

## Datenbeschreibung und Datendokumentation in der historischen Sozialforschung: Problembeschreibung und Empfehlungen für die Forschung

Reinke, Herbert

Veröffentlichungsversion / Published Version

Sammelwerksbeitrag / collection article

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

GESIS - Leibniz-Institut für Sozialwissenschaften

### Empfohlene Zitierung / Suggested Citation:

Reinke, H. (1986). Datenbeschreibung und Datendokumentation in der historischen Sozialforschung: Problembeschreibung und Empfehlungen für die Forschung. In M. Thaller (Hrsg.), *Datenbanken und Datenverwaltungssysteme als Werkzeuge historischer Forschung* (S. 125-143). St. Katharinen: Scripta Mercaturae Verl. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-340316>

### Nutzungsbedingungen:

*Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.*

*Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.*

### Terms of use:

*This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.*

*By using this particular document, you accept the above-stated conditions of use.*

**Datenbeschreibung und Datendokumentation  
in der historischen Sozialforschung.**

**Problembeschreibung und Empfehlungen für die Forschung.**

**1. Vorbemerkung: Datenbeschreibung und Datendokumentation in der Praxis der sozialwissenschaftlichen Datenarchive**

Zentraler Bestandteil der sozialwissenschaftlichen Infrastruktur vor allem in Westeuropa, in den skandinavischen Ländern und in Nordamerika ist seit den sechziger und den siebziger Jahren ein Netz von Datenarchiven. Diese Einrichtungen archivieren für die sozialwissenschaftliche Forschung maschinenlesbare Daten, in erster Linie Forschungsdaten, und stellen sie für weitere Untersuchungen zur Verfügung. Anfänglich hauptsächlich mit der Archivierung und Bereitstellung von Umfragedaten befaßt, erfassen die sozialwissenschaftlichen Datenarchive inzwischen auch andere für die sozialwissenschaftliche Forschung relevante Datentypen, z.B. Daten amtlicher statistischer Zählungen. Historische Daten für eine sich sozialwissenschaftlich verstehende historische Forschung, für die historische Sozialforschung, gehören bislang nur zu einem geringen Teil zum Aufgabenbereich der sozialwissenschaftlichen Datenarchive. Aufgrund der Entstehungsbedingungen und auch der aktuellen Situation der meisten sozialwissenschaftlichen Datenarchive – Umfragedaten machen immer noch die Masse der archivierten und für Sekundäranalysen bereitgestellten Daten aus – ist die von anderen Datentypen verschiedene Problematik einer Archivierung und Bereitstellung sozialwissenschaftlich relevanter historischer Daten gar nicht oder nur kaum reflektiert worden. Das Zentrum für historische Sozialforschung hat im Rahmen seiner – von der Deutschen Forschungsgemeinschaft geförderten – Archivierung und Bereitstellung maschinenlesbarer historisch-sozialwissenschaftlicher Forschungsdaten Ansätze für ein Instrumentarium entwickelt, das die Erfordernisse an eine Archivierung und Bereitstellung sozialwissenschaftlich relevanter historischer Daten abdeckt.

Im vorliegenden Beitrag wird ein Teil dieser Bemühungen um die Schaffung eines datenspezifischen Instrumentariums aufgegriffen: die Empfehlungen für Datenbeschreibungs- und Datendokumentationselemente in der historischen Sozialforschung. Die dabei formulierten und bereits praktizierten Empfehlungen decken einen Datenbeschreibungs- und Datendokumentationsbedarf ab, der in Teilen Beschreibungs- und Dokumentationsanforderungen der auf der Grundlage von Umfrage- und anderen Daten arbeitenden Sozialwissenschaften und der traditionellen Geschichtswissenschaft entspricht. Diese Verknüpfung von empirischen Sozialwissenschaften und Geschichts-

wissenschaft manifestiert sich darin, daß sich in der historischen Sozialforschung theoriegeleitete Forschungsfragen und die Verwendung von historischen Quellen, bzw. von prozeß-produzierten Daten miteinander verknüpfen. Die Beschreibungs- und Dokumentationsanfordernisse der historischen Sozialforschung gehen insofern über den entsprechenden Bedarf der Geschichtswissenschaft hinaus, als in der historischen Sozialforschung Informationen aus historischen Quellen in theoretisch relevante Sachverhalte transformiert werden. Dementsprechend besteht nicht nur die Notwendigkeit, vergleichbar den in der Geschichtswissenschaft entwickelten Standards, Quellen aufzulisten und ein Quellenverzeichnis zu erstellen, sondern es ist ein Beschreibungs- und Dokumentationsbedarf abzudecken, der auf ein umfassenderes "Verständnis" der Daten abzielt.

Die sozialwissenschaftlichen Datenarchive verfahren bei der Datenbeschreibung und -dokumentation gemäß den Standards, die in der empirischen Sozialforschung üblich geworden sind, ohne daß allerdings in jedem Fall nach ihnen gehandelt wird. Diese Standards betreffen:

1. Die Beschreibung der Grundgesamtheit oder Population. Die Grundgesamtheit ist in der empirischen Sozialforschung die wichtigste Bezugsgröße, die Einheit, auf die hin Aussagen bzw. Verallgemeinerungen beabsichtigt sind. "Die Grundgesamtheit (häufig auch als Universum oder Population bezeichnet) kann sich etwa auf die Bevölkerung eines Staates oder einer Gemeinde beziehen, kann aber auch nach einer kategorialen Zugehörigkeit von Personen bestimmt werden: etwa Richter, Sportler einer bestimmten Disziplin oder Minister in Deutschland. Die Grundgesamtheit kann sich aber auch Kollektive oder Artefakte beziehen: z.B. auf Organisationen, wie Unternehmen der Stahlindustrie oder die Finanzverwaltung des Bundes und der Bundesländer. Artefakte können beispielsweise Bilder des 19. Jahrhunderts sein, die Zeitungsberichterstattung über eine gewisse Klasse von Ereignissen (z.B. die Dreyfusaffäre oder Berichterstattung über Wissenschaft), aber auch Dokumente, wie z.B. Personalakten von Behörden." (Heine von Alemann, Der Forschungsprozeß, Stuttgart: Teubner 1977, S. 74)
2. Die Beschreibung bzw. Dokumentation derjenigen Merkmale, über die im Verlauf der Datenerhebung quantitative Informationen erhoben werden. Idealerweise ist der Ausgangspunkt dieses Vorgangs die Umsetzung einer theoretisch abgeleiteten Forschungsfrage in meßbare Größen, die wiederum in die Formulierung eines adäquaten Erhebungsinstruments (z.B. Fragebogens) mit (abhängig von der "Aufmachung" des Erhebungsinstruments) komplementärem Codeplan überleitet.

