

# ARCHIVIERUNG VON STATISTISCHEN DATEN

*Von Niklaus Bütikofer*

## 1. Einleitung<sup>1</sup>

«Der ist der Herr der Erde, wer ihre Tiefe misst.» Diese Erkenntnis, welche Novalis als Student der Bergwissenschaften Ende des 18. Jahrhunderts in eine poetische Form brachte, ist heute selbstverständlich. Sehr vieles wird heute gemessen oder gezählt, um unsere Herrschaft über die Erde zu vergrössern, und vieles muss heute auch gemessen werden, um die Folgen dieser Herrschaft einigermaßen unter Kontrolle zu halten. In unserer komplexen, unübersichtlich gewordenen Welt sind wir in vielen Bereichen angewiesen auf Daten, welche ein möglichst realitätsnahes Bild wichtiger Aspekte unserer Wirklichkeit liefern. Ohne diese Informationen wären kaum mehr vernünftige Entscheidungen möglich. Die finanziellen Auswirkungen einer Erhöhung der staatlichen Renten beispielsweise könnten nicht abgeschätzt werden ohne Wissen über die altersmässige Zusammensetzung der Bevölkerung. Ebenso sind genaue Angaben über das Vorhandensein belastender Stoffe eine notwendige Grundlage für Massnahmen im Umweltbereich.

Obwohl bereits im Altertum Volkszählungen durchgeführt wurden, ist doch das Bemühen, alle wesentlichen Aspekte unserer Welt in Zahlen

zu fassen und in ihren mengenmässigen Verhältnissen zu untersuchen, eine relativ junge Erscheinung. Ungefähr seit dem 18. Jahrhundert lassen viele Staaten regelmässig und in ständig zunehmendem Masse statistische Erhebungen durchführen.<sup>2</sup> Heute enthält das Statistische Jahrbuch der Schweiz Angaben über fast alle Bereiche des modernen Lebens.<sup>3</sup> Dabei weist das Jahrbuch nur diejenigen Daten aus, welche von der öffentlichen Verwaltung und von einzelnen Verbänden produziert werden. Aber auch Universitäten und private Forschungsinstitute erheben wichtige Daten von allgemeinem Interesse.

Numerische Massendaten eignen sich sehr gut für die maschinelle Verarbeitung. Bereits anlässlich der Volkszählung von 1920 setzte das damalige Eidgenössische Statistische Bureau Loch- und Sortiermaschinen mit mechanischen Zählwerken ein. 1960 wurden die Volkszählungsdaten dann erstmals mit elektronischen Maschinen verarbeitet; als Speichermedium kam die Lochkarte zum Einsatz.<sup>4</sup> Heute ist die Behandlung quantitativer Massendaten ohne EDV nicht mehr denkbar. Die neuen Möglichkeiten der Computer haben die Verarbeitung und Auswertung der Daten enorm beschleunigt und verfeinert. Sie sind mit ein Grund dafür, dass die Masse der produzierten Daten stark gewachsen ist. Als noch keine elektronischen Rechner und Speichermedien zur Verfügung standen, war die Auswertung einer Erhebung so aufwendig, dass sie eine einmalige Angelegenheit blieb. «Wenn alle geplanten Tabellen ausgezählt waren, wurden die Daten – nur schon aus Platzgründen – vernichtet. Eine nachträgliche Zusatzauswertung war dann nicht mehr möglich.»<sup>5</sup> Die Verbreitung von elektronischen Rechnern aller Grössenordnungen erlaubt es aber heute fast jedermann, umfangreiche Daten in sehr kurzer Zeit nach spezifischen, individuell bestimmten Gesichtspunkten auszuwerten. Dementsprechend hoch ist heute auch die Nachfrage nach dem «Rohstoff» Daten in maschinenlesbarer Form.

Es ist keine Frage, dass viele dieser Daten dauernd wertvoll sind. Für die politische Steuerung ist es oft unerlässlich, anhand von längeren Zeitreihen die Dynamik bestimmter Entwicklungen zu beurteilen, und für alle historisch ausgerichteten Wissenschaften bieten die Daten eine Grundlage, deren grosser Wert uns erst in Zukunft ganz bewusst werden wird, wenn es gilt, die Hypothesen unserer Zeit, die gerade im Umweltbereich immer deutlicher werden, abzutragen. Auch die

Geschichtswissenschaft hat sich seit längerer Zeit komplexen Zusammenhängen zugewandt und benötigt quantitative Methoden und serielle Daten über vergangene ökonomische, soziale und neuerdings auch ökologische Zustände und Entwicklungen. Dabei spielt die elektronische Datenverarbeitung ebenso eine zentrale Rolle wie in der amtlichen Statistik. Die Computer haben der Geschichte neue Forschungsfelder und Quellenbestände erschlossen, die bisher wegen ihrer Masse als unbearbeitbar galten.<sup>6</sup>

Damit sind auch diejenigen Institutionen angesprochen, welche sich hauptsächlich mit der Überlieferungsbildung und -bewahrung befassen, die Archive. Im Sinne einer Dienstleistung für die staatliche Verwaltung, die historisch ausgerichteten Wissenschaften und die interessierte Öffentlichkeit sichern und verwahren die Archive dauernd wertvolle Akten und Daten, welche die vorwiegend staatlichen Organe ihres jeweiligen Zuständigkeitsbereichs bei ihrer Tätigkeit produzieren.<sup>7</sup> Ihre schwierigste Aufgabe ist es, zu bestimmen, welche in der Gegenwart produzierten Unterlagen auch in Zukunft in welcher Form zur Verfügung stehen sollen. Sie haben dabei die neuen Möglichkeiten der elektronischen Datenverarbeitung in ihre Überlegungen einzubeziehen und ihre Arbeitsweise entsprechend anzupassen.

Die Beschäftigung mit Fragen der Bereiche *maschinenlesbare Massendaten und Archive* bzw. *quantitative Geschichte und Archive* ist nicht etwa neu; eine erste Welle von Aufsätzen zu diesem Problemkreis ist bereits Ende der 60er und im Laufe der 70er Jahre veröffentlicht worden.<sup>8</sup> Die Überlegungen waren ausgelöst worden durch den Aufschwung quantitativer Methoden in den Geschichtswissenschaften nach dem Zweiten Weltkrieg und das Erscheinen von Computern in den Schreibstuben der Historiker in den 60er Jahren. Gleichzeitig konnten Archivare und Historiker beobachten, wie in anderen Fachbereichen Dokumentationseinrichtungen entstanden, welche Wissenschaftler und interessierte Öffentlichkeit mit Informationen aus ihrem Fachgebiet versorgen wollten. Für die Archivare war dies Anlass, über «die zukünftige Funktion der öffentlichen Archive in einem Informationsbankensystem» nachzudenken.<sup>9</sup> Man erkannte die drohende Gefahr, dass die Archive im «Informationsmarkt» an den Rand gedrängt würden. Der Aufbau von integrierten Fachinformationssystemen ging allerdings nicht so schnell vorwärts, wie man sich das noch vor zwanzig Jahren

vorgestellt hatte, aber trotzdem sind bereits beachtliche Resultate erzielt worden, auch im Bereich der Geschichtswissenschaften. Verschiedene historische Datenbanken in maschinenlesbarer Form sind bereits vorhanden und viele sind im Aufbau begriffen.<sup>10</sup> Damit verbunden hat sich eine eigentliche historische Fachinformatik konstituiert und in der *Association for History and Computing* zusammengeschlossen.<sup>11</sup> Die staatlichen Archive sind dabei zumeist nicht über die Rolle von Zaungästen hinausgekommen.

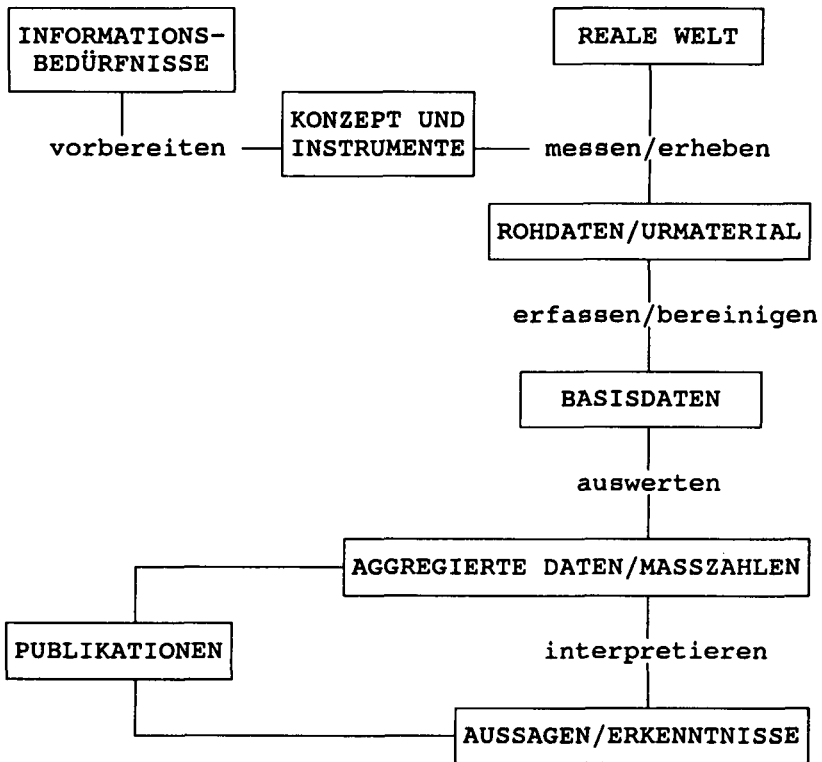
Der vorliegende Artikel soll ein Beitrag dazu sein, die archivischen Verfahren im Bereich der statistischen Daten weiterzuentwickeln und den neuen Gegebenheiten, die vor allem durch die Anwendung der elektronischen Datenverarbeitung entstanden sind, anzupassen. Unter dem Begriff «statistisch» verstehe ich hier in einem sehr weiten Sinne alle Bereiche, in denen Informationen über unsere Wirklichkeit mittels quantitativer Verfahren erhoben und verarbeitet werden. Das Spektrum reicht von der Volkszählung über die Messung von Wetterdaten bis zur Befragung von Rekruten. In den folgenden Kapiteln versuche ich ausgehend von einer Darstellung des statistischen Produktionsprozesses und einiger in diesem Zusammenhang wichtiger Eigenschaften historischer Forschung Leitlinien für die archivische Arbeit mit statistischen Daten herauszuarbeiten. Im letzten Kapitel schliesslich gehe ich auf die Frage ein, welche Funktionen die Archive, insbesondere die staatlichen, im Bereich statistischer Daten zwischen den Datenproduzenten und Datenanbietern auf der einen Seite und den Forschern als Datenbenutzern auf der anderen Seite wahrzunehmen haben.

