

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

Reinke, Herbert

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

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

GESIS - Leibniz-Institut für Sozialwissenschaften

#### Empfohlene Zitierung / Suggested Citation:

Reinke, H. (1982). Datenbeschreibung und Datendokumentation in der historischen Sozialforschung: Problemaufriss und Empfehlungen für die Forschung. *Historical Social Research*, 7(4), 4-24. <https://doi.org/10.12759/hsr.7.1982.4.4-24>

#### Nutzungsbedingungen:

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

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

#### Terms of use:

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

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

DATENBESCHREIBUNG UND DATENDOKUMENTATION IN DER  
IN DER HISTORISCHEN SOZIALFORSCHUNG  
PROBLEMAUFRISS UND EMPFEHLUNGEN FÜR DIE FORSCHUNG  
TEIL I: DAS PROBLEM

Herbert Reinke<sup>+</sup>

One of the prerequisites for understanding data in historical social research, their appropriate description and documentation, is still met unsufficiently. The current article assesses the problem (part I) and gives recommendations for an adequate description and documentation of machine-readable historical data (part II). The proposals made concern better description and documentation of target-frame solutions in historical social research, description of the sample drawn, statements about the indicator quality of the sources used, and variable comments.

1. DATENBESCHREIBUNG UND DATENDOKUMENTATION IN SOZIAL-  
WISSENSCHAFTLICHEN CURRICULA UND LEHRBÜCHERN

Sieht man einschlägige Veröffentlichungen zur Einführung in die Verfahren der historischen Sozialforschung durch, so stellt man fest, daß ein breites Spektrum von Problemen behandelt wird; Probleme, die im Verlaufe des Forschungsprozesses in der Sozialforschung aufgeworfen werden: Von der Formulierung einer Forschungsfrage, bis hin zu den statistischen Auswertungen von Datenmaterial als dem Versuch, Antworten auf die anfangs gestellte Forschungsfrage zu finden, werden meistens keine Abschnitte des Forschungsprozesses ausgelassen. Zwar werden die Akzente von Publikation zu Publikation unterschiedlich gesetzt, werden bestimmte Abschnitte des Forschungsprozesses gegenüber anderen besonders betont; in der Regel wird jedoch die statistische Analyse von Daten vorrangig behandelt. (1)

Der Vorrang, den die statistische Analyse von Daten in der Darstellung des sozialwissenschaftlichen Forschungsprozesses häufig einnimmt, ist durchaus gerechtfertigt; ist doch diese statistische Interpretation das "Herz" jedweder quantitativen Untersuchung. Demgegenüber nehmen die Voraussetzungen, die diese statistische Analysen von Daten gewährleisten sollen, häufig in der Darstellung des Forschungsprozesses nur einen geringen Raum ein. Zwei Voraussetzungen für eine Datenanalyse sind hier vor allem zu nennen: Die Datensammlung und die Datenbeschreibung und Datendokumentation. Beide Voraussetzungen scheinen selbstverständlich: Ohne Daten keine Analyse derselben, ohne Datenbeschreibung und Datendokumentation kein "Verständnis" der Daten. Während jedoch die Datensammlung in der empirischen Sozialforschung, soweit sie auf den klassischen Erhebungsver-

---

<sup>+</sup> Address all communications to: Herbert Reinke, Zentrum für historische Sozialforschung e.V., Universitätsstraße 20, 5000 Köln 41