Der Datenbeschreibungs- und Dokumentationsbedarf wird in der empirischen Sozialforschung in der Weise abgedeckt, daß einerseits eine Definition bzw. eine Beschreibung der Grundgesamtheit gegeben wird; daneben die Merkmalsdimensionen der jeweiligen Untersuchung durch eine Wiedergabe des Erhebungsinstruments (Fragebogen und Codeplan) dokumentiert werden. Die sozialwissenschaftlichen Datenarchive decken diesen Datenbeschreibungs- und Datendokumentationsbedarf ab, indem sie neben einer Wiedergabe des Original-Fragebogens und des Codeplans detaillierte Informationen über die Grundgesamtheit der betreffenden Studie und über die jeweils gezogene Stichprobe zur Verfügung stellen. Die Beschreibung von Grundgesamtheit und Stichprobe wird im Rahmen der sog. "Studienbeschreibung" vorgenommen, einem Instrument, das von verschiedenen westeuropäischen Datenarchiven gemeinsam zur Beschreibung und Dokumentation von Umfragedaten entwickelt worden ist. Mit der Wiedergabe des Original-Fragebogens und des Codeplans und gegebenenfalls einer Auflistung der Variablen werden in der Regel die Anforderungen an eine Beschreibung und Dokumentation der Merkmalsdimensionen von Umfragedaten als erfüllt angesehen. Dies nicht zuletzt deshalb, weil in der Sozialforschung auf der Grundlage von Umfragedaten die Umsetzung der theoretisch abgeleiteten Forschungsfrage in Meßoperationen, d.h. die Operationalisierung, idealiter identisch ist mit der Fragebogenkonstruktion. Zudem wird ein zusätzlicher Beschreibungs- und Dokumentationsbedarf auch deshalb nicht gesehen, weil davon ausgegangen wird, daß die Inhalte eines Fragebogens bzw. die Variablen aus dem zeitgenössischen Alltagsverständnis deutlich werden. Dies kann für die historischen Kontexte der Inhalte der historischen Sozialforschung nicht geltend gemacht werden, sind doch die Alltagskonnotationen der erforschten Zeit von denen der Zeit des Forschers verschieden. Diese Probleme stellen sich jedoch auch zunehmend für die Sozialforschung auf der Basis von Umfragedaten. In der empirischen Sozialforschung enthalten Codebücher über die Bezeichnung der Variablen und über die Bezeichnung bzw. das Etikett der Variablenausprägung normalerweise keine weiteren Informationen zum inhaltlichen "Verständnis" der Variablen. Die Wiedergabe der Variablenbezeichnung, in der Umfrageforschung mit der einzelnen Frage des Fragebogens identisch, wird häufig noch als ausreichend angesehen; entsprechen sich doch Fragebogenformulierung und Bezugssystem des Alltags. Dieses Selbstverständnis der Umfrageforschung wird schon brüchig, sobald Sekundäranalysen von Umfragen z.B. der frühen fünfziger Jahre durchgeführt werden sollen. Die Bedeutung einzelner Variablen aus Untersuchungen der damaligen Zeit ist nicht ohne weiteres transportabel, entspricht doch das Bezugssystem der damaligen

Zeit nicht mehr unserem. Die Alltagskonnotationen einer Fragebogenformulierung aus den fünfziger Jahren sind uns nicht mehr bekannt.

## **2. Zentrale Datenbeschreibungs- und Datendokumentationsprobleme in der historischen Sozialforschung**

### **2.1 Grundgesamtheit, Auswahlgesamtheit, Erhebungseinheit und Quelle in der historischen Sozialforschung**

In der historischen Sozialforschung stellt sich verstärkt das Problem, inwieweit Grundgesamtheit bzw. Zielgesamtheiten durch ein vorhandenes Quellenmaterial adäquat abgebildet werden können. In der empirischen Sozialforschung hat dieses Problem einen weitaus geringeren Stellenwert: der Festlegung der Grundgesamtheit (Bevölkerung der Stadt XYZ), der Auswahlgesamtheit und der eventuellen Ziehung einer repräsentativen Auswahl aus der Auswahlgesamtheit folgt die Untersuchung im "Feld", d.h. die Phase der Datensammlung z.B. in Form von Interviews. Dabei ist idealiter die Relation zwischen ausgewählten Einheiten und der angestrebten Ziel- bzw. Grundgesamtheit unproblematisch, können doch in der empirischen Sozialforschung in der Vorphase der Datenerhebung mögliche Inkongruenzen zwischen Erhebungseinheit und Zielgesamtheit in der Regel korrigiert werden. Diese Chance der Korrektur ist in der historischen Sozialforschung nicht gegeben. Die auf die Definition der Zielgesamtheit folgende Bestimmung der Auswahlgesamtheit und der Erhebungseinheit ist abhängig davon, inwieweit diese Erhebungseinheiten in dem zur Verfügung stehenden historischen Quellenmaterial "wiederzufinden" sind und in welchem Maße diese Erhebungseinheiten der angestrebten Zielgesamtheit entsprechen.

Dieses Problem soll an Hand eines Forschungsbeispiels verdeutlicht werden: Gegeben sei eine Forschungsfrage, die Ursachen, Formen und Ausmaße delinquenten Verhaltens der preußischen Bevölkerung im Zeitalter der Hochindustrialisierung, also in der Zeit des Kaiserreichs, untersuchen soll. Ziel- bzw. Grundgesamtheit ist also in diesem Falle die delinquente Bevölkerung Preußens im Zeitraum 1871-1914. Die Festlegung der Erhebungseinheit könnte theoretisch in verschiedenen Varianten resultieren: Einerseits könnte als Erhebungseinheit das einzelne delinquente Individuum herangezogen werden, in gleichem Maße könnten aber auch Aggregate von Personen als Erhebungseinheit fungieren. Quellenmaterial, an Hand dessen sich entsprechende Erhebungseinheiten festlegen und für Analysen in Untersuchungen aufnehmen lassen, sind vorhanden: Würde man versuchen, das einzelne delinquente Individuum als Erhebungseinheit in Betracht zu ziehen, so hätte man auf gerichtliche Untersuchungsakten zu rekurrieren; würde man als Erhebungs-

einheit Aggregate von Personen heranzuziehen versuchen, so könnte man auf die amtlichen Zählungen der gerichtlichen Kriminalstatistik des Statistischen Reichsamts und des Reichsjustizamtes zurückgreifen. Auf diese Quellen beschränkt sich im wesentlichen das Material zur obengenannten Forschungsfrage; in dieser Beschränkung ist aber zugleich auch der Grund dafür zu suchen, daß eine adäquate Abbildung einer Zielgesamtheit "delinquente Bevölkerung Preußens in der Hochindustrialisierung" nicht zu erreichen ist, sondern nur Annäherungswerte möglich sind.