## **2. Eigenschaften statistischer Arbeit und statistischer Daten<sup>12</sup>**

Die Gewinnung und Auswertung von statistischen Daten ist ein langer und aufwendiger Arbeitsprozess, in dessen Verlauf eine Menge verschiedener Unterlagen und Daten anfallen. Der Ablauf ist schematisch in *Grafik 1* dargestellt. In dieser verallgemeinerten Form kann das Schema sowohl für quantitative Erhebungen gesellschaftlicher Sachverhalte als auch für instrumentelle Messungen natürlicher Phänomene gelten.

Am Anfang jeder Erhebung bzw. Messung stehen bestimmte *Informationsbedürfnisse* über die *reale Welt*. Die heutige amtliche Statistik hat vielerlei Zwecken zu genügen, und ihre Informationsaufträge sind recht allgemein formuliert wie beispielsweise in der jüngsten Volkszählungsverordnung: «Die Volkszählung soll den Behörden des Bundes, der Kantone und der Gemeinden, den verschiedenen Kreisen der Wirtschaft, des sozialen, politischen und kulturellen Lebens, den Hochschulen und übrigen Trägern von Forschung und Lehre sowie allen weiteren Interessenten statistische Daten zur Verfügung stellen, die als Grundlage für Planungen und Entscheide, zu Zwecken der Forschung oder zur Information der Öffentlichkeit erforderlich sind.»<sup>13</sup> Die ersten Bevölkerungszählungen beispielsweise hatten aber den konkreten Zweck, die Wehr- oder Steuerkraft eines Staates zu erfassen. Es ist

Grafik 1: Produktionsprozess statistischer Daten



naheliegend, dass dieses spezifische Interesse des Staates die Antwortbereitschaft der Bevölkerung und damit die Qualität der Daten nicht unwesentlich beeinflusst hat.<sup>14</sup> Auch die moderne Sozialforschung will mit ihren Umfragen ganz gezielt bestimmte Hypothesen über soziale Zusammenhänge überprüfen und richtet die Anordnung der Erhebung auf diesen Zweck aus.

Statistische Daten werden unter Berücksichtigung bestimmter Grundsätze und Verfahren produziert, welche die Wissenschaft der Statistik und die einzelnen Bereichswissenschaften (zum Beispiel die Demografie oder die empirische Sozialforschung) entwickelt haben. Diese Grundsätze und Verfahren werden während der Vorbereitungsarbeiten in einem *Erhebungs-* bzw. *Mess-Konzept* konkretisiert. Dieses legt in der Regel fest, was genau auf welche Weise und mit welchem Instrumentarium gezählt, gemessen und ausgewertet werden soll. Dabei müssen die Informationsbedürfnisse, die zur Verfügung stehenden Verfahren und die reale Welt in optimaler Weise aufeinander abgestimmt werden. Charakteristisch dabei ist, dass sich die Wirklichkeit immer nur über starke Vereinfachungen in Begriffe und quantitative Angaben fassen lässt. Grundsätzlich muss sich jede Erhebung und Messung auf das beschränken, was im Hinblick auf die Informationsbedürfnisse wesentlich ist und was mit dem zur Verfügung stehenden Instrumentarium überhaupt erfasst werden kann. Dass diese konzeptuelle Arbeit schwierig und für die Qualität der erhobenen Daten entscheidend ist, zeigt indirekt die Kritik, die etwa den Preisindices in der Öffentlichkeit immer wieder erwächst. Gerade hier sind die Statistiker gezwungen, aus einer riesigen Anzahl von einzelnen Preisbewegungen, die niemals alle erfasst werden können, eine kleine Auswahl zu treffen und diese auf die Teuerungsrate der Konsumentenpreise zu reduzieren, welche die «Preisentwicklung der für die privaten Haushalte bedeutsamen Waren und Dienstleistungen» in ausreichendem Masse repräsentieren sollte.<sup>15</sup>

Selbst einfache Ausdrücke wie «Bevölkerung» bedürfen bei näherer Betrachtung genauerer Bestimmung. Es ist nicht von vornherein klar, ob «Bevölkerung» nur die am Stichtag der Zählung an ihrem offiziellen Wohnsitz angetroffenen Personen einschliesst oder ob Leute, die unterwegs sind, Touristen, die sich gerade im Lande aufhalten, oder Einwohner, die vorübergehend im Ausland weilen, ebenfalls mit eingeschlossen werden müssen. Zur Bestimmung einer komplexen Grösse,

wie sie beispielsweise das Bruttosozialprodukt darstellt, sind ganze Theoriegebäude erforderlich.<sup>16</sup> Die Naturwissenschaftler haben es nicht etwa leichter: die Lufttemperatur zum Beispiel ist 10 cm über dem Boden oft ziemlich verschieden von der in einem Meter Höhe gemessenen. Hinzu kommt, dass neben der Messhöhe auch der Standort und die verwendeten Messgeräte einen erheblichen Einfluss auf die Ergebnisse haben.

Wenn die Daten erhoben sind, dann müssen sie heute zunächst auf maschinenlesbare Speichermedien übertragen werden, sofern dies nicht bereits automatisch geschieht. Anschliessend werden sie in verschiedenen Schritten aufbereitet:

- Verschlüsselung: Individuelle, umgangssprachliche Angaben müssen vereinheitlicht und allenfalls in eine feste vorgegebene Nomenklatur und einen alfanumerischen Kode übertragen werden.
- Kontrolle: Die Daten werden mittels bestimmter Verfahren auf Vollständigkeit und eindeutige Fehler überprüft (Plausibilitätstests).
- Korrektur und Ergänzung: Die festgestellten Lücken und Fehler werden entweder über gezielte Nacherhebungen oder mit bestimmten Schätzverfahren ergänzt und korrigiert.
- Anonymisierung: Aus personenbezogenen Daten werden heute in der Regel die Namen entfernt.

Das Ergebnis dieser Arbeiten sind die *bereinigten Basisdaten*, die als Grundlage für alle weiteren Berechnungen dienen.

Im nächsten Schritt werden diese Basisdaten im Hinblick auf die gewünschten Erkenntnisse ausgewertet. Die Wissenschaft der Statistik hat dafür eine Vielzahl von einzelnen Methoden entwickelt. Allen gemeinsam ist, dass sich mit ihnen die Basisdaten auf bestimmte Eigenschaften hin untersuchen lassen und dass als Ergebnis jeweils neue, verdichtete Daten entstehen. In der Regel ergibt sich dabei ein Informationsverlust, d.h. die verdichteten Daten enthalten weniger Detailangaben als die Ausgangsdaten, weshalb auch eine Rekonstruktion der Ausgangsdaten auf der Grundlage der verdichteten Daten gewöhnlich nicht mehr möglich ist. Die häufigste Form der Auswertung ist die Zusammenfassung der Einzeldaten in Tabellen. Aus den Individualdaten einer Volkszählung kann beispielsweise eine Tabelle mit der Bevölkerungszahl aller Schweizer Gemeinden erstellt werden. Die

Ergebnisse der vielen möglichen Auswertungen werden in *Grafik 1* vereinfachend als *aggregierte Daten* und statistische *Masszahlen* bezeichnet.

In einem abschliessenden Schritt müssen die Ergebnisse der Auswertung interpretiert werden. Dabei ist etwa die Aussagekraft der Auswertungsdaten zu beurteilen, räumliche und zeitliche Veränderungen bestimmter Werte sind zu erklären, oder die Zusammenhänge verschiedener Merkmale sind herauszuarbeiten. Die Ergebnisse dieser Arbeiten bezeichne ich hier verallgemeinernd als *Aussagen* und *Erkenntnisse*.

Meist werden die Erkenntnisse und Teile der erarbeiteten Daten über verschiedene Kanäle und auf verschiedenen Medien veröffentlicht. Am bekanntesten sind die verschiedenen Druckschriften-Reihen der amtlichen Statistik.<sup>17</sup> Gewöhnlich finden aber nur aggregierte und ausgewertete Daten Eingang in *Publikationen*. Für wissenschaftliche Zwecke und unter bestimmten Bedingungen geben die Statistik-Stellen auch Basisdaten auf maschinenlesbaren Datenträgern ab. Die Druckschriften bieten meist in ihrer Einleitung wichtige Informationen über die jeweilige Datenerhebung und die Datenqualität.

