

Die Archivierung elektronischer Informationen am Beispiel statistischer Daten

Autor(en): **Bütikofer, Niklaus**

Objektyp: **Article**

Zeitschrift: **Geschichte und Informatik = Histoire et informatique**

Band (Jahr): **2 (1991)**

PDF erstellt am: **22.07.2024**

Persistenter Link: <https://doi.org/10.5169/seals-4259>

Nutzungsbedingungen

Die ETH-Bibliothek ist Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Inhalten der Zeitschriften. Die Rechte liegen in der Regel bei den Herausgebern.

Die auf der Plattform e-periodica veröffentlichten Dokumente stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Einzelne Dateien oder Ausdrucke aus diesem Angebot können zusammen mit diesen Nutzungsbedingungen und den korrekten Herkunftsbezeichnungen weitergegeben werden.

Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die systematische Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber.

Haftungsausschluss

Alle Angaben erfolgen ohne Gewähr für Vollständigkeit oder Richtigkeit. Es wird keine Haftung übernommen für Schäden durch die Verwendung von Informationen aus diesem Online-Angebot oder durch das Fehlen von Informationen. Dies gilt auch für Inhalte Dritter, die über dieses Angebot zugänglich sind.

Die Archivierung elektronischer Informationen am Beispiel statistischer Daten*

Niklaus Bütikofer, Schweizerisches Bundesarchiv, Bern

1. Einleitung

Die Statistik gehört zusammen mit der Buchhaltung zu den ersten Bereichen, in denen die Verwaltung die elektronische Datenverarbeitung (EDV) erfolgreich einsetzte. Die Erhebungsdaten der Volkszählung 1960 wurden beispielsweise erstmals auf Lochkarten erfasst und mit elektronischen Rechnern ausgewertet. Desgleichen ist die Bearbeitung quantitativer Daten der erste Bereich, in dem die EDV Einzug in die Geschichtswissenschaften gehalten hat. Schnell haben einige wenige Historiker erkannt, dass man mit dem neuen Werkzeug Computer viele Massenquellen, die bisher wegen ihres Umfangs als unbearbeitbar galten, nun mit sinnvollem Aufwand auswerten konnte. Der Computer mit seiner neuen Rechenkraft hat ausserdem die Weiterentwicklung quantitativer Methoden in starkem Masse vorangetrieben, wovon Statistiker und Historiker gemeinsam profitiert haben.

Die Archive, die gleichsam die Vermittlungsstelle zwischen amtlicher Statistik und Geschichtswissenschaften sind, haben mit etwas Verspätung von dieser Entwicklung Kenntnis genommen. Dabei galt ihre Aufmerksamkeit in Sachen EDV zuallererst ihren eigenen Automatisierungsbedürfnissen, die meist bei ihren umfangreichen Findmitteln lagen. Erst in einer zweiten Phase bemerkten die Archivare, dass die Verwaltungen das künftige Archivgut in laufend grösserem Umfang nicht mehr als Papierakten, sondern als elektronische Daten auf verschiedenen, ständig ändernden Datenträgern produzieren und ablegen. Die Schwierigkeiten, aber auch die neuen Möglichkeiten, welche diese Entwicklung den Archiven bietet, möchte ich in diesem Beitrag am Beispiel der statistischen Daten schlaglichtartig aufzeigen. Ich gehe dabei von den allgemeinen Aufgaben eines Archivs aus und entwickle daraus die wesentlichen Züge eines Archivierungskonzepts für Daten der amtlichen Statistik.

* Dieser Beitrag basiert im wesentlichen auf dem Aufsatz: Niklaus Bütikofer: Archivierung von statistischen Daten. In: Studien und Quellen, Nr. 16/17, hg. vom Schweizerischen Bundesarchiv, Bern 1991, S. 263-294. Einzelne Textstellen und Grafiken sind unverändert übernommen worden. Ausführliche Literaturhinweise finden sich ebenfalls in diesem Aufsatz.