Als generelles Problem ist zusätzlich zu konstatieren, daß die genannten Quellen nicht auf die delinquente Person allgemein, sondern auf *erfaßte* delinquente Personen abstellen, was für sich allein genommen die Beantwortung der Forschungsfrage – so wie sie gestellt worden ist – schon schwierig erscheinen läßt. Von dieser allgemeineren Frage der Messung von Delinquenz einmal abgesehen, stellen sich noch weitere Probleme: Auf das Individuum als Erhebungseinheit und damit auf Fallakten der Justiz zu rekurrieren, wird mit Sicherheit dazu führen, daß nicht alle erfaßten delinquenten Personen in Preußen im genannten Zeitraum, bzw. eine repräsentative Auswahl, falls eine Stichprobe gezogen werden sollte, in die Untersuchung einbezogen werden können; zu groß wird der Verlust an Untersuchungsakten durch Kassation oder durch Kriegsverlust sein, ja es ist vorstellbar, daß Untersuchungsakten für ganze Gebietseinheiten verlorengegangen sein könnten. Im Falle der publizierten Kriminalstatistiken des Deutschen Reiches ist das Überlieferungsproblem nicht relevant, weil die publizierten Tabellen aus der Reihe der "Statistik des Deutschen Reiches" in den wissenschaftlichen und sonstigen Bibliotheken im Laufe der Zeit kaum Verluste erlitten haben; und wenn Verluste eingetreten sind, sind diese in der Regel ersetzt worden. Jedoch erreichen auch die Kriminalstatistiken nur eine Annäherung an die Zielpopulation "delinquente Bevölkerung Preußens im Zeitalter der Hochindustrialisierung". Dennoch bedeuten derartige "Defekte" im Quellenmaterial nicht unbedingt, daß nunmehr die Fragestellung geändert werden muß. Die Möglichkeit, fehlende Informationen per Hochrechnung oder per Schätzung auszugleichen, ist sicherlich in vielen Fällen gegeben; allerdings verlangen derartige Korrekturen eine genaue Kenntnis des Materials.

Wenn man auch hier einmal davon absieht, daß bei den Kriminalstatistiken *erfaßtes* delinquentes Verhalten wiedergegeben wird, so bleibt als Problem das der inhaltlichen Eingrenzung: Die Kriminalstatistiken erfassen nur Verstöße gegen Reichsgesetze, ein breites Spektrum bundesstaatlicher Strafgesetzgebung unterhalb der Ebene der Reichsgesetze wird dagegen nicht erfaßt, d.h. eine bestimmte Gruppe der erfaßten delinquenten Bevölkerung

bleibt ausgeblendet. Damit ist auch hier wiederum nur eine Annäherung an die gesuchte Zielgesamtheit erreicht.

Ähnliche Beispiele lassen sich auch zu anderen Fragen aus anderen Bereichen der historischen Sozialforschung finden. Als grundsätzliches Fazit für den Datenbeschreibungs- und Datendokumentationsbedarf in der historischen Sozialforschung auf der Ebene von Zielgesamtheiten und Erhebungseinheiten läßt sich festhalten: Erörterungen über Aussagekraft und Reichweite herangezogener Quellen müssen – über den Rahmen der traditionellen Fragen der historischen Quellenkritik hinaus – Antworten auf Fragen nach der Abbildungsqualität von Quellen im Hinblick auf die angestrebte Zielgesamtheit beinhalten. Diese Antworten müssen dokumentiert werden. Ohne Ausführung über die Relation zwischen der auf bestimmten Quellen beruhenden Erhebungseinheit und der intendierten Zielgesamtheit kann die Gültigkeit der betreffenden Untersuchung nicht beurteilt werden.

## **2.2 Indikator und Quelle in der historischen Sozialforschung**

Die Indikatorenbildung ist in der Sozialforschung – neben der Definition von Grundgesamtheit, Auswahlgesamtheit und Erhebungseinheit – ein wichtiger Bestandteil der Konzeptualisierungsphase einer empirischen Untersuchung. Ausgangspunkt dieses Vorgangs ist in der Regel eine theoretische Forschungsfrage. Zumeist ist eine derartige Forschungsfrage nicht direkt beantwortbar, d.h. sie ist nicht ohne weiteres überprüfbar, meßbar. Dazu bedarf es der Umsetzung des theoretischen Sachverhalts in beobachtbare Größen, d.h. es werden Sachverhalte (Einstellungen, Verhalten, etc.) definiert, die den theoretischen Sachverhalt indizieren. Die Antwort auf eine Frage nach dem Indikator, der diese Funktion erfüllen kann, ist in den Modellvorstellungen der empirischen Sozialforschung primär abhängig von theoretischen Vorüberlegungen und Ergebnissen empirischer Vorarbeit. In der Forschungspraxis, insbesondere in der Sekundäranalyse, ist zwar das Verhältnis zwischen zu messendem Sachverhalt und Indikator häufig interaktiv; dennoch ist immer noch dieses deduktive Modell der Ermittlung von Indikatoren stilbildend. Dieser deduktive Ansatz hat Konsequenzen für das Verhältnis von Indikator und sozialem Sachverhalt, der mit Hilfe des Indikators gemessen werden soll. Die Definition von Indikatoren zwecks Messung der in der Forschungsfrage angesprochenen Sachverhalte führt im Ergebnis dazu, daß auf der Grundlage dieser Entscheidungen und daraus resultierender Meßoperationen die für eine Untersuchung relévante Realität bestimmt, "selektiert" wird. Man könnte mithin sagen, daß das Datenmaterial einer empirischen Untersuchung abhängig von den definitorischen Eingrenzungen der

die Untersuchung durchführenden Personen ist, daß die in der Untersuchung abgebildete Realität durch einen Selektionsvorgang in der Konzeptualisierungsphase determiniert wird. In der historischen Sozialforschung stellt sich dieses Problem anders. Zwar entspricht sich im Ansatz die Indikatorenbildung in beiden Zweigen der Sozialforschung, der Tatbestand jedoch, daß der historisch arbeitende Sozialforscher an bereits vorliegendes Material gebunden ist, hat weitergehende Rückwirkungen auf die Indikatorenbildung. Zwar kann im Idealfall, d.h. bei günstiger Quellenlage, in gewissem Maße auch eine "freie" Selektion der für die Untersuchung relevanten Realität durch die Indikatorenbildung stattfinden. Üblicherweise muß man in der historischen Sozialforschung aber davon ausgehen, daß die Indikatorenbildung an das vorhandene Material gebunden ist; wenn – vereinfacht formuliert – vorhandenes Einkommen als ein "starker" und Beruf ein "schwacher" Indikator für soziale Schichtung angesehen werden, die Quellen jedoch keine Einkommensangaben enthalten, so muß Beruf als Indikator für Schichtung berücksichtigt werden. Die für eine Untersuchung relevante Abbildung der Realität liegt in der historischen Sozialforschung somit bereits vor und wird nicht erst im Rahmen der Indikatorenbildung konstituiert. Damit stellt sich für die historische Sozialforschung verschärft das Problem, ob ein Indikator einen theoretischen Sachverhalt adäquat "übersetzen" kann.