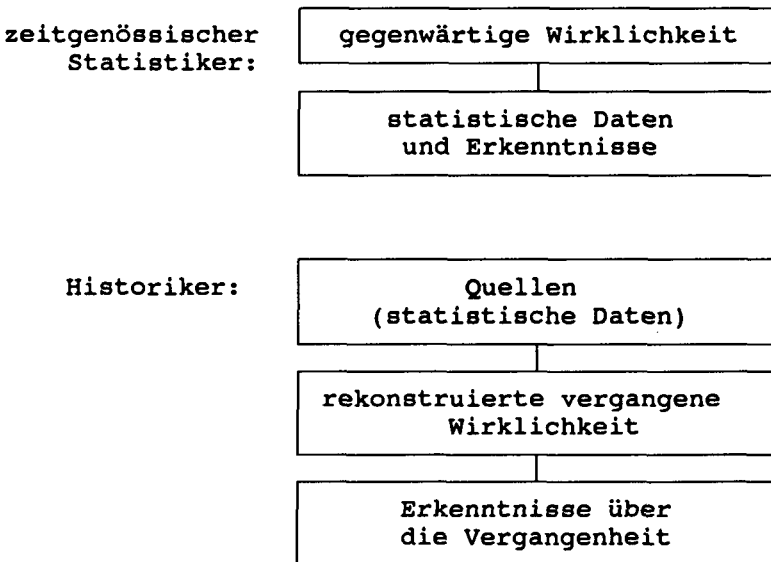
### **3. Eigenschaften historischer Quellen und historischer Forschung**

«Die Qualifizierung zur dauernden Aufbewahrung im Archiv, die Feststellung ihrer Archivwürdigkeit, ist der konstitutive Akt, der gesellschaftliche Daten zu ›historischen Materialien‹ werden lässt.»<sup>18</sup> Statistische Daten wandeln sich also im Archiv zur historischen Quelle, d. h. zu einem überlieferten Zeugnis vergangener Wirklichkeit. Wer auch immer diese Daten fortan verwendet, wird sich in der Rolle des Historikers finden, der daraus Informationen über vergangene Sachverhalte gewinnen will. Er wird den heute produzierten Unterlagen bei der Wiederbenützung nach einiger Zeit in ähnlicher Art entgegentreten müssen, wie heutige Historiker den Quellen der Vergangenheit. Ich setze voraus, dass es diesem künftigen Datenbenutzer, unabhängig davon, welche Einzelwissenschaft er betreibt, zunächst darum geht, ja

darum gehen muss, die Vergangenheit in ihren wesentlichen Zügen möglichst getreu zu rekonstruieren, bevor er den je wissenschaftsspezifischen Zielen und Methoden gemäss weiterarbeitet.

Die Situation des Historikers unterscheidet sich grundsätzlich von derjenigen des Statistikers, der die Erhebung durchgeführt und erstmals ausgewertet hat. Während der Zeitgenosse seine gegenwärtige Wirklichkeit als empirische Grundlage für seine Arbeit zur Verfügung hatte, muss sich der Historiker mit den Resultaten des Statistikers als empirischer Grundlage für seine Untersuchung begnügen (vgl. Grafik 2). Wir haben davon auszugehen, dass die zukünftigen Historiker andere Fragen haben und andere Methoden verwenden als der einstige Datenproduzent. Da sie aber die Erhebung nicht mehr wiederholen können, müssen sie versuchen, mit Hilfe quellenkritischer Methoden aus dem überlieferten Material diejenigen Informationen über die vergangene Wirklichkeit herauszuholen, welche Antworten auf ihre Fragen erlauben.

**Grafik 2: Untersuchungsebenen des zeitgenössischen Statistikers und des Historikers**



Die Quellenkritik ist der Kern der historischen Methode. Sie untersucht die Beziehung zwischen historischer Wirklichkeit und Quelle mit dem Ziel, deren Aussagewert und Informationsgehalt zu erschliessen.<sup>19</sup> Der Historiker braucht zu diesem Zweck Zusatzinformationen, die ihm Auskunft über die Entstehung und die «Lebensgeschichte» einer Quelle geben können. Im Falle von statistischen Daten müssen diese Zusatzinformationen alle Stufen des Entstehungsprozesses der Daten dokumentieren, damit die Zuverlässigkeit und die Reichweite der Daten sowie immer auch deren Vergleichbarkeit mit Daten aus räumlich und zeitlich verschiedenen Erhebungen beurteilt werden kann. Die Zusatzinformationen sollen dem Historiker auch die Möglichkeit bieten, allfällige nachträglich festgestellte systematische Fehler korrigieren zu können. Im allgemeinen muss davon ausgegangen werden, dass der Historiker mehr explizite Zusatzinformationen zum Verständnis von Daten benötigt als der Zeitgenosse, da für diesen gerade im Bereich der Definition von Merkmalen und Werteklassen vieles zum selbstverständlichen Alltagswissen gehörte, welches sich seither, ohne deutliche Spuren zu hinterlassen, verändert hat.<sup>20</sup>

Welche konkreten quellenkritischen Probleme sich stellen, wenn man heute zum Beispiel den Informationsgehalt von Volkszählungen des 19. Jahrhunderts erschliessen will, hat W. Lüönd kürzlich in einer Arbeit am Beispiel des Kantons Zug gezeigt.<sup>21</sup> Er weist unter anderem darauf hin, dass Zeitpunkt und Dauer der Zählungen wichtig sind, «denn je länger eine Erhebung dauert, um so mehr muss mit Auslassungen und Doppelzählungen gerechnet werden».<sup>22</sup> Es kamen auch bei praktisch jeder Zählung und zum Teil sogar im Verlaufe einer einzigen Zählung unterschiedliche Zählkriterien zur Anwendung, so dass Vergleiche beinahe verunmöglicht werden. Es ist ausserdem nicht immer klar, ob gewisse Kategorien von Personen, beispielsweise politische Flüchtlinge oder Heimatlose, die es damals in recht grosser Zahl gegeben hat, jeweils mitgezählt wurden oder nicht.

Natürlich arbeitet die amtliche Statistik heute viel sorgfältiger und differenzierter als im 19. Jahrhundert. Aber auch sie kommt nicht darum herum, ihre Zählkriterien den Veränderungen der realen Welt anzupassen. Ein gutes Beispiel sind die Nomenklaturen der Berufe oder Wirtschaftszweige; ihre verschiedenen Revisionen widerspiegeln die sozio-ökonomischen Veränderungen und erschweren damit gleichzeitig

den für Historiker wichtigen Vergleich über grössere Zeiträume hinweg. Das Begriffsnetz, das die Statistik bei ihren Erhebungen über die soziale Wirklichkeit wirft, und sein Wandel ist sogar zum Gegenstand für spezielle historische Untersuchungen geworden.<sup>23</sup>

#### **4. Leitlinien für die Archivierung**

Wie sollte nun die Überlieferung von statistischem Datenmaterial beschaffen sein, damit der zukünftige Historiker daraus möglichst korrekt und einfach die wesentlichen Züge unserer Gegenwart rekonstruieren kann? Zur Auswahl für die dauernde Aufbewahrung steht grundsätzlich alles Material, das im Laufe des statistischen Produktionsprozesses anfällt, insbesondere natürlich die drei grundsätzlichen Verarbeitungsstufen Urmaterial oder rohe Erhebungs- bzw. Messdaten, bereinigte Basisdaten und aggregierte Daten.

Vier allgemeine Eigenschaften werden – soweit wir das voraussehen können – den Wert einer Quelle für den zukünftigen Historiker bestimmen:

1. Die Relevanz der durch diese Quelle dokumentierten Sachverhalte für die zukünftigen Forschungsinteressen.
2. Der Gehalt an Informationen über die jeweilige Wirklichkeit.
3. Das Vorhandensein ausreichender Dokumentation über die Quelle und ihren Kontext, welche es erlaubt, den Informationsgehalt der Quelle vollumfänglich zu erschliessen und kritisch zu beurteilen.
4. Die Form und das Speichermedium der Quelle, welche beide so beschaffen sein sollten, dass sich der Forscher mit möglichst wenig Aufbereitungsaufwand Zugang zu den Informationen verschaffen kann.

Die Bedeutung statistischer Daten über vergangene Sachverhalte für die zukünftige Forschung ist bereits in den vorangehenden Kapiteln hervorgehoben worden. Es soll aber damit nicht geleugnet werden, dass es heute, in einer Zeit, in der fast alles auf irgendeine Weise quantifiziert wird, auch statistische Erhebungen über völlig unbedeutende Bereiche gibt. Der Archivar wird also vor allem bei Daten, die ausserhalb der

amtlichen Statistik produziert worden sind, die in Punkt 1 geforderte Relevanz für künftige Forschungsfragen im Einzelfall feststellen müssen. Er wird sich aber weitgehend auf die Regel verlassen können, dass die Daten aus der amtlichen Statistik generell archivwürdig sind, sofern die Beurteilung nach den übrigen Kriterien ebenfalls positiv ausfällt.

Das zentrale Bewertungskriterium ist der *Gehalt an Information* über die jeweilige Wirklichkeit. Es hilft dem Archivar, zu entscheiden, in welcher Verarbeitungsstufe er statistische Daten übernehmen soll. In den folgenden Überlegungen gehe ich von drei Voraussetzungen aus. Die erste ist die Annahme, dass es aus Gründen der Wirtschaftlichkeit und der Übersichtlichkeit nicht sinnvoll ist, alles statistische Material dauernd im Archiv aufzubewahren, sondern dass es vielmehr erstrebenswert ist, Redundanzen im Datenmaterial zu vermeiden. Zweitens setze ich voraus, dass das statistische Interesse und die statistischen Methoden einem steten Wandel unterworfen sind und dass deshalb der künftige Historiker nicht dieselben Auswertungen vornehmen will wie der heutige Datenproduzent. Und drittens gehe ich schliesslich davon aus, dass sich die Leistungsfähigkeit der elektronischen Datenverarbeitung weiterhin verbessert und sich der jetzt schon kleine Aufwand für die Herstellung von Aggregationen und anderen rechnerischen Auswertungen noch mehr vermindert.