fahren wie Interview und Beobachtung und Experiment beruht, diskutiert wird, gilt dieses für die Sozialforschung, die sich auf prozeß-produzierte bzw. historische Daten stützt, noch nicht. Noch weniger als die Verfahren der Datensammlung werden Datenbeschreibung und Datendokumentation in der historischen Sozialforschung zu Themen. Die Gründe für die fehlende Diskussion der Verfahren der Datensammlung in der historischen Sozialforschung sind vielfach: Die in der Regel vorhandene generelle Vertrautheit mit prozeß-produzierten Daten bzw. mit historischem Quellenmaterial, verbunden mit einem Defizit im Hinblick auf die Anwendung statistischer Analyseverfahren hat dazu geführt, daß in Einführungen in das Instrumentarium der Quantifizierung historischer Daten die Darstellung statistischer Verfahren dominiert. Während man jedoch inzwischen hinsichtlich der Anwendung statistischer Verfahren in der historischen Sozialforschung konstatieren kann, daß sie für die Disziplin kein grundsätzliches Problem mehr sind, sondern nur noch auf der Ebene mangelnder, aber erlernbarer Kompetenz behandelt werden, sind Datenbeschreibungs- und Datendokumentationsprobleme für die historische Sozialforschung immer noch ein grundsätzliches Problem. Es gibt in der historischen Sozialforschung so etwas wie einen "Regelkodex", wie er in der Geschichtswissenschaft Anwendung findet, noch nicht. Zentraler Bestandteil dieses "Regelkodexes" in der Geschichtswissenschaft ist der sogenannte wissenschaftliche Apparat. (2) Im wissenschaftlichen Apparat manifestiert sich in der "herkömmlichen" Geschichtswissenschaft der Dokumentationsbedarf dieser Disziplin. Dieser wissenschaftliche Apparat besteht in der Geschichtswissenschaft in erster Linie aus der Auflistung der verwendeten Primärquellen und der herangezogenen Sekundärliteratur. Im strengen Sinne nicht unbedingt ein Teil des wissenschaftlichen Apparats, aber dennoch ein Bestandteil der Quellenproblematik einer Studie, sind diejenigen Teile einer Arbeit, in denen der Autor seine quellenkritischen Überlegungen wiedergibt. Dazu gehören vor allem die zur äußeren Quellenkritik gehörenden Authentizitätsprüfungen und unter Umständen auch die der inneren Quellenkritik zuzurechnenden Bemühungen, eine Quelle adäquat zu interpretieren. Diese Dokumentationsstandards der Geschichtswissenschaft sind nun allerdings für den Dokumentationsbedarf der historischen Sozialforschung unzureichend. In der historischen Sozialforschung werden Informationen aus Quellen in Daten über theoretisch relevante Sachverhalte transformiert, dementsprechend besteht nicht nur die Notwendigkeit vergleichbar der Geschichtswissenschaft, Quellen aufzulisten, sondern es ist ein Beschreibungs- und Dokumentationsbedarf vorhanden, der auf das "Verständnis" der Daten abzielt.

Die Situation der empirischen Sozialforschung ist der der historischen Sozialforschung insofern vergleichbar, als dort ebenfalls Daten über theoretisch relevante soziale Sachverhalte gesammelt werden. Zwar werden die Beschreibungs- und Dokumentationsnotwendigkeiten auch in Lehrbüchern der empirischen Sozialforschung vernachlässigt, aber dennoch gehört die Beschreibung, Dokumentation und die Diskussion des verwendeten Primärmaterials, neben einer Beschreibung der Vorgehensweisen bei der Datenerhebung, zu den Forschungsstandards der empirischen Sozialforschung. Die zentralen Bereiche der Datenbeschreibung und Datendokumentation erfassen in der empirischen Sozialforschung die folgenden Merkmale(3):

1. Die Beschreibung der Population (Bevölkerung, Objekte) bzw. der repräsentativen Teilmenge dieser Population, die die Grundlage der Datensammlung darstellt.
2. Eine Beschreibung bzw. Dokumentation derjenigen Inhalte/Variablen, die im Rahmen der Datensammlung über die oben genannte Population erhoben werden. Ausgangspunkt der Datenbeschreibung und Datendokumentation auf dieser Ebene ist die Umsetzung einer theoretischen Fragestellung in meßbare Größen, die in die Entwicklung eines entsprechenden Erhebungsinstruments (z.B. Fragebogen) überleitet und evtl. (abhängig von der "Aufmachung" des Fragebogens) einen separaten Codeplan enthalten kann. Damit sind diejenigen Informationen wiedergegeben worden, die notwendig und sinnvoll sind zum "Verständnis" von Daten dieser Variante der Sozialforschung.

Die historische Sozialforschung, in der sich theoriegeleitete Forschungsfragen und die Verwendung von historischen Quellen bzw. von prozeß-produzierten Daten miteinander verbinden, hat infolgedessen einen Datenbeschreibungs- und Datendokumentationsbedarf, der Anleihen sowohl aus der Geschichtswissenschaft als auch aus der Soziologie nimmt. Diese spezifischen Beschreibungs- und Dokumentationsanforderungen bestehen darin, daß theoriegeleitete Überlegungen über die Umsetzung in meßbare Größen bis hin zum Erhebungsinstrument und zum Codeplan beschrieben werden müssen, - was dem gängigen Verfahren der empirischen Sozialforschung entspricht. Über dieses gängige Verfahren hinaus aber muß das zeitgenössische Sinnverständnis der wiedergegebenen Inhalte deutlich werden. Diese Notwendigkeit wird in der empirischen Sozialforschung noch nicht so deutlich gesehen, wird doch unterstellt, daß die Inhalte von Variablen aus dem zeitgenössischen Alltagsverständnis heraus deutlich werden. Demgegenüber kann dies 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. Aus der Geschichtswissenschaft wiederum übernimmt die historische Sozialforschung als Disziplin, die ebenfalls an historischen Quellen orientiert ist, den Bedarf, auch die Quellengrundlage der eigenen Forschungen zu beschreiben und zu dokumentieren. Der Ansatz der historischen Sozialforschung, theoriegeleitete Fragen an historisches Quellenmaterial zu stellen, verknüpft den Umsetzungsprozeß von Theorie in Meßoperationen mit historischen Quellen und führt zu einem spezifischen Beschreibungs- und Dokumentationsbedarf. Dieser spezifische Bedarf soll im folgenden diskutiert werden. Vorab aber sollen kurz die bereits vorhandenen Datenbeschreibungs- und Datendokumentationsinstrumente aufgeführt werden.