In der Forschung wird dieser Problembereich auch im Rahmen der Diskussion über die Kontextgebundenheit historischer Daten bzw. prozeß-produzierter Daten diskutiert. Historische Quellen wurden in der Regel zu anderen Zwecken geschaffen als dem, auf Fragen der Sozialforschung von heute Antworten zu bieten. Die Vermögenseinteilung eines städtischen Steuerverzeichnisses des ancien régime genügt sicherlich den fiskalischen Interessen der damaligen Stadtschreiber und ihrer Auftraggeber; ob diese Einteilung aber auch als adäquat angesehen werden kann, einen Sachverhalt wie Stand oder soziale Schichtung zu indizieren, steht zur Diskussion. Dies hat zur Konsequenz, daß die Kontextabhängigkeit von historischem Quellenmaterial diskutiert und im Hinblick auf den im Rahmen einer Forschungsfrage gemeinten theoretischen Sachverhalt evaluiert werden muß. Die Diskussion und die Evaluation der Quelle müssen zum "Verständnis" von Daten und zur Analyse "weitertransportiert" werden, d.h., sie müssen dokumentiert werden. Diese Informationen ermöglichen eine Beurteilung der Qualität der Primärforschung, lassen aber auch zu, daß Daten auch im Rahmen von sekundäranalytischen Nutzungen verwendet werden können.

Aus der geschichtswissenschaftlichen Diskussion quellenkritischer Fragen, vor allem aber aus den Forschungen über die Verwendbarkeit zeitgenössischer prozeß-produzierter Daten lassen sich einige generelle Kriterien



ableiten, an Hand derer Indikatorenprobleme evaluiert werden können und die zum Verständnis der Daten und der Analyse einer Untersuchung dokumentiert werden müssen. Zu diesen Kontextmerkmalen gehören:

1. Merkmale, die die faktischen und normativen Entstehungsbedingungen und die "Geschichte" von Quellen betreffen (z.B. Aufgabenprogramme, aktenproduzierende Organisationen, Zweck der Aktenanlage, etc.).
2. Merkmale, die das formelle und informelle Verhalten des Personals quellen- bzw. aktenproduzierender Organisationen berühren (spezielle Formen der Aktenführung, "Codes" in Akten, all diejenigen Merkmale, die als "Realität eigener Art" von Akten bezeichnet werden)
3. Merkmale, die – im Falle personenbezogener Quellen bzw. Massenakten – auf spezifische Verhaltensweisen der erfaßten Personen gegenüber der aktenführenden Instanz rekurrieren (Bereitschaft, bestimmte Informationen weiterzugeben, in Relation zu den Absichten der aktenführenden Instanzen, etc.).

Ad 1: Akten als Hinterlassenschaft des Aufgabenvolumens staatlicher Organisationen sind hierfür ein klassisches Beispiel. Aktenvermerke und Berichte bayrischer Provinzbeamter über den "sittlichen" Verfall der Landbevölkerung in der ersten Hälfte des 19. Jahrhunderts sind dazu verwendet worden, Aussagen über den Wandel der Sexualmoral in dieser Zeit zu formulieren. Inwieweit solche Berichte derartige Schlüsse zulassen und nicht eher nur die Sichtweise der kontrollierenden Instanzen indizieren, ist von denjenigen Wissenschaftlern zu beantworten, die diese Forschungsfrage gestellt haben. Die Frage, ob diese quellen als Indikatoren für andere Sachverhalte Gültigkeit besitzen, kann nur dann beantwortet werden, wenn die Informationen über die Entstehungsbedingungen des Materials weitere Aussagen zulassen.

Ad 2: Eine weitere Dimension des Quellenkontextes im Rahmen der Diskussion über die Indikatorenqualität von Informationen aus Quellen betrifft das Verhalten, den Verwaltungsstil derjenigen Individuen, die die Quelle erstellt haben. Auch hierzu ein Beispiel: Aus einer Untersuchung über den Widerstand gegen das NS-Regime in Düsseldorf wissen wir z.B. über die Indikatorenqualität von Vernehmungprotokollen der Gestapo, daß die Akten häufig Sprachcodes enthalten, daß Sachverhalte systematisch ausgeblendet werden, indem sie in Handakten "versteckt" werden, oder aber indem für sie das Prinzip der Schriftlichkeit außer Kraft gesetzt wird. Derartige Tatbestände müssen wiederum sorgfältig diskutiert und dokumentiert werden, um die Indikatorenqualität des verwendeten Materials beurteilen zu können.

Ad 3: Antworten auf die Frage, ob das, was ursprünglich gemessen werden sollte, auch tatsächlich im Kontext der Quelle ermittelt wurde, sind nicht allein davon abhängig, wie sich Mitglieder aktenführender Instanzen verhielten. Das Verhalten auch derjenigen, über die Informationen erhoben wurden, spielt eine Rolle. Auch hierzu ein Beispiel: Besitzangaben in Quellen sind beliebte Informationen, um bestimmte Aspekte des sozio-ökonomischen Status von Personen zu bestimmen. Die Bereitschaft zur Offenlegung von Besitzangaben ist unterschiedlich. Besitzangaben, die dazu gedacht sind, die Höhe von steuerlichen Abgaben zu bestimmen, sind sicherlich anders einzuschätzen als Vermögensangaben, die die Grundlage für Wahlberechtigungen bilden. Wiederum einen anderen Stellenwert haben die Besitzangaben, mit denen Armenverwaltungen die Höhe von Zuteilungen begründen.

### **3. Empfehlungen für eine Verbesserung der Datenbeschreibung und Datendokumentation in der historischen Sozialforschung**

#### **3.1 Vorbemerkung zu den Empfehlungen**

Im vorausgegangenen Teil ist grundsätzlich der Datenbeschreibungs- und Datendokumentationsbedarf formuliert worden, der bei Untersuchungen der historischen Sozialforschung akut wird. Derartige Datenbeschreibungen und Datendokumentationen sind in zweierlei Hinsicht notwendig: Einerseits ist zur Bewertung und Diskussion von Ergebnissen der Forschung ein "Verständnis" der Daten einer Untersuchung erforderlich. Darüber hinaus erleichtert eine Optimierung derartiger Datenkennungen die weitere Nutzung maschinenlesbarer historischer Forschungsdaten im Rahmen von Sekundäranalysen. Allgemein sollen die im folgenden vorgeschlagenen Richtlinien dazu dienen, den bisher unzureichenden Zustand von Datenbeschreibung und Datendokumentation in der historischen Sozialforschung zu verbessern. Denn: Verbesserte Datenkennungen können helfen, die Archivierung maschinenlesbarer historischer Forschungsdaten zu erleichtern und angemessen archivierte und dokumentierte Daten für Sekundäranalysen zur Verfügung zu stellen. Die vorliegenden Empfehlungen stellen die Erfahrungen mehrerer Jahre intensiver Beschäftigung mit maschinenlesbaren historischen Forschungsdaten dar und verstehen sich als Versuch, sowohl zu einem besseren "Verständnis" von Forschung als auch zu einem leichteren Zugang zu Forschungsdaten zwecks Sekundäranalysen zu gelangen, so wie auch die Verwendung von Forschung für Forschung in einem praktischen Sinne zu verbessern.

Die Strukturierung der Beschreibungs- und Dokumentationsitems in der vorliegenden Empfehlung folgt dem bereits genannten Instrument zur Be-

schreibung und Dokumentation von Umfragedaten. Vor einigen Jahren ist dieses Instrument auf Anregung des Zentrums für historische Sozialforschung um Bestandteile erweitert worden, die es ermöglichen, auch maschinenlesbare historische bzw. prozeß-produzierte Daten zu beschreiben und zu dokumentieren. Diese Standardisierung von Datenbeschreibung und -dokumentation auch für historische Daten soll nunmehr einen Datenaustausch auch über nationale Grenzen hinweg ohne größere Probleme erlauben.