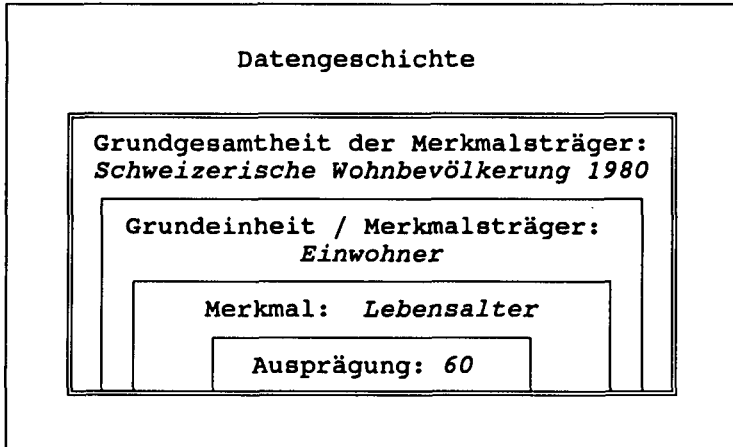
Bereits in *Kapitel 2* ist festgestellt worden, dass die meisten statistischen Auswertungen als Informationsverdichtungen betrachtet werden können und dass sich aus verdichteten Daten die detailreicheren Ausgangsdaten in der Regel nicht mehr zurückgewinnen lassen. In bezug auf das Kriterium Informationsgehalt sind also die Basis- bzw. die Rohdaten den aggregierten Daten vorzuziehen, auch wenn letztere sicher prägnantere und unter bestimmten Gesichtspunkten aussagekräftigere Informationen bieten. Die Entwicklung der elektronischen Datenverarbeitung legt ausserdem nahe, Basisdaten auf maschinenlesbaren Datenträgern als Einheit mit entsprechender Hard- und Software zu verstehen. So gesehen bilden Grunddaten zusammen mit Datenträger, Lese- und Rechenggerät sowie mit standardmässigen Algorithmen zum Verwalten und Auswerten der Daten ein einziges Informationssystem. Sämtliche durch die Software ermöglichten Aggregationen gehören somit als virtuelle Auswertungen zu den Grunddaten. Zu

diesen Argumenten für die Archivierung von Basisdaten kommt noch hinzu, dass diese meist eine Menge Angaben enthalten, die nicht in die Primärauswertung einbezogen werden und die deshalb verloren gingen, wenn nur die Ergebnisse archiviert würden.

Schwieriger ist die Frage zu entscheiden, ob die Rohdaten bzw. das Urmaterial oder die bereinigten Basisdaten für die dauernde Aufbewahrung ins Archiv zu übernehmen sind. Die Erfassung und Bereinigung der Rohdaten bewirkt zwar eine Informationsveränderung, aber nicht unbedingt einen Informationsverlust, wenn man von der aus Gründen des Datenschutzes heute meist unumgänglichen Anonymisierung von personenbezogenen Daten absieht.<sup>24</sup> Es werden in dieser Verarbeitungsphase beispielsweise fehlende Angaben oder offensichtlich falsche Werte durch Schätzungen ergänzt bzw. korrigiert. An den bereinigten Basisdaten ist im nachhinein nicht mehr festzustellen, welche Werte auf blossen Schätzungen beruhen. Dies ist aber kein Nachteil, wenn bekannt ist, wie hoch der Anteil der ergänzten und korrigierten Werte bei den verschiedenen erhobenen Merkmalen ist, denn damit lässt sich der Einfluss auf die Genauigkeit der Daten beziffern. Unter dem Aspekt des Informationsgehalts sind also zwischen Rohdaten und bereinigten Basisdaten keine wesentlichen Unterschiede auszumachen.

Das dritte der am Anfang dieses Kapitels genannten vier Kriterien für einen Archivierungsentscheid betrifft die Qualität der *Begleitdokumentation*. Statistische Daten bestehen in der Regel aus einer Menge von bestimmten Werten in einer bestimmten Anordnung. Ein Benutzer kann damit allein noch nichts anfangen. Er benötigt zusätzliche Informationen, die es ihm ermöglichen, die Daten korrekt zu verstehen, zu beurteilen und weiterzuverwenden. Daten, deren Schwachstellen man nicht genau kennt, verleiten gerne zu Fehlschlüssen, und es ist oft besser «auf ungenaue Zahlen, wenn sie einmal als solche entlarvt wurden, zu verzichten (...), als sie durch kunstreiche Manipulationen zu retten».<sup>25</sup> Die notwendige Dokumentation statistischer Daten besteht aus einer Dateibeschriftung, welche erst eigentlich die Bedeutung der Einzelwerte konstituiert, aus einer Datengeschichte, anhand welcher der Aussagewert der Daten beurteilt werden kann, sowie aus einer Beschreibung und Geschichte des Datenträgers und aus Hinweisen auf Möglichkeiten und Bedingungen der Benutzung.

Grafik 3: **Bedeutungsaufbau statistischer Daten** (in kursiven Lettern Beispiele einer Volkszählung)



Der Bedeutungsaufbau statistischer Daten ist in *Grafik 3* dargestellt. Die eigentliche *Dateibeschreibung* umfasst nur die innerhalb der doppelten Linie angeführten Ebenen (vgl. auch *Übersicht 1*). Darin muss zunächst die Grundeinheit der Erhebung definiert werden. Es wird sich dabei in der Regel um natürliche Elemente wie Einzelpersonen, Haushalte, Unternehmen, aber beispielsweise auch um Wetterstationen handeln. Von diesen statistischen Einheiten ist die Grundgesamtheit<sup>26</sup> in sachlicher, räumlicher und zeitlicher Abgrenzung genau zu bestimmen. Quantitative Angaben haben erst dann Informationswert, wenn bekannt ist, auf welche Grundgesamtheit sie sich beziehen. Die Merkmale, die bei den einzelnen Einheiten erhoben werden, sowie Art und Bereich der möglichen Merkmalsausprägungen (Wertebereich) sind ebenfalls genau anzugeben. Handelt es sich um Unterschiedsmerkmale in Form einer Kategorienskala, dann ist besondere Sorgfalt auf die Beschreibung der einzelnen Kategorien zu legen, weil sich hier am schnellsten Veränderungen ergeben, besonders im sozio-ökonomischen Bereich. Falls Werte in kodierter Form vorliegen, müssen die entsprechenden Codeverzeichnisse der Datenbeschreibung unbedingt beigelegt werden.

## Übersicht 1: Dokumentation statistischer Basisdaten bei der Archivierung

- 1. Dateibeschreibung:**  
(Enthält die Datei verschiedene Grundeinheiten, dann sind alle Angaben für jede einzelne davon zu machen.)
  - 1.1 Name und Umfang der Datei
  - 1.2 Statistische Grundeinheit (auch Entität oder Merkmalsträger genannt):  
Definition und Anzahl der Grundeinheiten.
  - 1.3 Grundgesamtheit der statistischen Einheiten:  
Eine genaue sachliche, räumliche und zeitliche Abgrenzung ist notwendig. Bei Stichproben ist zusätzlich das Verfahren und der Umfang der Stichprobenziehung anzugeben.
  - 1.4 Merkmale der Grundeinheit (auch Attribute genannt):  
Eine ausführliche Definition der Merkmale und aller jeweils gültigen Ausprägungen (Werte) dieser Merkmale sowie allenfalls für die Notierung verwendete Codes müssen angegeben werden.
  - 1.5 Anordnung der Daten auf dem Datenträger (Recordaufbau):  
Die Startposition und die Länge der einzelnen Merkmale sowie Struktur und Inhalt allfälliger "File Header Records" oder "End of File Records" müssen angegeben werden.
- 2. Datengeschichte:**
  - 2.1 Rechtliche Grundlagen der Erhebung
  - 2.2 Zweck der Erhebung (sofern dieser nicht bereits in den rechtlichen Grundlagen in genügendem Masse erläutert ist.)
  - 2.3 Verfahrensweise bei der Datenbeschaffung:  
Organisation und Zeitpunkt der Datenerhebung sollten angegeben werden. Bei Umfragen sind ein Originalfragebogen mit allfälligen Erläuterungen für den Befragten und Anweisungen für den Befragten beizulegen.
  - 2.4 Verfahrensweise bei der Erfassung und Bereinigung der Rohdaten:  
Die bei der Verschlüsselung, der Kontrolle, der Korrektur und Ergänzung der Rohdaten angewandten Verfahren sollten angegeben werden, ebenso der Umfang der bei jedem einzelnen Merkmal vorgenommenen Korrekturen und Ergänzungen.

- 2.6 **Beurteilung der Datenqualität:**  
Falls Urteile über die Repräsentativität und die Aussagekraft der Daten vorliegen, sollten sie der Dokumentation beigelegt werden. Allfällige Nacherhebungen und andere Verfahren der Qualitätskontrolle müssen ebenfalls dokumentiert werden.
- 2.7 **Veränderung der Daten durch den Datenproduzenten nach der Bereinigung:**  
Form, Umfang und Zweck von allfälligen Umstrukturierungen, Veränderungen oder Löschungen von Daten sowie allfällige unabsichtliche Verluste.
- 2.8 **Veränderung der Daten im Archiv:**  
Form, Umfang und Zweck von allfälligen Umstrukturierungen, Veränderungen oder Löschungen von Daten sowie allfällige unabsichtliche Verluste.
3. **Beschreibung und Geschichte des Datenträgers:**
  - 3.1 Technische Spezifikationen des Speichermediums und der Aufzeichnungsweise der Daten
  - 3.2 Hersteller und Herstelungsdatum
  - 3.3 **Beschreibungen und Löschungen:**  
Beschreibungen sollten mit Datum und Namen der aufgezeichneten Datei festgehalten werden. Bei Löschungen sollte zusätzlich zu Datum und Namen der gelöschten Datei das angewandte Lösungsverfahren angegeben werden. Wenn möglich sind die Logbücher einer jeden Bearbeitung des Datenträgers aufzubewahren.
4. **Benutzung:**
  - 4.1 **Benutzungsbedingungen:**  
Bedingungen, die von den entsprechenden Bestimmungen des Archivreglements abweichen, sind speziell zu erwähnen.
  - 4.2 **Verknüpfungsmöglichkeiten:**  
Möglichkeiten zur Verknüpfung der Datei mit anderen Dateien aufgrund von gemeinsamen Schlüsselmerkmalen sind anzugeben.
5. **Publikationen:**  
Publikationen und Auswertungen der Daten sind soweit möglich anzugeben.