2. Die Aufgabe der Archive

Ich setze voraus, dass die Aufgaben des Archivs von einer zunehmend informatisierten Umgebung in ihrem Kern nicht beeinträchtigt und verändert werden, weil es dazu auch in Zukunft kaum Alternativen geben wird. Nur professionelle Archive können meines Erachtens die Aufbewahrung und den Unterhalt von Akten und Daten über sehr lange Zeiträume hinweg einigermassen garantieren.

Die Aufgaben eines Archivs lassen sich in der folgenden Kurzformel zusammenfassen: Ein Archiv sichert, erschliesst, konserviert und vermittelt den organisch gewachsenen dokumentarischen Niederschlag der Tätigkeit derjenigen in der Regel staatlichen Dienststellen, denen es zugeordnet ist.

Die Dienststellen, denen das Schweizerische Bundesarchiv zugeordnet ist, sind hauptsächlich die Eidgenössischen Räte, der Bundesrat und die allgemeine Bundesverwaltung.

Zwei Elemente sind mir an der eben angeführten Definition wichtig: "dokumentarischer Niederschlag" und "Tätigkeit". Die Tätigkeit der Verwaltung hinterlässt gemäss diesen Elementen deutliche Spuren, die geeignet sind, diese Tätigkeit zu dokumentieren. Daran schliesst sich sofort die Frage, für welche Zwecke das Verwaltungshandeln dokumentiert werden muss. Die Dienststellen tun dies nicht etwa für die Archive, sondern in erster Linie aufgrund der folgenden eigenen Bedürfnisse:

1. Jede Verwaltungsstelle ist rechenschaftspflichtig sowohl vor ihren vorgesetzten Instanzen und der Öffentlichkeit als auch - im Streitfalle - vor Gerichten. Zu diesen Zwecken müssen die jeweiligen Handlungsvoraussetzungen, der Handlungsrahmen, die Handlungsintention, der Handlungsablauf und das Handlungsergebnis in ihren wesentlichen Zügen rekonstruierbar sein.
2. Alle Normen setzenden Entscheide (vom Gesetz bis zur einzelnen Verfügung) müssen aus Gründen der Rechtssicherheit mindestens bis zu ihrem Widerruf aufbewahrt werden.
3. Jede Verwaltungsstelle (natürlich nicht nur Verwaltungsstellen, sondern jeder, der in irgend einem Gebiet zielgerichtet arbeiten will) hat ein Interesse daran, auf früher ausgeführte Handlungen abzustellen, sei es als Präzedenzfall, der im Sinne der Rechtsgleichheit für alle nachfolgenden vergleichbaren Fälle massgebend ist, oder sei es als Voraussetzung und Arbeitsvorlage für das weitere Handeln.

Dieses Interesse der Verwaltung an von ihr selber oder von anderen

Dienststellen produzierten Unterlagen kann sich auf recht weit zurückliegende Zeiten richten. Es ist zudem klar, dass der "dokumentarische Niederschlag" grundsätzlich aus Schrift, Bild und Ton auf allen möglichen Trägermedien bestehen kann.

Erst im nachhinein kommen schliesslich historische Interessen ins Spiel. Wissenschaften historischer oder anderer Richtungen können verschiedene Aspekte des Verwaltungshandelns zu ihrem Gegenstand machen. Zu diesem Zwecke müssen sie natürlich die Tätigkeit und die Grundlagen der Verwaltung rekonstruieren können. Das Archiv, welches die spätere historisch-wissenschaftliche Auswertung vermittelt, ist gewissermaßen darauf angewiesen, dass die Dienststellen ihre Akten und Daten für ihre eigenen vorrangigen Zwecke ablegen, d.h. aufbewahren. Das Archiv kann eben nur diejenigen Akten und Daten für die Langzeitarchivierung sichern, welche in der Dienststelle nicht sofort nach unmittelbarem Gebrauch vernichtet worden sind.