## 2. ZU DEN STANDARDS DER DATENBSCHREIBUNG UND DER DATEN-DOKUMENTATION IN DER EMPIRISCHEN SOZIALFORSCHUNG

Im vorangegangenen Abschnitt ist darauf verwiesen worden, daß Standards der Datenbeschreibung und Datendokumentation der empirischen Sozialforschung, deren Basis das Interview ist, auch für die historische Sozialforschung relevant sind. Diese Standards sind zwar "common knowledge" in der empirischen Sozialforschung, trotzdem ist

die Aufarbeitung dieser Richtlinien der Datenbeschreibung und Datendokumentation in der deutschsprachigen Literatur unzureichend.(4) In der englischsprachigen Literatur existieren zwar umfangreiche Darstellungen des Datenbeschreibungs- und Datendokumentationsbedarfs in der Sozialforschung, jedoch sind diese im deutschsprachigen Raum bislang noch kaum rezipiert worden.(5) Auch die Ansätze der westeuropäischen Datenarchive, im Rahmen der Archivierung vor allem von Umfragedaten, den Datenbeschreibungs- und Datendokumentationsbedarf zu definieren und zu einem international kompatiblen Instrument der Datenbeschreibung und Datendokumentation zu gelangen, sind bislang nicht bis in die Literatur über Methoden und Verfahren des Forschungsprozesses in der Sozialforschung diffundiert.(6)

Zu den Elementen der Beschreibung der Population bzw. entsprechender repräsentativer Teilmengen gehören in der empirischen Sozialforschung Aussagen über Grund- bzw. Zielgesamtheiten von Untersuchungen, über die Erhebungs- bzw. Analyseeinheiten und über eine evtl. Stichprobenziehung bzw. Auswahl aus der betreffenden Grundgesamtheit. Dabei wird in der Regel ausführlich auf das angewandte Stichprobenverfahren und auf die daraus resultierenden Ergebnisse hinsichtlich Stichprobengröße und Repräsentativität im Hinblick auf die angestrebte Grundgesamtheit angegeben. Im Rahmen der Beschreibung der Inhalte der Forschung werden in der empirischen Sozialforschung, die auf der Grundlage von Umfragedaten beruht, der Originalfragebogen und der Codeplan wiedergegeben. Damit werden normalerweise die Anforderungen an eine Beschreibung und Dokumentation der inhaltlichen Aspekte der Daten als gegeben angesehen, zumal in der Sozialforschung auf der Grundlage von Umfragen die Umsetzung der theoretisch orientierten Forschungsfrage in Meßoperationen, d.h. die Operationalisierung identisch ist mit der Fragebogenkonstruktion.

### 3. ZENTRALE DATENBESCHREIBUNGS- UND DATENDOKUMENTATIONS- PROBLEME IN DER HISTORISCHEN SOZIALFORSCHUNG

#### 3.1 ZIELGESAMTHEIT, ERHEBUNGSEINHEIT UND QUELLE IN DER HISTORISCHEN SOZIALFORSCHUNG (7)

Für die historische Sozialforschung stellt sich das Problem, inwieweit Zielgesamtheiten durch ein vorhandenes Quellenmaterial adäquat abgebildet werden können. In der empirischen Sozialforschung hat dieses Problem einen weitaus geringeren Stellenwert: Der Festlegung der Zielgesamtheit (Bevölkerung der Stadt XYZ) und der eventuellen Ziehung einer repräsentativen Auswahl aus dieser Ziel- oder Grundgesamtheit folgt die Untersuchung im "Feld", d.h. die Phase der Datensammlung z.B. in Form von Interviews schließt sich an. Dabei ist idealiter die Relation zwischen untersuchten Einheiten und der angestrebten Ziel- bzw. Grundgesamtheit unproblematisch, können doch in der empirischen Sozialforschung in der Vorphase der Datenerhebung mögliche Inkongruenzen zwischen Erhebungseinheiten 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 Erhebungseinheiten ist abhängig davon, inwieweit diese Erhebungseinheiten in dem zur Verfügung stehenden historischen Quellenmaterial "wiederzufinden" sind und in welchem Maße diese Erhebungseinheiten für