Die *Datengeschichte* dient dem späteren Benutzer in erster Linie zur Beurteilung der Qualität der Erhebung. Eine Datengeschichte sollte den Auftrag bzw. den Anlass zur Erhebung, das Konzept mit Angaben über die praktische Durchführung und die Aufbereitung der Daten bis zum Stand der bereinigten Basisdaten dokumentieren. Im Grunde entspricht diese Datengeschichte einem heute allgemein akzeptierten Forschungsstandard, welcher fordert, dass Erkenntnisprozesse möglichst lückenlos rekonstruierbar und damit kritisierbar sein müssen (Intersubjektivität).<sup>27</sup> Wichtig ist auch, dass die Art und die Resultate allfälliger nachträglicher Qualitätsprüfungen etwa in Form von Nacherhebungen mit den Daten selber überliefert werden. Natürlich gehört auch die Zeit von der Erstellung der Basisdateien bis zur jeweiligen Wiederbenutzung zur Datengeschichte. Wichtige Eingriffe in die Dateien wie etwa die Übertragung in eine andere Struktur oder die Anpassung an neue Kategorienskalen bzw. Kodeverzeichnisse müssen ebenfalls dokumentiert werden, unabhängig davon, ob sie beim Datenproduzenten oder bereits im Archiv erfolgen. Wie detailliert die Dokumentation zu sein hat, ist jeweils an der Genauigkeit abzumessen, die bei statistischen Daten möglich und sinnvoll ist. In *Übersicht 1* sind die verschiedenen Elemente, die eine Dokumentation enthalten sollte, aufgelistet.

Wesentlich für die Archivierung sind auch die *technische Beschreibung und Geschichte des Datenträgers*. Die technische Beschreibung muss sicherstellen, dass die Daten überhaupt gelesen werden können, und die Geschichte des einzelnen Datenträgers soll dem Archivar in erster Linie erlauben, aufgrund von Herstellungsqualität, Alter und bisherigen Beanspruchungen die Zuverlässigkeit und die weitere Lebensdauer des Datenträgers abzuschätzen.

Das letzte der eingangs erwähnten Bewertungskriterien fordert, dass die Form der Daten und der Datenträger dem Benutzer den Zugriff möglichst leicht machen sollten. Um das Kriterium sinnvoll anwenden zu können, müssen die Benutzerbedürfnisse eingermassen bekannt sein. Ich gehe hier davon aus, dass sich diese auf zwei Typen reduzieren lassen:

1. Die «Grundlagenforschung», welche die Daten möglichst «von Grund auf» unter neuen Gesichtspunkten auswerten will und deshalb in erster Linie an Basisdaten interessiert ist.

2. Die Benutzer, welche sich nur einen raschen Überblick verschaffen wollen und in erster Linie an vorhandenen Ergebnissen interessiert sind.

Der «Grundlagenforscher» wird quantitative Untersuchungen heute nicht mehr ohne elektronische Datenverarbeitung durchführen; er wird praktisch darauf angewiesen sein, dass die Daten in maschinenlesbarer Form zur Verfügung stehen. Er wird weiter wünschen, dass die Daten einen möglichst hohen Informationsgehalt haben, möglichst gut dokumentiert sind und problemlos von Hard- und Software seines Rechenzentrums verarbeitet werden können.

Für das Archiv bedeutet dies, dass die verwendeten Archiv-Datenträger nicht nur möglichst zuverlässig und dauerhaft sein sollten, sondern auch durch möglichst viele Geräte lesbar sein müssen. Seit Jahrzehnten bereits entspricht das Magnetband diesen Anforderungen am besten. Auch die Daten sollten so strukturiert sein, dass sie von möglichst vielen Softwarepaketen eingelesen und verarbeitet werden können. Am vielseitigsten verwendbar sind die sogenannten «flachen Dateien», in denen die Daten als Tabelle organisiert sind. Jede Zeile in diesen Tabellen entspricht einer einzelnen statistischen Grundeinheit, und eine jede Spalte (bestimmte Position auf jeder Zeile) entspricht einem bestimmten Merkmal.

Ob nun für den erwähnten «Grundlagenforscher» Rohdaten oder bereinigte Basisdaten geeigneter sind, ist wohl eine Ermessensfrage. Ich nehme an, dass der spätere Benutzer eher wieder zu statistischen Daten greifen wird, wenn die Bereinigungsarbeit, die auch mit Hilfe der elektronischen Datenverarbeitung einen beträchtlichen Aufwand verursacht, bereits geleistet ist. Ausserdem bietet sich die Wahl nur noch selten, weil heute das Urmaterial aus Gründen des Datenschutzes häufig vernichtet werden muss. Der Archivar, der zwischen Urmaterial und bereinigten Daten entscheiden kann, sollte sich allerdings versichern, dass alle Angaben der Erhebungsbogen übernommen worden sind und dass aus der Begleitdokumentation hervorgeht, wie hoch der Anteil an ergänzten und korrigierten Werten ist.

Der Benutzer, der nur Überblicksdaten und Ergebnisse benötigt, wird sich an die Publikationen halten können und müssen. Sie enthalten

Daten auf hoher Aggregationsstufe mit Erläuterungen und Interpretationen. Sie sollten dauernd aufbewahrt werden, da sie einen raschen Überblick über die wesentlichen Ergebnisse einer Erhebung geben und auch wirkungsgeschichtlich von Interesse sind, weil statistische Erhebungen meist in dieser Form in einer breiteren Öffentlichkeit wirksam werden.

Die Ausführungen in diesem Kapitel legen nahe, die bereinigten Basisdaten und die Publikationen aller relevanten statistischen Erhebungen, insbesondere diejenigen der amtlichen Statistik, zur dauernden Aufbewahrung in die zuständigen Archive zu übernehmen.<sup>28</sup> Eine Archivierung sollte allerdings nur dann erfolgen, wenn jeder der oben genannten Faktoren in einem minimalen Ausmass zutrifft. Wenn die Daten ungenügend dokumentiert sind oder wenn sie nur mit sehr hohem Aufbereitungsaufwand überhaupt wieder benutzt werden können, dann erscheint eine Archivierung wenig sinnvoll.

## 5. Die Rolle der Archive

Nach diesen grundsätzlichen Überlegungen möchte ich nun einen Blick auf die konkrete Situation im Bereich der Archivierung statistischer Daten werfen, bevor ich anschliessend versuchen will, die Aufgabe der Archive in einem grösseren Umfeld zu bestimmen. Verallgemeinerungen sind hier nur sehr schwer möglich, da keine entsprechenden Erhebungen zur Verfügung stehen. Ich muss mich deshalb im folgenden auf einige Schlaglichter beschränken, die aber meines Erachtens doch in einem gewissen Masse repräsentativ sind.

Das *Schweizerische Bundesarchiv* beispielsweise bewahrt verschiedene Bestände mit statistischen Daten auf. Im Vergleich zu den Massen von Daten, welche die amtliche Statistik auf Bundesebene produziert hat, ist aber der Umfang dieser Überlieferung bescheiden (vgl. die Zusammenstellung in *Übersicht 2*). Der grösste Teil der älteren archivierten Daten sind Tabellen, also aggregierte Daten. Das dürfte vor allem damit zusammenhängen, dass sich die Wirtschafts- und Sozialgeschichte erst seit wenigen Jahrzehnten vermehrt auf quantitative Datenserien abstützt und dass vor der Einführung der elektronischen Datenverarbei-

tung die Massen von Rohdaten allein aufgrund des Aufwandes, den eine Sekundärauswertung verursacht hätte, als nicht archivwürdig angesehen wurden. Was die neueren Daten betrifft, so hat sich die Situation nun geändert. Die bereinigten Basisdaten der Statistik des jährlichen Bevölkerungsstandes werden mit zugehöriger Dokumentation regelmässig zur dauernden Aufbewahrung abgeliefert, und es laufen Bestrebungen, die Archivierung in weiteren Bereichen auf ähnliche Weise zu regeln.

Übersicht 2: **Archivbestände des Schweizerischen Bundesarchivs mit statistischen Unterlagen**<sup>29</sup>.

<b>E 3321</b>	<b>Bundesamt für Statistik</b>	
	Initiativen (Specimen)	1971-1991
	Referenden (Specimen)	1975-1989
	Initiativen, Referenden	1969-1980
	Nationalratswahlen, Referenden, Initiativen	1935-1971
	Nationalratswahlen	1971, 1975, 1979, 1983, 1987
	Amtsarchiv	1800-1950
	Familiennamenbuch der Schweiz	1939-1970
	Formulare der Eidg. Volkszählungen	1969-1989
	<i>Eidg. Volkszählung (Personen- und Stammrecords)</i>	1980
	Kontroll-Listen zur Volkszählung	1960, 1970, 1980
	<i>Statistik des jährlichen Bevölkerungsstandes</i>	1969-1989
	<i>(Wanderungen, Ausländer, Geborene, Gestorbene, Heiraten)</i>	
	Eidg. Betriebszählung	1905, 1929, 1939, 1955, 1965
	Kontroll-Listen für die Erhebung der Haushaltrechnungen von Unselbständigen und Rentnern	1989
	Pensionskassen-Statistik, Fragebogen	1988-1989
<b>E 7181 (A)</b>	<b>Bundesamt für Industrie, Gewerbe und Arbeit, Sozialstatistik</b>	
	Löhne und Einkommen	1969-1979
	Kleinhandelspreise, Lebenskosten, Mietpreise	1914-1973
	Fabrikstatistik	1929-1957
<b>E 7181 (C)</b>	<b>Bundesamt für Industrie, Gewerbe und Arbeit, Wirtschaft und Statistik</b>	
	Arbeitsvermittlung, Arbeitsmarkt (AVAM)	1985-1990
	Arbeitslosen-Informationssystem (ALIS)	1984-1990
	Gemeindearbeitslosigkeit (GAL)	1977-1990

Ein ähnliches Bild wie im Bundesarchiv ergibt sich im *Bundesamt für Statistik*, dem wichtigsten Datenproduzenten auf Bundesebene. Dieses bewahrt in seinem Amtsarchiv noch zahlreiche Daten auf, die zum Teil bis in die Anfänge unseres Jahrhunderts zurückreichen. Allerdings handelt es sich auch hier nur um Aggregationen in Tabellenform. Rohdaten (Urmaterial) oder bereinigte Einzeldaten, die älter als zwanzig Jahre sind, werden, soweit mir bekannt ist, nicht mehr aufbewahrt; sie dürften zum grössten Teil verloren sein. Selbst die Lochkarten, welche in der ersten Phase der elektronischen Datenverarbeitung in den 1960er Jahren erstellt wurden, sind vor nicht allzulanger Zeit vernichtet worden. Die jüngeren Daten aber dürften noch mehr oder weniger vollständig als bereinigte Basisdaten und zumeist auf maschinenlesbaren Datenträgern vorhanden sein. Allerdings sind auch diese Daten gefährdet, und zwar nicht so sehr durch die physische Vernichtung, als vielmehr durch den Verlust der zugehörigen Datenbeschreibung und Datengeschichte oder durch mangelhafte Nachführung der Datenbeschreibung bei der Übertragung der Daten in neue Strukturen. Die Ursachen für diese Gefährdung liegen wohl im starken Aktualitätsbezug der amtlichen Statistik. Sie hat den grössten Teil ihrer knappen Kräfte auf die Erarbeitung der von Verwaltung und Öffentlichkeit geforderten «neuesten Zahlen» zu konzentrieren, so dass den Daten, die bereits einige Jahre alt sind, nicht mehr genügend Aufmerksamkeit zuteil wird. Die Nachfrage nach diesen Basisdaten ist auch vergleichsweise klein, da die meisten Bedürfnisse mit den Angaben der wichtigsten Tabellen befriedigt werden können.

