Histoire quantitative et histoire sérielle. In: Cahiers V., Pareto 3, S. 165-175.
- CHAUNU, P. (1970), L'histoire sérielle - bilan et perspective. In: Revue historique 94, S. 297-320.
- CHEN, P.P.S. (1976), The Entity-Relationship-Model - Towards a Unified View of Data. In: ACM TODS 1/1, S. 9-36.
- CODD, E.F. (1970), A Relational Model of Data for Large Shared Data Banks. In: ACM (13), S. 377-387.
- CODD, E.F. (1979), Extending the Database Relational Model to Capture More Meaning. In: ACM TODS (4), S. 397-434.

- DATE, C.J. (1986), An Introduction to Database Systems, 2. vol., 4. ed., Reading.
- HAUSMANN, F./HÄRTEL, R./KROPAC, LH./BECKER, P. (Hg.) (1987), Datenetze für die historischen Wissenschaften. Probleme und Möglichkeiten bei Standardisierung und Transfer maschinenlesbarer Daten, Graz.
- JARAUSCH, K.H. (1986), Some Reflections on Coding. In: Thaller, M., Datenbanken und Datenverwaltungssysteme als Werkzeuge historischer Forschung, St. Katharinen, HSF 20, S. 175-178.
- JARITZ, G. (1987), Offenheit zur Öffnung mit der Standardisierung? In: Hausmann/Härtel/Kropac/Becker (Hg.), Datennetze für die historischen Wissenschaften, Graz, S. 129-134.
- SCHRENK, C. (1987), Standardisierte Erfassung, Dokumentation und Auswertung von neuzeitlichen Urbaren. In: Hausmann/Härtel/Kropac/Becker (Hg.), Datennetze für die historischen Wissenschaften, Graz, S. 211-219.
- SCHULTZ, H. (1986), Probleme sozialökonomischer Klassifikation. In: Thaller, M., Datenbanken und Datenverwaltungssysteme als Werkzeuge historischer Forschung, St. Katharinen, HSF 20, S. 179-185.
- SIEGLERSCHMIDT, J. (1986), Überlegungen zu einer Datenbank 'Historische Demographie'. In: Thaller, M., Datenbanken und Datenverwaltungssysteme als Werkzeuge historischer Forschung, St. Katharinen, HSF 20, S. 31-41.
- THALLER, M. (Hg) (1986), Datenbanken und Datenverwaltungssysteme als Werkzeuge historischer Forschung, St. Katharinen, HSF 20.
- TSICHRITZIS, D.C./LOCHOVSKY, F.H. (1982), Data Models, Englewood Cliffs.
- VETTER, M. (1982), Aufbau betrieblicher Informationssysteme, Stuttgart.
- VINEK, G./RENNERT, P.F./TJOA, A.M. (1982), Datenmodellierung: Theorie und Praxis des Datenbankentwurfs. Würzburg/Wien.
- XÉNOPOL, A.D. (1900), Les faits de répétition et les faits de succession. In: Revue de synthèse historique 1, S. 121-136.
- XÉNOPOL, A.D. (1901), La classification des sciences et l'histoire. In: Revue de synthèse historique 2, S. 264-276.
- XÉNOPOL, A.D. (1902), Les sciences naturelles et l'histoire. In: Revue de synthèse historique 4, S. 276-292.
- WEYRAUCH, E. (1980), Über soziale Schichtung. In: Bători, I. (Hg.): Städtische Gesellschaft und Reformation. (Spätmittelalter und Frühe Neuzeit 12), Stuttgart, S. 5-57.