### **3.2 Empfehlungen zur Beschreibung von Grundgesamtheit und Erhebungseinheiten**

Ein Instrument zur Beschreibung und Dokumentation maschinenlesbarer historischer Forschungsdaten muß notwendigerweise die Bereiche abdecken, die in den grundsätzlichen Ausführungen zu Beginn dieses Beitrags als beschreibungs- und dokumentationsbedürftig für das "Verständnis" von Daten aus der Sozialforschung insgesamt definiert worden sind. Dazu gehören Aussagen über

- A. Die Ziel- bzw. Grundgesamtheit, auf die die Aussagen einer Untersuchung bezogen sind, und die evtl. Auswahl bzw. Stichprobe, die aus einer derartigen Gesamtheit gezogen worden ist.
- B. Die Merkmalsdimensionen (Variablen) einer Untersuchung / Kriterien zur Beurteilung der Indikatorenqualität der herangezogenen Quelleninformationen.

Den Informationen über die Grund- bzw. Zielgesamtheit einer Studie bzw. über die daraus gezogene Auswahl bzw. Stichprobe sind Informationen über Art und Herkunft der verwendeten Daten vorangestellt. Diese Kategorie, die in dem ursprünglich für die Beschreibung von Umfragedaten konzipierten Instrument nicht vorhanden war, ist für die historische Sozialforschung von zentraler Bedeutung. Neben der Auflistung der verwendeten Quellen, spezifiziert nach den verschiedenen eventuell verwendeten Quellenarten und ihrer Herkunft und Aufbewahrung, sollte an dieser Stelle vermerkt werden, inwieweit verschiedene Quellen zu einem gemeinsamen Datensatz zusammengefügt worden sind.

Die Auflistung von Art und Herkunft des verwendeten Quellenmaterials unterschieden nach unveröffentlichten und veröffentlichten Aktenmaterialien, nach veröffentlichten und unveröffentlichten Zensusmaterialien, nach "Sekundärmaterialien" oder anderen Materialien, ist an sich nichts Neues, insoweit es die Arbeit mit historischem Datenmaterial betrifft. Diese Regel ist jedoch hier noch einmal aufgeführt, um sie ausdrücklich in den Zusammenhang der Datenbeschreibung und Datendokumentation für die historische

Sozialforschung einzufügen. Zugleich soll sie aber dahingehend erweitert werden, daß die in der historischen Sozialforschung nicht unübliche Praxis der Kombination von Quellen zwecks Einfügung fehlender Informationen "am Ort", d.h. bei der Auflistung der herangezogenen Quellen, dokumentiert wird. Dieser Basisinformation folgt der Abschnitt der Datenbeschreibung und Datendokumentation, dessen Inhalt Aussagen über Ziel- und Grundgesamtheit einer Untersuchung und die daraus evtl. gezogene Auswahl bzw. Stichprobe sind.

Die Datenbeschreibung und -dokumentation beschränkt sich in diesem Abschnitt jedoch nicht auf diese beiden Fragen, sondern bezieht eine Reihe weiterer Probleme mit ein. Vorrangig in diesem Beschreibungskomplex ist die Definition der Ziel- bzw. Grundgesamtheit. Damit ist der Rahmen abgesteckt, innerhalb dessen sich das Forschungsinteresse der betreffenden Untersuchung lokalisieren läßt. Mit der Ziel- bzw. Grundgesamtheit werden in der Regel Auswahlgesamtheit und Erhebungseinheit definiert. Zur Berichterstattungspflicht des Wissenschaftlers in der historischen Sozialforschung gehört es, exakte Explikationen der Auswahlgesamtheit und der Erhebungseinheiten zu geben. Inwieweit diese dann die angestrebte Ziel- bzw. Grundgesamtheit abdecken, ist eine Frage, die der Forscher im Rahmen seiner quellenkritischen Überlegungen zu entscheiden hat. Die Ergebnisse dieser Überlegungen sind "mitteilungspflichtig", lassen sie doch Schlüsse darüber zu, inwieweit die aus den Quellen definierten Erhebungseinheiten adäquat (im Sinne der Repräsentativität) die angestrebte Zielgesamtheit wiedergeben und für welche anderen Fragestellung sie eventuell eingesetzt werden könnten.

Eine Kennzeichnung der Relation zwischen angestrebter Zielgesamtheit und den durch die vorhandenen Quellen festgelegten Erhebungseinheiten erlaubt eine Beurteilung der Reichweite (der intendierten und bei günstiger Quellenlage auch erreichten) Forschungsaussage und ist für eine weitere Verwendung der jeweils beschriebenen und dokumentierten Daten unerlässlich. Es ist durchaus vorstellbar, daß Erhebungseinheiten im Rahmen sekundär-analytischer Nutzungen auch für anderen Ziel- bzw. Grundgesamtheiten repräsentativ sein können als für die, die im Rahmen der Originalerhebung definiert worden sind. Die Notwendigkeit, Korrespondenz von angestrebter Ziel- bzw. Grundgesamtheit und den Erhebungseinheiten zu explizieren, ist auch dann gegeben, wenn keine Stichprobe gezogen, sondern eine Total- bzw. Vollerhebung vorgenommen worden ist. Da die Erhebungseinheiten durch das vorliegende Quellenmaterial präjudiziert werden, ist dieser Problembereich grundsätzlicher Natur, d.h. er berührt die Quellenproblematik von Untersuchungen der historischen Sozialforschung.

Ist jedoch aus dem vorliegenden Quellenmaterial eine Stichprobe gezogen worden, so ist die Kennzeichnung des angewandten Stichprobenverfahrens und der Vorgehensweise bei der Stichprobenziehung Teil der Berichtspflicht des Forschers. Entsprechende Angaben erlauben es, die Repräsentativität der untersuchten Teilmengen einer Untersuchung im Hinblick auf die angestrebte Ziel- bzw. Grundgesamtheit zu beurteilen. Bestandteil der Kennzeichnung der Auswahl sollten Angaben sein, die eine Beurteilung der Entsprechung zwischen Merkmalen der ausgewählten Einheiten und der Zielgesamtheit zulassen. Diese Frage kann eventuell auch auf das Problem der Entsprechung von Zielgesamtheit und den durch das vorliegende Quellenmaterial bestimmten Erhebungseinheiten zurückführen, andererseits aber auch ein Verweis auf Probleme der Vorgehensweise bei der Stichprobenziehung sein, deren Kennzeichnung nicht nur für das "Verständnis" der Daten, sondern erneut auch für sekundäranalytische Nutzungen von Daten von Bedeutung ist.

In der Diskussion über die grundsätzlichen Probleme der Abbildung einer bestimmten Grund- bzw. Zielgesamtheit durch bestimmte Quellen wurde auf die Überlieferungsproblematik als einem möglichen Faktor verwiesen, der zu einer verzerrten Abbildung der Grundgesamtheit durch die Quellen führen kann. Um derartige Verzerrungen zu erfassen, wurden auf der Grundlage der vom Zentrum entwickelten Datenbeschreibungs- und Datendokumentationsmerkmale besondere Kennungen der Quellenüberlieferung aufgenommen. Damit sollte es in Zukunft möglich sein, Verzerrungen im Hinblick auf die zu bildende Grundgesamtheit zu beurteilen.