die angestrebte Zielgesamtheit repräsentativ sind. Dieses Problem soll anhand eines Forschungsbeispiels verdeutlicht werden(8): Gegeben sei eine Forschungsfrage, die Ursachen, Formen und Ausmaß 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 Untersuchungseinheit könnte theoretisch in verschiedenen Varianten resultieren: Einerseits könnte als Untersuchungseinheit das einzelne delinquente Individuum herangezogen werden, in gleichem Maße könnten aber auch Aggregate von Personen als Untersuchungseinheiten fungieren. Quellenmaterial, anhand dessen sich entsprechende Untersuchungseinheiten festlegen und für Analysen in Untersuchungen aufnehmen lassen, sind vorhanden: Würde man versuchen, das einzelne delinquente Individuum als Untersuchungseinheit in Betracht zu ziehen, so hätte man auf gerichtliche Untersuchungsakten zu rekurrieren, würde man als Untersuchungseinheiten Aggregate von Personen heranziehen versuchen, so könnte man auf die amtlichen Zählungen der 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 vorab 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 etwas allgemeineren Frage der Messung von Delinquenz einmal abgesehen, stellen sich noch weitere Probleme: Auf das Individuum als Untersuchungseinheit 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 aber 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.

### 3.2 INDIKATOR UND QUELLE IN DER HISTORISCHEN SOZIALFORSCHUNG

Die Indikatorenbildung ist in der Sozialforschung - neben der Definition von Grundgesamtheit und von Untersuchungseinheit - ein weiterer wichtiger Bestandteil in 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 Einheiten gesucht, die den theoretischen Sachverhalt indizieren. Die Entscheidung darüber, welcher Indikator diese Funktion erfüllen kann, ist in den Modellvorstellungen der empirischen Sozialforschung allein abhängig von theoretischen Vorüberlegungen und Ergebnissen empirischer Vorarbeiten. Diese Vorgehensweise stellt den Untersuchungsansatz dar, wie er in fast jedem Methodenlehrbuch zur empirischen Sozialforschung als Teil des Forschungsprozesses postuliert wird. In der Forschungspraxis, insbesondere in der Sekundäranalyse, ist zwar das Verhältnis zwischen zu messendem Sachverhalt und Indikatoren häufig interaktiv; dennoch ist immer noch das 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 relevante Realität bestimmt, "selektiert" wird. Man könnte mithin sagen, daß das Primärmaterial, auf dem in der empirischen Sozialforschung eine Untersuchung basiert, abhängig von den Definitionen 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 dar. Im Ansatz entspricht

die Indikatorenbildung in diesem Zweig der Sozialforschung der geschichtlichen Vorgehensweise der empirischen Sozialforschung. Der Tatbestand jedoch, daß der historisch arbeitende Sozialforscher an bereits vorliegendes Material gebunden ist, hat 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 - Einkommen ein "starker" und Beruf ein "schwacher" Indikator ist, 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. prozeßproduzierter Daten diskutiert. Historische Quellen wurden in der Regel zu anderen Zwecken geschaffen, als die, 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 Steuereinteilung 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 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 sekundäranalytischer Nutzung verwendet werden können.

Welche allgemeine Kriterien können nun dazu dienen, den Indikatorwert von Primärmaterial zu evaluieren? Da historisches Quellenmaterial bei theoretisch orientierten Forschungsfragen in der Regel nicht im Rahmen ihrer kontextgebundenen Eigenbedeutung genutzt, sondern als Indikatoren für eben diese Forschungsfragen genommen werden, gilt, daß

"ohne eine angemessene Berücksichtigung und Einbringung des je zeit- und raumspezifischen Entstehungs- und Bedeutungskontextes ... aus geschichtlichen Quellen kaum zuverlässige und gültige historische und prozeß-produzierte Daten gewonnen werden (können)". (9)

Welches sind nun diese Kontextmerkmale? Teilweise aus der geschichtswissenschaftlichen Diskussion quellenkritischer Fragen, vor allem aber aus Forschungen über die Verwendbarkeit zeitgenössischer prozeßproduzierter Daten lassen sich die Kontextmerkmale herauskristallisieren, anhand derer Indikatorenprobleme evaluiert werden



können und die zum "Verständnis" der Daten und der Analyse einer Untersuchung dokumentiert werden müssen. Zu diesen Kontextmerkmalen gehören (10):

1. Merkmale, die die faktischen und normativen Entstehungsbedingungen und die "Geschichte" von Quellen betreffen (z.B. Aufgabenprogramme aktenproduzierender Organisationen, Zweck der Aktenanlage etc.).
2. Merkmale, die das formelle und informelle Verhalten des Personals quellen- bzw. aktenproduzierender Organisationen betreffen (spezielle Formen der Aktenführung, "Kodes" in Akten, all diejenigen Merkmale, die als die "Realität eigener Art" (11) 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 Instanz etc.).