Im elektronischen Bereich birgt genau diese Situation die grosse Gefahr in sich, dass durch voreiliges Löschen von Daten riesige Lücken in der Überlieferung entstehen. Um hier etwas Abhilfe zu schaffen, versucht das Bundesarchiv seit einiger Zeit bereits, durch Beratung und Erlass von Weisungen Einfluss auf die Schriftgutverwaltung und auf die Konzipierung von EDV-Systemen zu nehmen, getreu dem Grundsatz, dass man im EDV-Bereich den Lebenszyklus der Daten vor ihrer Produktion festlegen sollte. Die jüngsten statistischen Daten im Schweizerischen Bundesarchiv betreffen denn auch bereits das Jahr 1990 (vgl. Übersicht 2 im letzten Teil dieses Beitrages).

3. Archivierung statistischer Daten

Die Grundfrage für den Archivar lautet: Wie ist die Arbeit der amtlichen Statistik langfristig zu dokumentieren, d.h. welche Akten und Daten sind in welcher Form und auf welchem Datenträger zu archivieren, damit sowohl derjenige, der sich für die Geschichte der Statistik interessiert, als auch derjenige, der die Statistiken selber als zentrale Daten über die sozio-ökonomische Wirklichkeit einer bestimmten Zeit auswerten will, eine optimale Quellenlage vorfinden werden?

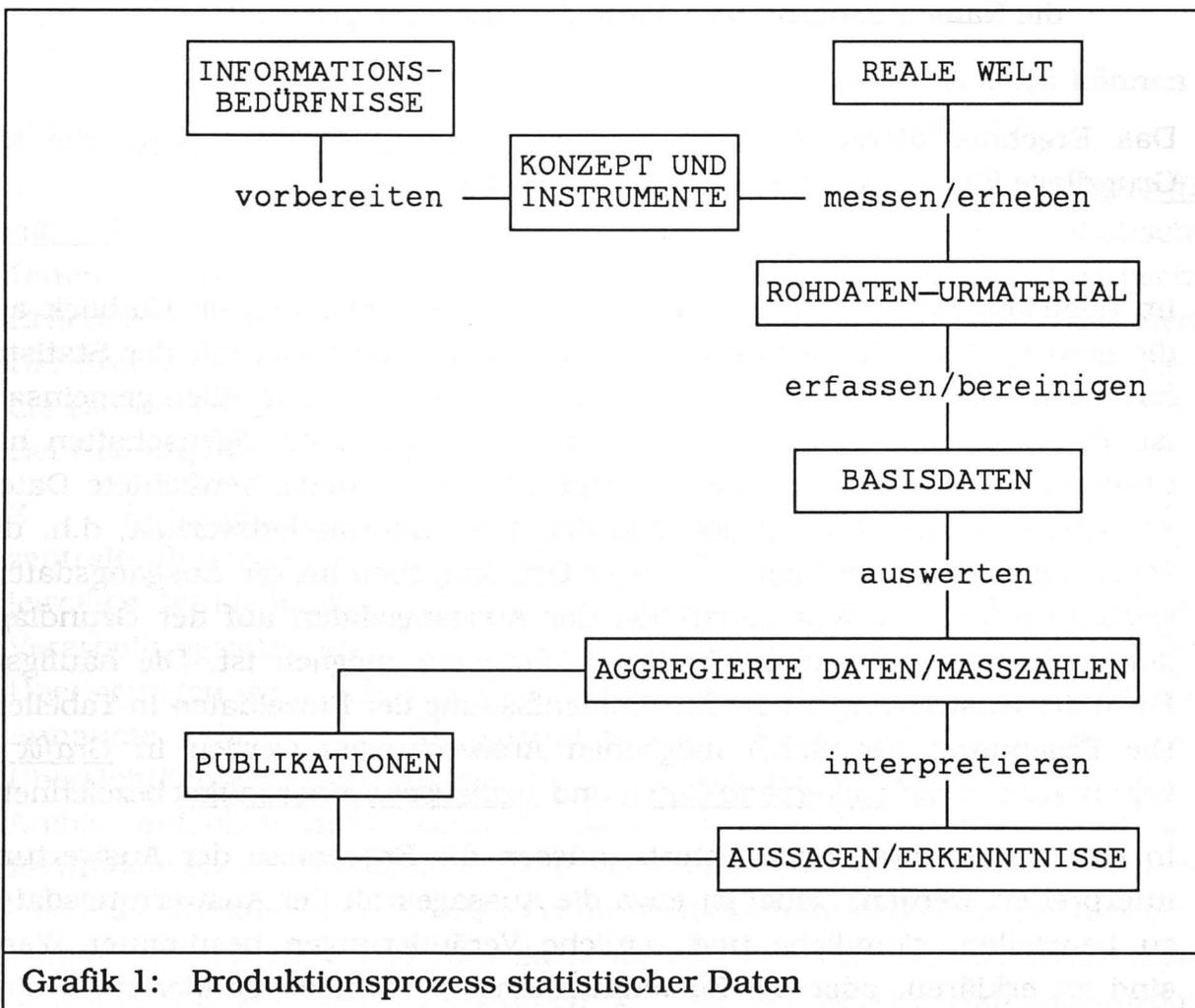
Ausgangspunkt der Überlegungen muss die Verwaltungshandlung sein, in deren Verlauf verschiedene Materialien als "dokumentarischer Niederschlag" anfallen. In meinem Beispiel hier ist es die Produktion von Statistiken.