Seit 1987 bietet das Bundesamt für Statistik als zukunftsweisende Neuerung die Datenbank STATINF an.<sup>30</sup> Sie enthält einen grossen Teil der aggregierten statistischen Daten des Bundes und kann über öffentliche Datennetze im Direktzugriff abgefragt werden. Die Daten sind als zweidimensionale Tabellen strukturiert und können entweder direkt mit den in einer Methodenbank zur Verfügung stehenden Auswertungsalgorithmen bearbeitet oder aber über das Datennetz auf einen lokalen Rechner übertragen und mit eigener Software ausgewertet werden.

Auch von der nichtamtlichen Forschung an *Universitäten* und anderen, meist *privaten Instituten* werden in beträchtlichem Umfange wertvolle quantitative Daten erhoben, die grundsätzlich von dauerndem Interesse sind, aber keiner Abgabepflicht an staatliche Archive unterstehen. Das

Spektrum reicht von Marktforschungsunterlagen bis zu mittelalterlichen Preisreihen. Die Daten dürften in der Regel gewisse Zeit in den Instituten aufbewahrt werden; eine dauernde Archivierung ist aber höchst ungewiss, es sei denn, die Daten gelangen in ein Universitätsarchiv oder in eines der wenigen Institute, die sich auf die Archivierung von sozialwissenschaftlichen Daten spezialisiert haben, wie z.B. das Zentralarchiv für Empirische Sozialforschung an der Universität Köln oder das Steinmetz Archiv in den Niederlanden<sup>31</sup>. Im allgemeinen dürfte aber «die bei Wissenschaftlern und Technikern ausgeprägte <Wegwerfmentalität>, sobald Unterlagen nicht mehr der Bearbeitung unmittelbar anstehender Projekte dienen», zutreffen.<sup>32</sup> Dauernd wertvolle Grundlagen gehen dadurch verloren.

Unter diesen nichtamtlichen Datensammlungen sind die grösseren *historischen Datenbanken* besonders hervorzuheben. Ihre Daten werden nicht über direkte Erhebungen oder Messungen gewonnen, sondern durch methodische Auswertung von alten statistischen Unterlagen und von seriellen Quellen, die sich quantifizieren lassen. Bei dieser Auswertung fliesst sehr viel quellenkritische und interpretatorische Arbeit ein, die ihre methodische Grundlage aus der neu entstandenen Historischen Fachinformatik bezieht.<sup>33</sup> Sie hat auch den Begriff der <Metaquelle> geprägt, um damit eine zentrale, durch die elektronische Datenverarbeitung ermöglichte Neuerung zu bezeichnen, nämlich die Verknüpfung von verschiedenen Serien quantifizierbarer Quellen.<sup>34</sup> «Durch den Aufbau von Metaquellen wird eine zusätzliche, auf die Bedürfnisse des Forschers zugeschnittene Quellenbasis geschaffen, welche einen neuen, den Zeitgenossen in dieser Form unbekanntem (...) Teilaspekt der Realität abbildet.»<sup>35</sup> Diese historischen Datenbanken entstehen meist in Zusammenhang mit einem bestimmten Forschungsprojekt, das die Daten unter spezifischen Gesichtspunkten auswertet, aber darüber hinaus stehen die Daten für alle Interessierten zur Verfügung, zum Teil sogar im Direktzugriff über Datennetze wie etwa die recht umfassend angelegte Datenbank BERNHIST<sup>36</sup> oder die Historische Datenbank der Forschungsstelle für schweizerische Sozial- und Wirtschaftsgeschichte der Universität Zürich<sup>37</sup>. Das Problematische an dieser geschichtswissenschaftlichen «Halbfertigproduktion» ist der Umstand, dass für den Sekundärbenutzer der Quellenbezug weitgehend verloren ist, da es aus praktischen Gründen nicht möglich ist, die Metaquellen in der Datenbank derart detailliert zu beschreiben, dass

die Forderungen der Quellenkritik erfüllt werden können.<sup>38</sup> Wer die Daten mit anderen Forschungsfragen unter neuem Gesichtspunkt benutzen will, wird zumindest unsicher werden, wenn er die Aussagekraft der Zahlen für seine Frage beurteilen will. Die langen Zeitreihen, die oft aus ganz unterschiedlichen Quellentypen zusammengeführt werden, täuschen eine Homogenität vor, die leicht zu ungenauen oder gar falschen Interpretationen führen kann. Eine gründliche historische, d.h. unter anderem auch quellenkritische Arbeit wird sich auf jeden Fall eingehend mit der Beziehung zwischen den eigentlichen Quellen und den neuen Metaquellen befassen müssen. Historische Datenbanken sind zwar für viele Zwecke geeignet und für den Benutzer sehr praktisch, aber sie ersetzen die eigentlichen Quellen nicht. Ausserdem stellt sich hier wie bei den übrigen universitären oder privaten Forschungen dieselbe Frage nach dem langfristigen Schicksal der Daten. Die finanziellen und personellen Ressourcen der meisten dieser Unternehmen sind befristet und Schwankungen bzw. Wechseln ausgesetzt. Die Gefahr, dass dabei Daten oder auch nur Wissen über die Daten verloren gehen, ist recht gross.

Diese Ausführungen zeigen, dass einerseits die langfristige Sicherung eines grossen Teils der statistischen Grunddaten nicht gewährleistet ist und dass andererseits ein Bedarf nach aufbereiteten aggregierten Daten besteht, den viele Datenproduzenten in zunehmendem Masse mit Datenbanken befriedigen, die entweder im Direktzugriff über Daten-netze oder auf gängigen maschinenlesbaren Datenträgern benutzbar sind.

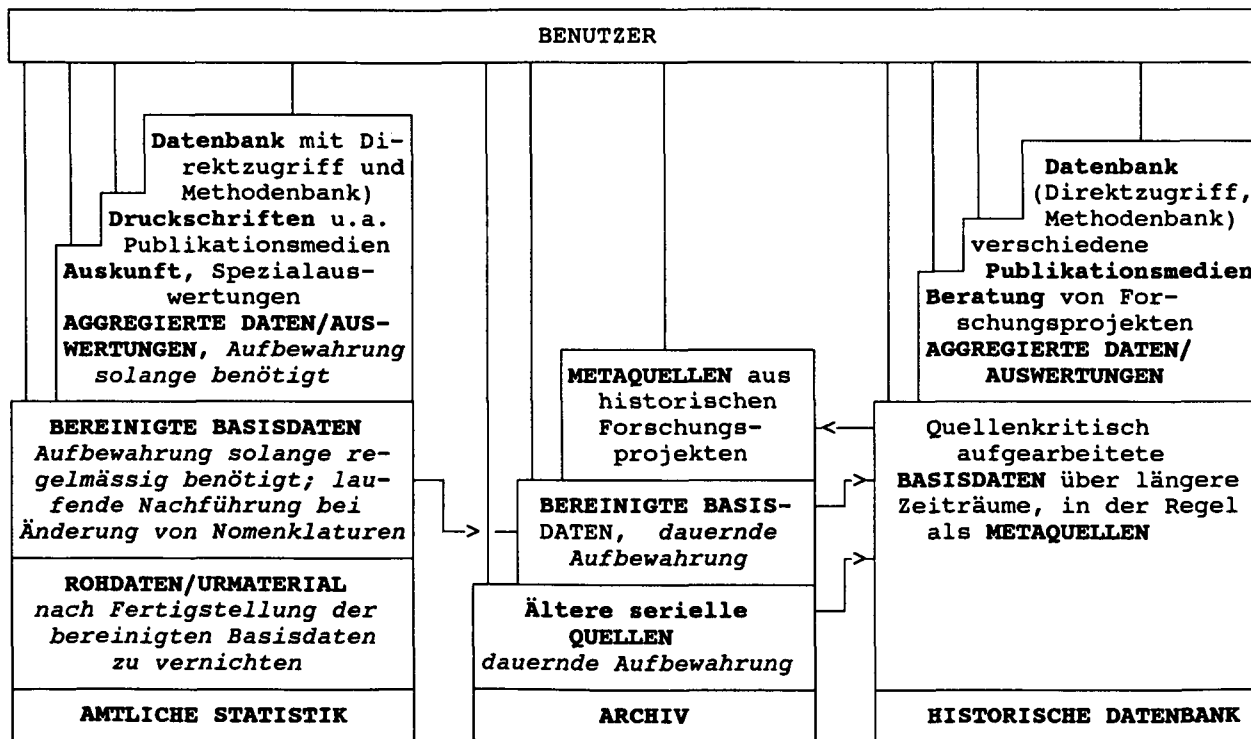
Für die öffentlichen Archive stellen sich in diesem Zusammenhang verschiedene strategische Fragen. Wie können sie die grundlegenden Daten der amtlichen Statistik in ihrem Zuständigkeitsbereich besser sichern? Sollen sie die Basisdaten von universitären oder privaten Forschungsprojekten ebenfalls übernehmen und dauernd aufbewahren? Müssten Archive ihre quantitativen Daten nicht ebenfalls aggregieren und als Datenbank für interessierte Benutzer anbieten? Sollen sie gar ihre Dienstleistungen soweit ausbauen wie viele auf quantitative Massendaten spezialisierte Archive, welche den Benutzern nicht nur die Quellen, sondern auch die nötigen Instrumente und die fachliche Beratung für die Auswertung und Interpretation dieser Quellen zur Verfügung stellen?