Ad. 1:

In der Regel werden durch normative Bestimmungen (Gesetze, Verwaltungsvorschriften, sonstige Anordnungen) Quellen, insofern sie im Verwaltungsgang einer Behörde entstanden sind, konstituiert. Ähnliches gilt aber auch für eine ganze Reihe anderer Quellen. So weiß man, daß viele Haushaltslisten des Ancien Régime erstellt worden sind, um die Steuerfähigkeit und die Abgabefähigkeit der Bevölkerung einzuschätzen. Ob derartiges Material auch ausreichende Handhaben bietet, allgemeine Bevölkerungscharakteristika zu untersuchen, muß sorgfältig geprüft werden, und falls die Frage positiv entschieden wird, mit der gleichen Sorgfalt dokumentiert werden, damit die Indikatorenqualität des verwendeten Materials bewertet werden kann. Ein zusätzliches Beispiel soll das hier vorhandene Beschreibungsproblem weiter erläutern:

Berichte bayerischer Provinzbeamter über den "sittlichen" Zustand der Landbevölkerung in der ersten Hälfte des 19. Jahrhunderts, die von der bayerischen Staatsverwaltung initiiert worden sind, sind dazu verwendet worden, Aussagen über den Wandel der Sexualmoral in der ersten Hälfte des 19. Jahrhunderts zu formulieren. Inwieweit solche Berichte derartige Schlüsse zulassen und nicht eher nur die Sichtweise der kontrollierenden Instanzen indizieren, wird vom Wissenschaftler beantwortet, der diese Forschungsfrage gestellt hat. Die Beurteilung, ob diese Quellen als Indikatoren für den zu messenden Sachverhalt "Wandel der Sexualmoral", oder aber sogar im Rahmen sekundäranalytischer Fragen als Indikatoren für andere Sachverhalte Gültigkeit besitzen, hängt von Informationen über diese Quelle ab. (12)

Ad 2:

Eine weitere Dimension des Quellenkontextes, die diskutiert werden muß, um beurteilen zu können, ob anhand einer Quelle ein theoretisch gemeinter Sachverhalt indiziert werden kann, betrifft das Verhalten, den Verwaltungsstil derjenigen Individuen, die die Quelle erstellt haben. Aus einer Untersuchung über den Widerstand gegen das NS-Regime in Düsseldorf wissen wir über die Indikatorenqualität von Vernehmungsprotokollen der Gestapo, daß z.B. 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. (13)

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 aktiver Instanzen verhielten. Das Verhalten auch derjenigen, über die Informationen erhoben wurden, spielt eine Rolle. 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.

### 3.3. VARIABLENKOMMENTARE IN DER HISTORISCHEN SOZIALFORSCHUNG

In der empirischen Sozialforschung enthalten Kodexbücher über die Bezeichnung der Variable und über die Bezeichnung bzw. das Etikett der Variablenausprägung normalerweise keine weiteren Informationen zum inhaltlichen "Verständnis" der Variable. Die Wiedergabe der Variablenbezeichnung, in der Umfrageforschung mit der einzelnen Frage des Fragebogens identisch, wird häufig noch als ausreichend angesehen; entsprechen sich doch Frageformulierung und Bezugssystem des Alltags. Dieses Selbstverständnis der Umfrageforschung wird schon brüchig, sobald Sekundäranalysen von Umfragen z.B. der frühen 50er Jahre durchgeführt werden sollen. 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. Solche Kommentare können ein breites Spektrum von Informationen enthalten. Erläuterungen von Berufsangaben sind das prominenteste Beispiel aus der historischen Sozialforschung; allgemein läßt sich als Regel formulieren, daß in der historischen Sozialforschung das zeugenössische Sinnverständnis von Tatbeständen festgehalten werden muß, die sich in Variablen "widerspiegeln".

TEIL II:

EMPFEHLUNGEN FÜR EINE VERBESSERUNG DER DATENBESCHREIBUNG  
UND DATENDOKUMENTATION IN DER HISTORISCHEN SOZIALFORSCHUNG

1. VORBEMERKUNG

Im vorausgegangenen Teil ist grundsätzlich der Datenbeschreibungs- und Datendokumentationsbedarf formuliert worden, der akut wird, sobald im Rahmen quantitativer historischer Forschungen maschinenlesbare historische Daten anfallen. 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; ohne Datenbeschreibung und Datendokumentation kein "Verständnis" dieser Daten. 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 von drei Jahren 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, somit auch die Verwendung von Forschung für Forschung in einem praktischen Sinne zu verbessern.