3.1 Der Produktionsprozess statistischer Daten

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.

Jede amtliche statistische Erhebung beruht auf einer gesetzlichen Grundlage, die ihrerseits auf bestimmte Informationsbedürfnisse zurückgeht. Die Statistik-Produzenten konkretisieren ihren Auftrag in einem Erhebungskonzept, welches genau angibt, was wie gezählt und ausgewertet werden soll. Wer jemals alte Zählungen ausgewertet hat, weiss um die Schwierigkeiten, welche die Erhebungs-Kategorien dem quellenkritischen Historiker bieten. Im nachhinein ist es in der Regel schwierig, zu ermitteln, welche konkreten Erscheinungen mit den jeweiligen Begriffen erfasst wurden, ganz zu schweigen von den oft unbemerkten Veränderungen im Bedeutungsgehalt einzelner Worte über eine bestimmte Zeit hinweg.



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:

- Sie werden verschlüsselt, d.h. individuelle, umgangssprachliche Angaben müssen vereinheitlicht und allenfalls in eine feste vorgegebene Nomenklatur und einen alfanumerischen Kode übertragen werden.
- Sie werden kontrolliert, d.h. sie werden mittels bestimmter Verfahren auf Vollständigkeit und eindeutige Fehler überprüft (Plausibilitätstests).
- Sie werden korrigiert und ergänzt, d.h. die festgestellten Lücken und Fehler werden entweder über gezielte Nacherhebungen oder mittels bestimmter Schätzverfahren behoben.
- Sie werden anonymisiert, d.h. aus personenbezogenen Daten werden die Namen entfernt, was heute der Regel entspricht.

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

Im nächsten Bearbeitungsschritt 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. 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. 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.

3.2 Die Grundsätze der Archivierung

Wie sollte nun die Überlieferung von statistischen Unterlagen beschaffen sein, damit der zukünftige Historiker daraus möglichst korrekt und einfach sowohl die Geschichte der amtlichen Statistik als auch die wesentlichen Züge unserer Gegenwart, die heute statistisch erfasst wird, rekonstruieren kann? Zur Auswahl für die dauernde Aufbewahrung steht grundsätzlich alles Material, das gemäss dem Schema von Grafik 1 im Laufe des statistischen Produktionsprozesses anfällt.

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. Das heisst in bezug auf statistische Daten nichts anderes, als dass es auch völlig belanglose statistische Erhebungen gibt, die es nicht wert sind, langfristig aufbewahrt zu werden. Der Archivar wird sich aber weitgehend auf die Regel verlassen können, dass die Daten aus der amtlichen Statistik generell archivwürdig sind, sofern die Beurteilung nach den folgenden Kriterien ebenfalls positiv ausfällt.

2. Der Gehalt an Informationen über die jeweilige Wirklichkeit. Das zentrale Bewertungskriterium ist der Gehalt an Informationen 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.

Weiter oben ist bereits 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 Rechengerät 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. 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.