Um die Überlieferungsproblematik der verwendeten Quellen zu verdeutlichen und zu kennzeichnen, sollen die im folgenden aufgeführten Fragen nach Zustand und Erhaltung von Quellen beantwortet werden. Archivare, die in ihrer alltäglichen Arbeit weitaus intensiver mit den Problemen der Quellenüberlieferung befaßt sind, werden die hier vorgeschlagenen Kategorien sicherlich als nicht ausreichend empfinden. Demgegenüber muß aber darauf verwiesen werden, daß die vorgeschlagenen Kategorien Kompromisse zwischen einem wünschbaren Optimum an Informationen über Quellenbestände und den Einschränkungen, die aus der Standardisierung eines Instruments resultieren, darstellen. Ganz wesentlich ist auch, daß hier eine offene Frage gestellt wird, welche das Ausmaß der Beantwortung dem Einzelfall überläßt. Die Überlieferungsproblematik einer Quelle sollte im einzelnen entlang den folgenden Merkmalen gekennzeichnet werden. Abhängig vom "Zustand" der verwendeten Quelle können diese Kennzeichnungen durchaus mehrere der genannten Merkmale betreffen. Angaben sollten darüber gemacht werden:

- ob die verwendete Quelle vollständig erhalten ist.
- ob die verwendete Quelle nur teilweise erhalten ist, d.h. ob sie eventuell gemäß bestimmten, vielleicht auch nachvollziehbaren Kriterien vom Archivpersonal teilkassiert worden ist,
- ob die Quelle teilkassiert worden ist, ohne daß bestimmte Kassationskriterien angebbbar sind,
- oder ob die Quelle vielleicht teilzerstört ist, weil der Bestand aufgrund äußerer Einwirkungen (Feuer, Krieg, sonstige physische Beeinträchtigungen) beschädigt ist.

Neben der Bewertung der Fragen, in welchem Maße die verwendete Quelle erhalten ist, ist es auch wichtig, Hinweise auf die interne Organisation des Quellenmaterials zu geben. Die Abbildung einer Grundgesamtheit in den Quellen ist auch davon abhängig, wie sichtbar Informationen in Quellen sind. Dazu sind Angaben notwendig

- über die Form der Aktenhaltung beim Quellenproduzenten, soweit diese noch aus der Quelle sichtbar ist,
- oder aber über die Form und das Ausmaß der Umorganisation der Quelle im Archiv oder bereits bei der Stelle, die die Quellen produziert hat.

Die Notwendigkeit, die Abbildung einer Zielgesamtheit durch eine Quelle einschätzen zu müssen, verlangt über die Informationen über das Ausmaß und Art der Überlieferung hinaus auch Angaben über die Dokumentation und den Zugang zur Quelle. Eine unzulängliche Quelldokumentation, ein unvollständiges Findbuch, eine nicht gewährleistete vollständige Zugänglichkeit zur Quelle können durchaus in Abbildungsfehlern resultieren. Deshalb sollte die Datenbeschreibung/Datendokumentation Informationen darüber enthalten, ob die verwendete Quelle

- vollständig erschlossen und zugänglich ist,
- ob sie eventuell nur teilerschlossen, jedoch prinzipiell zugänglich ist,
- und ob sie vielleicht sogar noch völlig unerschlossen ist.

### **3.3 Beschreibung der Merkmalsdimensionen (Variablen)/Kriterien zur Beurteilung der Indikatorenqualität der herangezogenen Quelleninformationen**

In der grundsätzlichen Diskussion der Datenbeschreibungs- und Datendokumentationsprobleme der historischen Sozialforschung sind bereits die Probleme angesprochen worden, die bei einem dem "Verständnis" der Daten der historischen Sozialforschung adäquaten Beschreibungs- und Dokumentationsinstrument berücksichtigt werden müssen. Zwei Problembereiche, die diskutiert und dokumentiert werden sollen, wurden angesprochen:

- Die Beschreibung von Kontextmerkmalen der Quelle, um die Qualität der verwendeten Indikatoren evaluieren zu können.
- Der erweiterte Codeplan, um die zeitgebundene Bedeutung von Variablen zu kennzeichnen und festzuhalten.

Die Beschreibung des Kontextes von Quellen sollte in erster Linie folgende Merkmale berücksichtigen:

- Formelle Entstehungsbedingungen von Quellen, insbesondere Gesetze, Verwaltungsanordnungen, Geschäftsordnungen, etc. Die entsprechenden Regelungen sollten aufgeführt werden, vor allem sollten auch Operationalisierungen auf administrativer Ebene, z.B. die Umsetzung von Gesetzen in Ausführungsverordnungen dokumentiert werden.
- Charakteristika der Informationsverarbeitungsstrategien von Quellenproduzenten. An dieser Stelle sollten en détail diejenigen festgestellten Verfahren, Routinen, Schreibweisen, Auslassungsstrategien, etc. des oder derjenigen festgehalten werden, die die betreffende Quelle erstellt und bearbeitet haben.
- Daneben sind, soweit feststellbar, auch diejenigen Informationen eines Quellenkontextes zu berücksichtigen, die auf die Möglichkeit rekurrieren, daß Quellen auch von denen beeinflusst sein können, auf die sich die Inhalte der Quellen beziehen. Demententsprechend sollten die Ergebnisse der quellenkritischen Diskussion, die derartige Fragen anschnitten, hier dokumentiert werden.

Der erweiterte Codeplan ist für die historische Sozialforschung unerlässlich, um den Beschreibungs- und Dokumentationsbedarf auf der Variablebene abzudecken. In der empirischen Sozialforschung sind entsprechende Erfordernisse – trotz bestehender Probleme – bislang noch nicht aufgenommen worden.

An dieser Stelle sollen nur die Informationen formuliert werden, die geeignet erscheinen, eine Variable aus dem Bereich der historischen Sozialforschung adäquat zu "verstehen":

Codebuchinformation pro Variable

1. Nr. der Variable
2. Bezeichnung bzw. kompletter Text der Variable
3. Bezeichnung bzw. Etikett und Wertebereich der Variablenausprägungen
4. Position der Variable auf dem Datenträger (z.B. Kartenart, falls Lochkarte als Datenträger verwendet wurde, Spaltenkennungen)
5. Randauszählung
6. Variablenkommentar
7. Quellenangaben.

Während die Informationen 1 - 5 zu den üblichen Codebuchstandards der empirischen Sozialforschung gehören, gehen die beiden übrigen Informationen über die bis dato in diesem Bereich der Sozialforschung noch gültigen Beschreibungsstandards hinaus. Sie erfüllen den zusätzlichen Informationsbedarf der historischen Sozialforschung. Ein derartiger Variablenkommentar kann durchaus aus einer längeren Textstelle bestehen. Bei Angaben zu Aggregatdaten sind auch Informationen über das Meßniveau diejenigen amtlichen Zählungen einzubeziehen, auf deren Grundlage diese Daten erstellt worden sind. Ohne diese zusätzlichen Angaben sind Aggregatdaten nicht "verständlich".