Die Antwort für staatliche Archive muss angesichts der Datenflut, die nicht nur im Bereich der statistischen Daten im Anwachsen ist, Prioritäten setzen und das Notwendige vom Wünschbaren trennen. Wie bereits in der Einleitung erwähnt, liegt der Kern der archivischen Aufgaben in der Sicherung und dauernden Aufbewahrung der wesentlichen Informationen der Gegenwart. Seit es Archive gibt, also seit Jahrhunderten, ist diese Aufgabe dieselbe geblieben. Da nun aber der dringendste Handlungsbedarf im Bereich der statistischen Daten gerade in der Sicherung derjenigen Daten mit dem grössten Informationsgehalt besteht, sollten sich Archive darauf konzentrieren und nicht versuchen, dem Historiker Auswertungsarbeit abzunehmen. In *Grafik 4* ist die Funktion der Archive bei der Aufbewahrung und Vermittlung von statistischen Daten neben derjenigen der amtlichen Statistik und der im Entstehen begriffenen historischen Datenbanken dargestellt.

Im Sinne einer Schlussfolgerung möchte ich diesen Beitrag mit den folgenden Vorschlägen für die künftige Zusammenarbeit zwischen amtlicher Statistik, historischen Datenbanken und öffentlichen Archiven abschliessen:

- Die von der amtlichen Statistik produzierten Daten müssen möglichst frühzeitig gesichert werden. Die Basisdaten sollten unmittelbar nach der Bereinigung kopiert und mit der nötigen Dokumentation versehen an das zuständige Archiv abgegeben werden. Der jeweilige Datenproduzent behält eine Kopie der Basisdaten als Grundlage für seine Auswertungen und führt sie laufend nach, wenn Änderungen der Verschlüsselung notwendig werden. Er bewahrt die Daten auf, solange er sie zur Herstellung von Zeitreihen benötigt.
- Die öffentlichen Archive übernehmen quellenkritisch aufgearbeitete Basisdaten oder Metaquellen aus historischen Forschungsprojekten, die entweder abgeschlossen sind oder ihre Daten nicht mehr benötigen, sowie aus historischen Datenbanken, deren Überleben gefährdet ist oder die aus einem anderen Grund ihre Basisdaten archivieren wollen.
- Die Archive unterstützen nach Möglichkeit alle Projekte, die statistische Daten quellenkritisch aufarbeiten und historische Datenbanken aufbauen. Sie sollten aber ihre knappen Kräfte nicht in den Aufbau eigener solcher Metaquellenbanken stecken.
- Wünschbar wäre ausserdem, in der Schweiz eine zentrale Nachweisstelle für maschinenlesbare quantitative Daten aufzubauen, damit die

Grafik 4: Die Stellung der Archive bei der Aufbewahrung und Vermittlung statistischer Daten



grossen Investitionen, die in die Erhebung und Aufbereitung von Daten fliessen, besser verwertet werden können. Archive könnten dabei zumindest eine Ressource, nämlich das Fachwissen über das Erschliessen und Aufbewahren von Daten und Akten, zur Verfügung stellen.

## Anmerkungen

- <sup>1</sup> Für wertvolle Anregungen und kritische Durchsicht des Manuskriptes bin ich meinen Kollegen Christoph Graf und Andreas Kellerhals zu Dank verpflichtet.
- <sup>2</sup> Vgl. Andreas Kellerhals, «*Weisst du wieviel Sternlein stehen...?*». *Die protostatistischen Erhebungen im Kanton Bern zwischen 1528 und 1831*, unveröffentlichte Lizentiatsarbeit Universität Bern 1984.
- <sup>3</sup> Das *Statistische Jahrbuch der Schweiz 1990*, hg. vom Bundesamt für Statistik, Zürich 1989, enthält die Kapitel: Bevölkerung; Raum, Landschaft und Umwelt; Erwerbsleben; volkswirtschaftliche Gesamtrechnungen; Preise; Produktion, Handel und Verbrauch; Land- und Forstwirtschaft; Energie; Bau- und Wohnungswesen; Tourismus; Verkehr und Nachrichtenwesen; Geldmenge, Finanzmärkte und Banken; Versicherungen; Gesundheit; Bildung und Wissenschaft; Kultur, Lebensbedingungen und Sport; Politik; öffentliche Finanzen; Rechtspflege.
- <sup>4</sup> Die Angaben sind entnommen aus: Christian Lerch: *Die EDV als Hilfsmittel der Statistik: Erwartungen aus der Sicht des Statistikers*, und Dominique Renaud: *Eine kohärente EDV-Politik: Stellungnahme aus dem Gesichtswinkel des Informatikers*, beide in: *Forum Statisticum*, Nr. 14, Februar 1981, S. 16 bzw. S. 31.
- <sup>5</sup> Ch. Lerch, *EDV als Hilfsmittel*, S. 16. Die zitierte Feststellung gilt selbst noch für das «Lochkarten-Zeitalter»: «Die Speicherkapazität der damaligen Computer war so gering, das neben dem Tabellierungsprogramm keine Daten gespeichert werden konnten. Praktisch jede Tabelle musste in einem separaten Durchlauf aller Lochkarten erstellt werden. Diesem Tabellierungslauf gingen zeitlich aufwendige, z.T. mehrfache Sortierläufe auf speziellen Sortiermaschinen voraus.» (Ebd.)
- <sup>6</sup> Vgl. Karl Heinrich Kaufhold: *Datenverarbeitung und Geschichtswissenschaft – Probleme und Aufgaben*. In: K. H. Kaufhold/J. Schneider (Hg.): *Geschichtswissenschaft und elektronische Datenverarbeitung*. Wiesbaden 1988, S. 9–17.
- <sup>7</sup> Der Zuständigkeitsbereich des Schweizerischen Bundesarchivs zum Beispiel umfasst hauptsächlich die Dienststellen der allgemeinen Bundesverwaltung, den Bundesrat und die Eidgenössischen Räte. Vgl. auch das *Reglement für das Schweizerische Bundesarchiv vom 15. 7. 1966 (Stand vom 1. 1. 1982)*, in: *Systematische Sammlung des Bundesrechts (SR)*, 432.11.
- <sup>8</sup> Vgl. v.a. die Hinweise in: Wolf Buchmann: *Archive und elektronische Datenverarbeitung. Ein Diskussionsbeitrag zu den Folgen der Einführung einer neuen Technologie für die Archive*, in: F. Kahlenberg (Hg.): *Aus der Arbeit der Archive. Festschrift Hans Booms*, Boppard/Rh. 1989, S. 243–256.
- <sup>9</sup> Friedrich Kahlenberg: *Informationsbankensysteme ohne Archive? Bemerkungen zum Funktionswandel öffentlicher Archive*, in: *Archivalische Zeitschrift* 68, 1972, S. 125–133; Zitat auf S. 132.

- <sup>10</sup> Vgl. etwa die Darstellungen und Übersichten in der Zeitschrift *Historical Social Research – Historische Sozialforschung*.
- <sup>11</sup> Vgl. dazu v.a. die Zeitschrift *History and Computing*, 1989ff.
- <sup>12</sup> Die nachfolgenden Ausführungen stützen sich hauptsächlich auf Günter Buttler/Reinhold Stroh: *Einführung in die Statistik*. Reinbek b. Hamburg 1980; Werner Haug, *Bevölkerungsstatistik*, in: Peter Bohley/Armin Jans (Hg.): *Einführung in die Wirtschafts- und Sozialstatistik der Schweiz*. Bern und Stuttgart 1990, sowie auf Jürgen Bortz: *Lehrbuch der Statistik für Sozialwissenschaftler*. 2. Aufl., Berlin u. a. 1985.
- <sup>13</sup> *Verordnung über die Eidgenössische Volkszählung 1990 vom 26. Okt. 1988 (SR 431.112.1)*, Art. 1, Abs. 1.
- <sup>14</sup> Wilhelm Bickel zitiert in seiner *Bevölkerungsgeschichte und Bevölkerungspolitik der Schweiz seit dem Ausgang des Mittelalters*, Zürich 1947, S. 27, den Landvogt von Saanen, der im Jahre 1653 ein Bevölkerungsverzeichnis erstellen wollte: «Diese Nachforschung hat ungleiche Gedanken und gefährliche Einbildung bei den Untertanen verursacht, ja soweit, dass ein gemein Geschrei allhier ausgebreitet worden, solche Verzeichnis geschehe darum, dass Ihr Gnädigen Herren einer jeden Haushaltung 6 Pfund Kontribution aufzulegen Vorhabens seien.» Vgl. auch Christian Simon: *Hintergründe bevölkerungsstatistischer Erhebungen in Schweizer Städteorten des 18. Jahrhunderts. Zur Geschichte des demographischen Interesses*, in: *Schweizerische Zeitschrift für Geschichte* 34, 1984, S. 186–205.
- <sup>15</sup> Vgl. Dieter Koch: *Preisstatistik*. In: P. Bohley/A. Jans, *Einführung*, S. 283.
- <sup>16</sup> Vgl. Armin Jans: *Volkswirtschaftliche Gesamtrechnung*. In: P. Bohley/A. Jans, *Einführung*, S. 323–375.
- <sup>17</sup> Vgl. etwa die Hinweise in den *Statistischen Jahrbüchern*.
- <sup>18</sup> Hans Booms: *Gesellschaftsordnung und Überlieferungsbildung*. In: *Archivalische Zeitschrift* 68, 1972, S. 3–40, Zitat auf S. 8.
- <sup>19</sup> Jörn Rüsen: *Rekonstruktion der Vergangenheit. Grundzüge einer Historik II: Die Prinzipien der historischen Forschung*, Göttingen 1986, S. 107ff.
- <sup>20</sup> Vgl. Herbert Reinke: *Datenbeschreibung und Datendokumentation in der historischen Sozialforschung. Problemaufriss und Empfehlungen für die Forschung*, in: *Historical Social Research – Historische Sozialforschung* 24, 1982, S. 4–24. Er schreibt über Umfragen der frühen 50er Jahre: «Die Bedeutung einzelner Variablen ist uns durchaus nicht mehr verständlich, entspricht doch das Bezugssystem der damaligen Zeit nicht mehr unserem. Die Alltagskonnotationen einer Fragebogenformulierung aus den 50er Jahren sind uns nicht mehr bekannt. Je weiter entfernt dieses Bezugssystem ist, um so unabdingbarer sind Kommentare, die Variablen verständlich machen.» (S. 12)