3. Das Vorhandensein ausreichender Dokumentation über die Quelle und ihren Kontext ist ein weiteres wichtiges Kriterium, das über den Wert einer Quelle entscheidet. Die Dokumentation muss ermöglichen, den Informationsgehalt der Quelle vollumfänglich zu erschliessen und kritisch zu beurteilen. 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 ganz zu verzichten, als sie durch ausgeklügelte Bearbeitung zu retten. Die notwendige Dokumentation statistischer Daten besteht aus einer Dateibeschreibung, welche erst eigentlich die Bedeutung der Einzelwerte konstituiert, aus einer Datengeschichte, anhand welcher er Aussagewert der Daten beurteilt werden kann, sowie aus einer Beschreibung und Geschichte des Datenträgers und aus Hinweisen auf Möglichkeiten und Bedingungen der Benutzung (vgl. Übersicht 1).

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). 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.

4. Die Form und das Speichermedium der Quelle ist der letzte wesentliche Faktor, welcher bei der archivischen Überlieferung statistischer Daten zu berücksichtigen ist. Form der Quelle und Speichermedium sollten beide so beschaffen sein, dass sich der Forscher mit möglichst wenig Aufbereitungsaufwand Zugang zu den Informationen verschaffen kann. Wesentlich für die Archivierung sind die technische Beschreibung und die 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 ermöglichen, aufgrund von Herstellungsqualität, Alter und bisherigen

Beanspruchungen die Zuverlässigkeit und die weitere Lebensdauer des Datenträgers abzuschätzen.

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 Befragter 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 vorliegen, sollten sie der Dokumentation beigelegt werden. Allfällige Daten 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 Herstellungsdatum**
- 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.

Das letzte der eingangs erwähnten Bewertungskriterien fordert, dass der Datenträger und die Form der Daten dem Benutzer den Zugriff möglichst leicht machen sollten. Um das Kriterium sinnvoll anwenden zu können, müssen die Benutzerbedürfnisse einigermaßen 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 (eine 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 mit zugehöriger ausführlicher Dokumentation und die Publikationen aller relevanten statistischen Erhebungen, insbesondere derjenigen der amtlichen Statistik, zur dauernden Aufbewahrung in die zuständigen Archive zu übernehmen. 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.

3.3 *Der Stand der Archivierung auf Bundesebene*

Nach diesen grundsätzlichen Überlegungen möchte ich nun einen kurzen Blick auf die konkrete Situation im Bereich der Archivierung statistischer Daten werfen. 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 Datenverarbeitung 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 (Die kursorische Übersicht umfasst nur die beiden wichtigsten Ämter unter den vielen Statistik-Produzenten der Schweizerischen Bundesverwaltung. Unterlagen auf maschinenlesbaren Datenträgern sind kursiv gesetzt.)

E 3321 (-) Bundesamt für Statistik

Initiativen und Referenda	1969-1991
Nationalratswahlen, Referenda, Initiativen	1935-1969
Nationalratswahlen	1966-1987
<i>Eidg. Volkszählung (Personen- und Stammrecords)</i>	1980
Formulare der Eidg. Volkszählungen	1970-1980
Kontroll-Listen zu Volkszählungen	1960-1980
<i>Statistik des jährlichen Bevölkerungsstandes (Wanderungen, Ausländer, Geborene, Gestorbene, Heiraten)</i>	1969-1990
Amtsarchivakten	1800-1950
Familiennamenbuch der Schweiz	1939-1970
Eidg. Betriebszählungen	1905, 1929, 1939, 1955, 1965
<i>Eidg. Betriebszählungen</i>	1975, 1985
Kontroll-Listen für die Erhebung der Haushaltrechnungen von Unselbständigen und Rentnern	1989
Pensionskassen-Statistik (Fragebogen)	1988-1989
	==>>

**E 7181 (A) Bundesamt für Industrie, Gewerbe und Arbeit,
Sozialstatistik**

Löhne und Einkommen	1969-1979
Kleinhandelspreise und Lebenskosten, Mietpreise	1914-1973
Fabrikstatistik	1929-1957

**E 7182 (-) Bundesamt für Industrie, Gewerbe und Arbeit,
wirtschafts- und sozialstatistische Unterlagen**

Oktoberloohnerhebungen	1977-1990
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 Amtsbücherei 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.