Die Quelleninformation pro Variable hat Kontrollfunktion, sie dient dazu, bei Datenbereinigungen die wiedergegebenen Variablen evtl. anhand der Quelle zu überprüfen.

#### **4. Kurzfassung der Empfehlungen für eine Verbesserung der Datenbeschreibung und Datendokumentation in der historischen Sozialforschung**

A) Beschreibung von Ziel- bzw. Grundgesamtheit, von Erhebungs- bzw. Beobachtungseinheiten und eventueller Stichproben aus der Quelle

A1) Definition der Ziel- bzw. Grundgesamtheit:

Auf welche Ziel- bzw. Grundgesamtheit bezieht sich die Fragestellung Ihrer Untersuchung? Beispiele möglicher Antworten: Ziel- bzw. Grundgesamtheit ist die Bevölkerung der Stadt Osnabrück in der zweiten Hälfte des 19. Jahrhunderts; Ziel- bzw. Grundgesamtheit ist der Grundbesitz in der Stadt Köln den Jahren 1850 bis 1910 etc.

A2) Explikation der Erhebungs- bzw. Beobachtungseinheiten

Anhand welcher (in den Quellen erfaßter) Einheiten haben Sie versucht, Aussagen über Ihre Ziel- bzw. Grundgesamtheit zu machen? Mögliche Antworten dazu im Rahmen der o.g. Beispiele: Erhebungs- bzw. Auswahlinheit sind alle in den Adreßbüchern der Stadt aufgeführten Haushalte, oder aber alle in den Katasterverzeichnissen der Stadt Köln vorgenommenen Eintragungen über Besitzverhältnisse.

A3) "Repräsentativität" der Erhebungseinheiten im Hinblick auf die abzubildende Zielgesamtheit:

Hier ist die Frage zu beantworten, ob und wie die Erhebungseinheiten (die oftmals durch das vorliegende Quellenmaterial bedingt sind) die Zielgesamtheit abbilden. Um weiterhin im Rahmen eines unserer Beispiele zu argumentieren: Bilden Adreßbücher der Stadt Osnabrück deren Sozialstruktur



im gegebenen Zeitraum adäquat ab, oder muß auch in diesem Fall konstatiert werden, daß, wie häufig bei diesem Quellentyp zu beobachten, Unterschichtenangehörige nicht erfaßt werden? Wie groß schätzen Sie diese Dunkelziffer?

A4) Angaben über das verwendete Stichprobenverfahren und über die Vorgehensweise bei der Ziehung der Stichprobe:

Falls Sie eine Stichprobe aus Ihrer Quelle gezogen haben, beschreiben Sie die Vorgehensweise bei der Ziehung der Stichprobe. Die Antwort auf diese Frage sollte überaus detailliert Aufschlüsse über die verschiedenen Schritte einer Stichprobenziehung und über die Qualität der gezogenen Stichprobe zulassen. Als Beispiel einer derartigen Beschreibung (einer komplizierten Vorgehensweise) sei hier der entsprechende Abschnitt aus der bereits genannten Untersuchung über den Widerstand in Düsseldorf gegen das NS-Regime angeführt:

“Die ‘Personalakten’ der Gestapoleitstelle Düsseldorf waren nach dem Stand vom 1. Mai 1977 zu über zwei Drittel in mehreren Karteien systematisch erschlossen. In einer ‘Ortskartei Düsseldorf’ waren ca. 5000 Karteikarten über Personen gesammelt, die zwischen 1933 und 1945 in Düsseldorf ‘staatspolitisch in Erscheinung getreten’ sind und über die die Gestapo Düsseldorf eine (oder mehrere) ‘Personalakten’ angelegt hatte. Diese Ortskartei ist nach 52 ‘Hauptschlagwörtern’ in Anlehnung an den Geschäftsverteilungsplan des Geheimen Staatspolizeiamtes in Berlin vom 1.7.1939 untergliedert. 41 Schlagwörter wurden in die Erhebung einbezogen, elf blieben dagegen un bearbeitet. Als Grundgesamtheit, aus der die Auswahl gezogen werden sollte, wurden schließlich die ca. 3.770 Karteikarten in den 41 relevanten Schlagwörtern der ‘Ortskartei Düsseldorf’ definiert. Eine (personenbezogene) Karteikarte, auf der die Nummer(n) der in Frage kommenden Gestapoaakten verzeichnet ist (sind), bildete die Auswahlinheit, während eine Person die Analyseeinheit darstellt. Die 41 ‘Hauptschlagwörter’ wurden in vier Kategorien gruppiert und für jede Kategorie Quoten festgelegt; als Stichprobenumfang ergab sich eine Fallzahl von 825 Personen. Im Zuge einer ersten Auswahlstufe wurde aus den ‘Hauptschlagwörtern’ mit mehr als 90 Karteikarten eine Wahrscheinlichkeitsauswahl von 10 % gezogen (Ziehen jeder 10ten Karteikarte mit Zufallsstart). Aus Schlagwörtern mit weniger als 90 Karteikarten wurden die Karteikarten ‘vollständig’ erhoben; allerdings blieben sog. ‘Sammelsachakten’ zunächst unberücksichtigt, so daß auch hier praktisch systematische Auswahl gezogen wurden. Durch diese Verfahren gelang es, das breite Spektrum unterschiedlicher Tatbestände in der Auswahl zu erfassen. Die umfangreichen Tatbestandsgruppen, z.B. im ‘Hauptschlagwort’ ‘Heimtücke’, sind in der Auswahl zugunsten der Tatbestände mit geringen

Fallzahlen unterrepräsentiert. Dies gilt in besonderem Maße für die KPD: Die Karteikarten in diesem Schlagwort umfassen ca. 28 % aller relevanten Karteikarten, in unserem Sample stammen dagegen nur knapp 14 % der Fälle aus diesem Schlagwort, das somit zu 100 % unterrepräsentiert ist" (Reinhard Mann, Protest und Kontrolle im Dritten Reich. Nationalsozialistische Herrschaft im Alltag einer rheinischen Großstadt, Köln und Wien: Böhlau Verlag, erscheint 1985).

A5) Angaben über das "Überleben" der Quelle:

Hier sind die Angaben darüber zu machen, ob die Quelle eventuell vollständig erhalten ist, oder ob sie eventuell teilkassiert oder zerstört worden ist. Soweit Teilkassationen vorgenommen worden sind, sind Ihnen die Kassationskriterien bekannt (geworden)? Ist vielleicht systematisch ein Teilbestand kassiert worden, oder aber streuen sich die Kassationen über den gesamten Bestand? Kann man etwas über den Effekt sagen, den eventuelle Teilerstörungen hatten? Falls Sie eine Stichprobe gezogen haben, wie haben Sie derartige Kassationen/Zerstörungen berücksichtigt?