- <sup>21</sup> Werner Lüönd: *Die Volkszählungen des 19. Jahrhunderts im Kanton Zug*. In: *Tugium* 1990, S. 70–96.
- <sup>22</sup> Ebd. S. 70.
- <sup>23</sup> Vgl. dazu Alain Desrosières: *Comment faire des choses qui tiennent: histoire sociale et statistique*. In: *Histoire & Mesure*, 1989, S. 225–242.
- <sup>24</sup> Die Anonymisierung ist für die Forschung in der Regel kein Verlust, da statistische Daten aus Interesse an quantitativen Verhältnissen benutzt werden und nicht aus Interesse an bestimmten Einzelpersonen.
- <sup>25</sup> W. Lüönd, *Volkszählungen*, S. 71.
- <sup>26</sup> Die Grundgesamtheit umfasst «alle potentiell untersuchbaren Einheiten, die ein gemeinsames Merkmal (oder eine gemeinsame Merkmalskombination) aufweisen». Bortz, *Lehrbuch*, S. 111.
- <sup>27</sup> Heinrich Best/Wilhelm H. Schröder: *Quantitative historische Sozialforschung*. In: Ch. Meier/J. Rösen (Hg.): *Historische Methode*. München 1988, S. 246.
- <sup>28</sup> Die Archivierung der nicht aggregierten Grunddaten wird auch von amerikanischen Autoren gefordert. Vgl. Charles Dollar: *Appraising, machine-readable records*. In: *American Archivist* 41, 1978, S. 424, und Meyer H. Fishbein: *Reflections on appraising statistical records*. In: *American Archivist* 50, 1987, S. 229.
- <sup>29</sup> Eine Übersicht über die mit Statistikaufgaben betrauten Dienststellen der Eidgenössischen Bundesverwaltung gibt W. Haug, *Bevölkerungsstatistik*, S. 4.
- <sup>30</sup> Bundesamt für Statistik: *Was ist STATINF?* Bern 1986 (Materialien zur Statistik. Amtliche Statistik der Schweiz Nr. 68).
- <sup>31</sup> Vgl. *Historical Social Research – Historische Sozialforschung* 14, 1989, S. 118–121: «The Steinmetz Archive is a data archive which collects the basic data of (predominantly) Dutch social science research. By basic data we mean the (anonymous) «raw data» from quantitative empirical research or statistical administrations, as recorded in machine-readable form (on magnetic tapes, disks, punched cards, or other media).»
- <sup>32</sup> Vgl. Helmuth Trischler: *Aus der Sicht eines wissenschaftlichen Archivbenützers. Referat am 61. Deutschen Archivtag, Arbeitssitzung zum Thema «Die naturwissenschaftlich-technische Überlieferung in der Bundesrepublik Deutschland – Probleme ihrer archivischen Sicherung*. In: *Der Archivar* 44, 1991, Sp. 68–73.
- <sup>33</sup> Manfred Thaller: *Gibt es eine fachspezifische Datenverarbeitung in den historischen Wissenschaften?* In: K. H. Kaufhold/J. Schneider (Hg.): *Geschichtswissenschaft und elektronische Datenverarbeitung*. Wiesbaden 1988, S. 45–83, sowie Konrad Jarausch/Gerhard Arminger/Manfred Thaller: *Quantitative Methoden in der Geschichtswissen-*

schaft. *Eine Einführung in die Forschung, Datenverarbeitung und Statistik*, Darmstadt 1985.

- <sup>34</sup> Vgl. dazu Peter Becker: *Formen und Möglichkeiten der Standardisierung bei Metaquellen*. In: F. Hausmann u. a. (Hg.): *Datennetze für die historischen Wissenschaften?* Graz 1987, S. 18–27.
- <sup>35</sup> Christian Pfister, unter Mitarbeit von Hannes Schüle: *Metaquellen als Grundlagen zur Abgrenzung und Typisierung historischer Agrarzonen. Das Beispiel des Kantons Bern im späten 18. und 19. Jahrhundert*, In: *Itinera* 10, 1989, S. 28–57, Zitat auf S. 29.
- <sup>36</sup> Vgl. bspw. Christian Pfister/Hannes Schüle: *Encompassing «Geo-Histoire». Methodological dimensions and historiographical implications of the «BERNHIST» interdisciplinary information system*, in: R. van der Voort (Hg.): *The Tindberg Volume*. Amsterdam 1990.
- <sup>37</sup> Kurzbeschreibung in: P. Bohley/A. Jans, *Einführung*, S. 385.
- <sup>38</sup> P. Becker, *Formen*, gibt eine umfangreiche Liste dieser Anforderungen.

## Résumé

Dès les années 1960, l'emploi de l'informatique s'est répandu dans les administrations publiques et les milieux de la recherche pour le traitement des données quantitatives de masse. Le phénomène a non seulement accru le volume des données mais il a aussi modifié la relation avec elles. Les Archives doivent tenir compte de ce changement et réfléchir à leur propre mode de procéder. La première partie de cette contribution essaie de définir des lignes directrices opérationnelles pour le traitement archivistique des données statistiques, à partir d'une analyse de leur processus de production et de leur spécificité. Dans cette démarche, il convient aussi de prévoir le changement que subissent les données lorsqu'elles deviennent sources historiques et objet de recherches historiques. Le résultat de ces réflexions montre, d'une part, que c'est un stade bien défini du traitement des données, à savoir les données de base mises au net, qui se prête le mieux à l'archivage, et, d'autre part, que la valeur d'information des données ne peut être conservée de manière durable que si des exigences minima assez considérables sont satisfaites relativement à la documentation qui doit accompagner leur versement aux archives.

La deuxième partie de l'article aborde la question de savoir si les Archives ne doivent pas également réexaminer leur rôle sur le «marché de l'information», face au développement du traitement électronique des données et de leur communication. Dans le domaine des données quantitatives de masse, tant l'administration publique que les Universités, ou les institutions qui leur sont proches, ont commencé à constituer des banques de données comprenant des séries de données remontant loin dans le temps et à les offrir à travers des réseaux publics de données ou sur des supports pour micro-ordinateurs. Ce phénomène se passe largement en dehors de la participation des Archives et sous une forme qui, bien que satisfaisant pleinement l'utilisateur actuel, laisse craindre une diminution croissante de leur valeur d'information par rapport à l'idéal que l'on voudrait atteindre avec les lignes directrices, dont il est question dans la première partie de cet article. Il est absolument nécessaire que là aussi les Archives maintiennent leur rôle traditionnel séculaire de conservatrices du patrimoine et assurent, en commun avec les producteurs de données et les utilisateurs, la pérennité d'informations de valeur.

## Compendio

Dagli anni Sessanta nell'amministrazione pubblica e nella ricerca si ricorre all'elaborazione elettronica dei dati che ha comportato non solo un aumento quantitativo dei medesimi ma anche modifiche nel loro trattamento. Una svolta, questa, che gli archivi devono prendere in considerazione rivedendo le proprie procedure. La prima parte del presente contributo tenta di delineare alcuni criteri adeguati per il lavoro archivistico con dati statistici, in base ad un'analisi del loro processo di produzione e della loro specificità. In tale contesto va previsto anche il mutamento cui i dati sono soggetti quando diventano fonte ed oggetto di ricerca storica. Da queste considerazioni emerge, da un lato, che i più idonei all'archiviazione sono i dati sottoposti ad un determinato grado d'elaborazione, segnatamente i dati depurati di base, d'altro lato appare chiaro che una conservazione a lungo termine del contenuto informativo dei dati è possibile solamente se riguardo alla loro documentazione viene soddisfatto un insieme relativamente ampio d'esigenze basilari.

La seconda parte affronta l'interrogativo se gli archivi debbano rivedere sostanzialmente anche il proprio ruolo nel «mercato delle informazioni» considerato lo sviluppo nell'elaborazione e comunicazione elettronica dei dati. Nel settore dei dati quantitativi di massa sia l'amministrazione pubblica che le università o istituzioni affini hanno iniziato ad allestire banche di dati cronologicamente lontani e a metterle a disposizione attraverso reti pubbliche di trasmissione o supporti per microcomputer. Ciò avviene senza la partecipazione degli archivi e in una forma che, seppur molto pratica per l'odierno utente, rischia, alla luce dei criteri emersi nella prima parte del presente contributo, di portare ad un crescente impoverimento del contenuto informativo dei dati. E quindi assolutamente indispensabile che anche in quest'ambito gli archivi restino fedeli al loro secolare ruolo di custodi della tradizione ed assicurino assieme a produttori ed utenti di dati una duratura salvaguardia di preziose informazioni.



## Archivierung von statistischen Daten

In	Studien und Quellen
Dans	Etudes et Sources
In	Studi e Fonti
Jahr	1991
Année	
Anno	
Band	16-17
Volume	
Volume	
Autor	Bütikofer, Niklaus
Auteur	
Autore	
Seite	263-295
Page	
Pagina	
Ref. No	80 000 103

Das Dokument wurde durch das Schweizerische Bundesarchiv digitalisiert.

Le document a été digitalisé par les Archives Fédérales Suisses.

Il documento è stato digitalizzato dell'Archivio federale svizzero.