2. ZUM HINTERGRUND DER VORLIEGENDEN EMPFEHLUNGEN

Die vorliegenden Empfehlungen "speisen" sich aus zwei Quellen: Zum einen die Erfahrungen, die durch mehrjährige Beobachtung der Forschung gemacht worden sind, Beobachtungen, die sich nicht allein aus der Archivierung und Bereitstellung maschinenlesbarer historischer Forschung sowie aus der Beratung der Forschung resultieren, sondern zum anderen aus der Arbeit des Zentrums bei der jährlichen Dokumentation von Arbeiten aus der historischen Sozialforschung und aus der Kenntnis der Erfordernisse an eine adäquate Datenbereitstellung für Lehrerfordernisse stammen. Daneben stellt auch die Einbindung der Bereitstellungen maschinenlesbarer historischer Forschungsdaten in einen internationalen Rahmen Erfordernisse an die Datenbeschreibung und Datendokumentation, insbesondere Kompatibilitäts-erfordernisse. Aus der ersten Quelle, der Beobachtung und Beratung der Forschung im weitesten Sinne, resultieren die inhaltlichen Merkmale, die zur Verwendung für die Beschreibung und Dokumentation der Daten der historischen Sozialforschung hier vorgeschlagen werden. Die Strukturierung der Beschreibungs- und Dokumentationsitems folgt einem vor einiger Zeit von den westeuropäischen Archiven entwickelten Instrument zur Beschreibung und Dokumentation von Umfragedaten. In den letzten Jahren ist dieses Instrument auf Anregung des Zen-

trums hin erweitert worden, um Bestandteile, die es ermöglichen, auch maschinenlesbare historische bzw. prozeß-produzierte Daten zu beschreiben und zu dokumentieren. Somit steht für die Forschung ein Instrument zur Verfügung, das es erlaubt, Datenbeschreibung und Datendokumentation so zu gestalten, daß maschinenlesbare historische Daten aufgrund standardisierter Informationen auch international verständlich werden können. Standardisierung bringt jedoch nicht immer Vorteile, sondern wirft häufig auch eine Reihe von Problemen auf, so auch in diesem Fall: Es ist nicht auszuschließen, daß durch die Standardisierung des Beschreibungsinstruments speziellen Beschreibungsproblemen nicht mehr ausreichend Rechnung getragen werden kann. Diese wird jedoch teilweise auch dadurch ausgeglichen, daß ein "Verständnis" von Daten auch im internationalen Kontext erreicht wird. Die Nachteile der Standardisierung werden auch dadurch gemildert, daß zwar ein Raster vorgegeben wird, das abgefragt wird, die Antworten aber durchaus "open ended" sein können.

### 3. 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 historischen Sozialforschung definiert worden sind. Dazu gehören Aussagen über

- A. Die Ziel- bzw. Grundgesamtheit, auf der die Aussagen einer Untersuchung bezogen sind, und die evtl. Auswahl bzw. Stichprobe, die aus einer derartigen Untersuchung gezogen worden ist.
- B. Die erfaßten Inhalte (Variablen) einer Untersuchung.

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 im ursprünglich als Instrument zur 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, insofern 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 Beobachtungs- bzw. Erhebungseinheit definiert. Zur Berichterstattung des Wissenschaftlers in der historischen Sozialforschung gehört es, exakte Explikationen der Beobachtungs- bzw. 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 Fragestellungen 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 sekundäranalytischer Nutzungen auch für andere Ziel- bzw. Grundgesamtheiten repräsentativ sein können als für die, die im Rahmen der Originalerhebung definiert worden ist. 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 der grundsätzlichen Probleme der Abbildung einer bestimmten 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. Auf deren Grundlage 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. Ein Archivar, der in seiner alltäglichen Arbeit weitaus intensiver mit den Problemen der Quellenüberlieferung befaßt ist, wird die hier vorgeschlagenen Kategorien sicherlich als nicht ausreichend empfinden. Demgegenüber muß aber erneut 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 sehr 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 angebar sind,
- oder ob die Quelle vielleicht teilzerstört ist, weil der Bestand aufgrund äußerer Einwirkungen (Feuer, Krieg, sonstige physische Beeinträchtigungen) beschädigt ist.

Neben der Beantwortung der Frage, in welchem Maße die verwendete Quelle erhalten ist, gilt es, auch die Frage zu beantworten, die die interne Organisation des Quellenmaterials betreffen. Die Abbildung einer Grundgesamtheit in den Quellen ist auch dadurch bedingt, 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 können, verlangt über die Informationen über Ausmaß und Art der Überlieferung hinaus auch Angaben über die

Dokumentation und den Zugang zur Quelle. Eine unzulängliche Quellendokumentation, 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.

#### 4. BESCHREIBUNG DER INHALTE DER FORSCHUNG

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 Problemereiche, 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 die folgenden 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. sollte 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. Dementsprechend sollten die Ergebnisse der quellenkritischen Diskussion, die derartige Fragen anschnitten, hier dokumentiert werden.