A6) Angaben über die Ordnung der Quelle:

Hier sind Fragen zu beantworten, die die Ordnung des Quellenmaterials betreffen. Falls das Quellenmaterial Ihrer Untersuchung personenbezogene Massenakten sind, wurden die Quellen alphabetisch, chronologisch oder geographisch oder in Kombinationen hiervon sortiert? Sind diese Sortierungen in der Vergangenheit eventuell verändert worden? Kennen Sie z.B. Anweisungen für die Aktenführung, die von der aktenhaltenden Instanz ausgegeben worden sind? Wenn ja, zitieren Sie diese Anweisungen! Sind einzelne Akten des Bestandes im Zeitverlauf ergänzt worden, nicht jedoch alle Akten insgesamt? Ist die Quelle nach der Abgabe an das Archiv in der Weise verändert worden, daß z.B. eine ursprünglich vorhandene Sortierung zugunsten einer anderen Sortierung verändert worden ist.? Welche Konsequenzen hat dies für Ihre Stichprobenbeziehung gehabt?

A7) Angaben über die Erschließung und die Zugänglichkeit einer Quelle:

Mit welchen Mitteln und in welchem Umfang war die von Ihnen herangezogene Quelle erschlossen? Existierte ein Findbuch, war die Quelle per Karteikarte teilweise oder komplett erschlossen? Gab es noch den ursprünglichen Aktenplan? Existieren überhaupt Informationen über die Sortierungskriterien der Quellen?

B) Beschreibungen der Merkmalsdimensionen, Kriterien zur Bewertung von Indikatoren, Variablenkommentare

B1) Angaben zu den formellen Entstehungsbedingungen von Quellen:

Mit welchen Intentionen wurde die für Ihre Untersuchung relevante Quelle ursprünglich geführt? Machen Sie hier bitte Angaben über den formellen Auftrag, der (in der Regel) zur Entstehung einer Quelle geführt hat (im Falle von Massenakten der öffentlichen Verwaltung) oder aber die Absicht, die hinter der Entstehung anderer Quellen (wie z.B. Adreßbücher) stand.

#### B2) Angaben über das Verhalten von Quellenproduzenten:

Ob anhand vorhandener Angaben in den Quellen das gemessen wird, was gemessen werden sollte, ist auch abhängig von verschiedenen Faktoren. Dazu sind Angaben über die faktischen Verfahrensweisen bei der Quellenerstellung notwendig. Ähnlich den Codes, die heute in Zeugnissen über ausscheidende Mitarbeiter einer Firma üblich sind, können auch historische Quellen spezifische Verfahrensweisen reflektieren, die beschrieben werden müssen, um die Indikatorfunktionen der Quelle beurteilen zu können.

#### B3) Angaben über mögliches Verhalten der "Objekte" von Quellen:

Die Frage, ob die Personen, über die Informationen in der betreffenden Quelle erhoben worden sind, Einfluß auf diese Informationen hatten, ist in diesem Kontext ebenfalls zu beantworten. Gab es ein Interesse dieser Personen, Informationen eventuell zu selektieren, z.B. die Höhe persönlichen Besitzes je nach Kontext unterschiedlich anzugeben? Konnten Sie derartige kontextabhängige Verzerrungen überprüfen; war es Ihnen möglich, eventuell Korrekturen vorzunehmen?

#### B4) Variablenkommentar im Codebuch:

Bitte, kommentieren Sie die Bezeichnung der Variablen und der Variablenausprägung so detailliert wie möglich. Geben Sie Quelleninformationen pro Variable. Listen Sie die folgenden Informationen pro Quelle auf:

1. Nr. der Variable, 2. Bezeichnung bzw. kompletter Text der Variable, 3. Bezeichnung bzw. Etikett und Wertebereich der Variablenausprägung, 4. Position der Variable auf dem Datenträger, 5. Randauszählung, 6. Variablenkommentar, 7. Quellenangabe.

Anbei noch ein Beispiel für einen Variablenkommentar. In diesem Kommentar wird die Variable "Politische Position der Abgeordneten" der französischen Nationalversammlung erläutert. Das Beispiel entstammt einer laufenden Untersuchung:

"In dieser Variablen wurden die im Primärmaterial enthaltenen Kennzeichnungen der politischen Position des Abgeordneten erfaßt. Diese Angaben sind vage und verwenden keine einheitliche Terminologie. Als Hinweis auf die politische Position des Abgeordneten führen die Autoren zuweilen Urteile zeitgenössischer Beobachter an, Robert/Couigny werten

zusätzlich Abstimmung aus (nach Raginel). Gelegentlich bleibt unklar, ob sich die Einordnung auf die Position des Abgeordneten während der Julimonarchie oder während der Revolution von 1848/49 bezieht. Problematisch bleibt insbesondere die Untergliederung des rechten Flügels in Legitimisten, Orleanisten, Bonapartisten und 'unspezifische Rechte'. In den häufigen Zweifelsfällen wurde der Abgeordnete der Kategorie 'Rechte' zugeordnet. Ähnlich zweifelhaft ist die Unterscheidung von 'Gemäßigten Linken' und 'Gemäßigten Republikanern'. Auf dem linken Flügel ist die Einordnung eindeutiger und zuverlässiger. Zu einer exakten Zuordnung reichen die Angaben jedoch nicht aus. Dies leisten die aus der seriellen Analyse des Abstimmungsverhaltens gewonnenen Indizes.

In der Variablen PPI wurde die (zeitlich) erste politische Position erfaßt, die Variable PP2 erfaßt die Zielposition derjenigen Abgeordneten, die ihren Standort gewechselt hatten. Auch über diese Positionswechsel informiert zuverlässiger die serielle Analyse des Abstimmungsverhaltens." (Heinrich Best, *Die Abgeordneten der Assemblée Nationale Constituante 1848*, Köln: Zentrum für historische Sozialforschung 1983 (= Datenhandbücher für die historische Sozialforschung, Bd. 2), S. 37 f.)

## Anhang

Probleme der Datenbeschreibung und Datendokumentation sind in sozialwissenschaftlichen Lehrbüchern bislang noch unterbelichtet. Ein Vorschlag für einen Katalog sozialwissenschaftlicher Datenbeschreibungselemente wurde Anfang der 80er Jahre vorgelegt:

- *Alice Robbin*: Technical Guidelines for Preparing and Documenting Data, In: *Robert Boruch et al. (eds.): Reanalyzing Program Evaluations*, San Francisco: Jossey Bass, 1981, S. 84-143.

Die Diskussion um die spezifische sozialwissenschaftliche Nutzung historischer und prozeß-produzierter Daten ist vor allem in den folgenden Arbeiten und Sammelbänden geführt worden:

- *Paul J. Müller (Hrsg.): Die Analyse prozeß-produzierter Daten*, Stuttgart: Klett-Cotta, 1977.
- *Wolfgang Bick, Paul J. Müller*: Probleme der Nutzung prozeß-produzierter Daten, Bonn: Bundesministerium für Forschung und Technologie, 1982.
- *Wolfgang Bick, Reinhard Mann, Paul J. Müller*: Sozialforschung und Verwaltungsdaten, Stuttgart: Klett-Cotta, 1984.