Der erweiterte Codeplan ist für die historische Sozialforschung unerlässlich, weil - zumindest gegenüber dem bislang noch dominierenden Selbstverständnis der empirischen Sozialforschung - ein größerer Beschreibungs- und Dokumentationsbedarf auf der Ebene der Variablen besteht, als dies zur Zeit noch für die empirische Sozialforschung geltend gemacht wird. Die Begründung für diesen aus-

fürlichen Variablenkommentar ist weiter oben ausgeführt worden und kann hier entfallen. Deshalb sollen an dieser Stelle nur die Informationen formuliert werden, die geeignet erscheinen, eine Variable aus dem Bereich der historischen Sozialforschung adäquat zu "verstehen":

Codebuchinformationen 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 derjenigen amtlichen Zählungen einzubeziehen, auf deren Grundlage diese Daten erstellt worden sind. Ohne diese zusätzlichen Angaben sind Aggregatdaten nicht "verständlich".

Die Quelleninformation pro Variable hat Kontrollfunktionen, sie dient dazu, bei Datenbereinigungen die wiedergegebenen Variablen evtl. anhand der Quelle zu überprüfen. (14)

##### 5. KURZFASSUNG DER EMPFEHLUNGEN FÜR EINE VERBESSERUNG DER DATENBESCHREIBUNG UND DATENDOKUMENTATION IN DER HISTORISCHEN SOZIALFORSCHUNG

A) Beschreibungen 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 Ihre 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 in 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 Osnabrück 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 Quellmaterial bedingt sind) die Zielgesamtheit abbilden.

Um weiterhin im Rahmen eines unserer Beispiele zu argumentieren: Bilden Adressbücher die Bevölkerung der Stadt Osnabrück 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(15):

"Die 'Personalakten' der Gestapoleitstelle Düsseldorf waren nach dem Stand vom 1. Mai 1977 zu über zwei Dritteln in mehreren Karteien systematisch erschlossen. In einer 'Ortskartei Düsseldorf' waren ca. 5000 Karteikarten über Personen gesammelt, die zwischen 1933 und 1945 in Düsseldorf 'staatspolizeilich 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 Staatspolizeiämtes in Berlin vom 1.7.1939 untergliedert (vgl. Tabelle 2). 41 Schlagwörter wurden in die Erhebung einbezogen, elf blieben dagegen unbearbeitet. 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 sind, bildete die Auswahl-einheit, während eine Person die Analyseinheit darstellt. Die 41 'Hauptschlagwörter' wurden in vier Kategorien gruppiert und für jede Kategorie Quoten festgelegt; als Stichprobenumfang ergab sich eine Fallzahl von 825 Personen (vgl. Tabelle 3).

Im Zuge einer ersten Auswahlstufe wurden 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 Auswahlen 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."

A5) Angaben über das "Überleben" der Quelle:

Hier sind Angaben darüber zu machen, ob die Quelle eventuell vollständig erhalten ist, oder ob sie eventuell teilkassiert oder teilzerstört 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 Teilzerstörungen 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 Stichprobenziehung 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 von Inhalten (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 Kodes, 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 Bezeichnungen der Variablen und der Variablenausprägung so detailliert wie möglich. Machen 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. Randauszahlung, 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 Positionen 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 zeugenössischer Beobachter an, Robert/Couigny werten zusätzlich Abstimmungen 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 PP1 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."(16)

#### ANMERKUNGEN

- 1 So in den prominentesten Darstellungen:  
Michael Drake, *The Quantitative Analysis of Historical Data*, Milton Keynes: The Open University Press 1974.  
Charles M. Dollar und Richard J. Jensen, *Historians Guide to Statistics. Quantitative Analysis and Historical Research*, New York: Holt, Rinehart and Winston 1971.  
Roderick Floud, *An Introduction to Quantitative Methods for Historians*, London: Methuen 1979 (2nd. printing), deutschsprachige Ausgabe: Einführung in quantitative Methoden für Historiker, Stuttgart: Klett-Cotta 1980.  
Jean Heffer, Jean-Louis Robert und Pierre Saly, *Outils Statistiques pour les Historiens*, Paris: Publications de la Sorbonne 1981.  
Norbert Ohler, *Quantitative Methoden für Historiker*. München: C.H. Beck 1980.  
Edward Shorter, *The Historian and the Computer*. Englewood Cliffs: Prentice Hall 1971.  
Neuerdings auch: Manfred Thaller, *Einführung in die Datenverarbeitung für Historiker*. Köln/Wien: Böhlau 1982.  
Auf den Stellenwert von Datenbeschreibung und Datendokumentation in Lehrpaketen verweist Henning Bauer in seinem Beitrag in diesem Heft.
- 2 Vgl. dazu Walter Eckermann et al. *Einführung in das Studium der Geschichte*, Berlin (DDR): Wissenschaftsverlag, 3. Aufl. 1979 und grundsätzlich: Ewald Standop, *Die Form der wissenschaftlichen Arbeit*, Heidelberg: UTB, 6. Aufl. 1975.
- 3 Obwohl Datenbeschreibung und Datendokumentation in kaum einem Lehrbuch der empirischen Sozialforschung explizit angesprochen werden, gehören die beiden hier genannten Beschreibungskomplexe zum Regelkodex der empirischen Sozialforschung.
- 4 In der bekanntesten deutschsprachigen Darstellung des Forschungsprozesses werden Datenbeschreibung und Datendokumentation explizit nicht behandelt: Heine von Alemann, *Der Forschungsprozeß*. Stuttgart: B.G. Teubner 1977.
- 5 Vgl. insbesondere: 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. Dort auch weitere Literaturhinweise.

- 6 Dieses Instrument wurde von westeuropäischen Datenarchiven unter maßgeblicher Beteiligung des Zentralarchivs für empirische Sozialforschung der Universität zu Köln entwickelt. Auf Anregung des Zentrums für historische Sozialforschung hin wurde dieses Instrument um Kategorien zur Beschreibung historischer und prozeß-produzierter Daten erweitert.  
Vgl. dazu auch: Herbert Reinke, Towards Standards for the Description of Machine-Readable Historical Data, in: Historical Social Research, No. 18/April 1981, S. 3-10. Bei der Entwicklung dieser Kategorien ist ein ehemaliger Mitarbeiter des Zentrums, Herbert Küppers, maßgeblich beteiligt gewesen.
- 7 Grundsätzliche Überlegungen zu der in diesem Abschnitt behandelten Problematik stellt Harald Rohlinger in seinem Beitrag "Quellen als Auswahl - Auswahl aus Quellen an".
- 8 Die folgenden Überlegungen resultieren vor allem aus der Beschäftigung mit der Kriminalstatistik des Deutschen Reiches, die im Rahmen eines Forschungsprojektes im Zentrum für historische Sozialforschung quellenkritisch bearbeitet und maschinenlesbar aufbereitet werden.
- 9 Erdmann Weyrauch, Datenverarbeitung als Quellenkritik, in: Paul J. Müller (Hrsg.), Die Analyse prozeß-produzierter Daten, Stuttgart: Klett-Cotta 1977, S. 141-178.
- 10 Die folgenden Ausführungen profitieren außerordentlich auch von den Arbeiten und Forschungen von Wolfgang Bick, Paul J. Müller und Reinhard Mann über Probleme der Analyse prozeß-produzierter Daten und der Aktenanalyse. Die "Essenz" der Überlegungen dieser drei Kollegen findet sich in dem von ihnen herausgegebenen Sammelband "Sozialforschung" und Verwaltungsdaten, Stuttgart: Klett-Cotta (erscheint 1983).  
Vgl. dazu auch den Beitrag von Paul J. Müller in diesem Heft.
- 11 Erhard Blankenburg, Die Aktenanalyse, in: E. Blankenburg (Hrsg.), Empirische Rechtssoziologie, München: Piper 1975, S. 193-198.
- 12 Edward Shorter, "La vie intime". Beiträge zu seiner Geschichte am Beispiel des kulturellen Wandels in den bayerischen Unterschichten im 19. Jahrhundert, in: Peter Christian Ludz (Hrsg.), Soziologie und Sozialgeschichte, Opladen: Westdeutscher Verlag 1973, S. 530-529.
- 13 Bei dieser Untersuchung handelt es sich um: Reinhard Mann, Non-konformes Verhalten und politische Kontrolle im nationalsozialistischen Deutschland. Düsseldorf im Dritten Reich. Die Arbeit des im vergangenen Jahr verstorbenen Reinhard Mann erscheint demnächst posthum.
- 14 Dieser Zweck der Quelleninformation im Variablenraster ist bereits früher von Thomas Schuler in seinem Aufsatz "Quantifizierung und Geschichtsstudium" angesprochen worden. Der Aufsatz ist erschienen in: Historical Social Research, No. 17, Januar 1981, S. 55-73.

- 15 Die Beschreibung der Vorgehensweise bei der Stichprobenziehung stammt aus der in Anmerkung 14 genannten Untersuchung von Reinhard Mann.
- 16 Der wiedergegebene Variablen-Kommentar stammt aus der Datenbeschreibung der folgenden Untersuchung: Heinrich Best, Die Abgeordneten der Assemblée Nationale Constituante 1848. Sozialprofil und Legislatives Verhalten/Datenhandbuch. Köln: Zentrum für historische Sozialforschung 1982. Zentrum-Report No. 13.