

### Focus im Rückblick: Massenakten als Datenbasis der empirischen Sozialforschung: methodische Voraussetzungen und institutionelle Erfordernisse

Bick, Wolfgang; Müller, Paul J.

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:

Bick, W., & Müller, P. J. (2002). Focus im Rückblick: Massenakten als Datenbasis der empirischen Sozialforschung: methodische Voraussetzungen und institutionelle Erfordernisse. *Historical Social Research*, 27(2/3), 227-252. <https://doi.org/10.12759/hsr.27.2002.2/3.227-252>

#### 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>

Focus im Rückblick: Massenakten als Datenbasis der empirischen Sozialforschung. Methodische Voraussetzungen und institutionelle Erfordernisse

*Wolfgang Bick und Paul J. Müller\**

1. Einführung

Zunehmend wird in der Sozialforschung das einst eindeutig dominierende Instrument, die Umfrage, ergänzt durch Verfahren der Datengewinnung, die auf nicht-selbst-erhobene Daten zurückgreifen. Bei diesen fremdproduzierten Daten spielen neben den Eigenerhebungen Statistischer Ämter administrative Daten die größte Rolle. Bei der Verwendung von Verwaltungsdaten oder prozeß-produzierten Daten – und hiermit sind die Daten gemeint, die als Aufzeichnungen öffentlicher und privater Organisationen im Rahmen ihrer Tätigkeit und nicht nur zum Zwecke wissenschaftlicher bzw. statistischer Auswertung gesammelt werden bzw. wurden – treten jedoch besondere Probleme auf, die gerade Folge dieser Entstehungsbedingungen sind.

Sie werden eben nicht im Rahmen von Datensammlungen erhoben, die vom Sozialforscher nach eigenen Bedürfnissen konzipiert und mit Hilfe von sozialwissenschaftlichen Methoden kontrolliert werden. Sie werden vielmehr in einem ganz anderen, „fremden“ Kontext gesammelt. Vollzugsverwaltungen

---

\* Stark gekürzter (insbesondere ohne Anmerkungen und Literatur) Wiederabdruck des gleichnamigen Beitrags in: Wolfgang Bick / Reinhard Mann / Paul J. Müller (Hg.): Sozialforschung und Verwaltungsdaten. Stuttgart 1984, S. 123-159. Dieser Sammelband basierte hauptsächlich auf einem Kolloquium, das die Arbeitsgemeinschaft QUANTUM e. V. 1980 veranstaltet hatte. Es war dies zwar das erste Treffen zwischen Archivaren und Sozialwissenschaftlern – hat jedoch seine Vorgeschichte einerseits in einem Gutachten aus dem Jahre 1965 (dem sogenannten Döll-Gutachten), andererseits durch Gründung von QUANTUM im Jahre 1975, in der neben Historikern und Sozialwissenschaftlern auch Archivare organisiert sind.

generieren diese Daten anläßlich der Abwicklung von Verwaltungsvorgängen, nicht jedoch zum Zwecke der Forschung oder Statistik.

Massenakten sind ein Teil dieser prozeß-produzierten Daten, die jedoch aufgrund bestimmter Struktureigenschaften einen besonderen Stellenwert im Rahmen von quantitativen historisch-sozialwissenschaftlichen Forschungen haben. Massenakten zeichnen sich durch eine besondere Art der Orientierung von Verwaltungen aus:

- Im Gegensatz zu Sachakten, in denen eine Mehrzahl von Objekten und die Relationen zwischen diesen Objekten wiedergespiegelt werden, sind Massenakten nach einzelnen Fällen organisiert. (Dies ist u. a. eine Bedingung für die ansonsten unwichtige Eigenschaft „massenhaft“ zu sein.)
- Massenakten sind dadurch gekennzeichnet, haben als Struktureigenschaft, daß erst einmal der Objektbereich der Verwaltung in Teile zerlegt wird: die einzelne natürliche oder juristische Person; das einzelne Objekt (Haus, Grundstück, KfZ etc.)
- Das Verhältnis zum einzelnen Teil (Objekte, Personen) ist dabei dadurch gekennzeichnet, daß eine Registrierung der „Klienten“ auf eine längere Dauer für nötig erachtet wird. (Dies im Gegensatz zu Verwaltungen, die nur anonyme Kunden, nicht aber namentlich festgemachte Klienten haben: z. B. Museum, Badeanstalten, Infrastrukturen.)
- Die Vorab-Zergliederung des Objektbereichs der Verwaltungstätigkeit ist jedoch nicht notwendigerweise mit einer Atomisierung identisch.

Die Nutzung von prozeß-produzierten Daten – also auch von Massenakten – für Forschungszwecke bedeutet, daß mit Material gearbeitet wird, in das faktisch noch eine unbekannte Zahl von Störfaktoren einwirkt. Erledigungsstrategien von Sachbearbeitern, der Legitimationscharakter von Akten, Verzerrungen durch das Abfassen von Schriftstücken in ihrer Wirkung auf übergeordnete Stellen, Informationsstrategien des Klienten: Dies sind nur einige der Faktoren, die dazu führen, daß bei der Nutzung häufig unklar bleibt, welche Aussagequalitäten aber auch welche Begrenztheiten prozeß-produzierte Daten haben.

In den vergangenen Jahren gab es immer wieder Versuche, diese Unklarheiten beim Umgang mit prozeß-produzierten Daten durch die Perfektionierung des Verwaltungsvollzugs gemäß den Wünschbarkeiten der Statistik, Forschung und Planung oder durch die Erfassung von zusätzlichen Informationen zu vermindern, die für planerische und statistische Zwecke besser geeignet sind. Diese sind zumeist an der Eigendynamik der Vollzugsverwaltungen gescheitert. Und seitdem es Datenschutz als gesetzliche Regelung gibt, ist es noch weitaus schwieriger geworden, den Verwaltungsvollzug für andere Zwecke etwas erheben zu lassen, was nicht für den Vollzug selbst von Interesse ist.

Diese geringe Beeinflußbarkeit der Vollzugsdaten selbst verstärkte jedoch das Interesse, gerade das in der Verwendung dieser Daten ausgleichen zu wollen, was in ihrer Entstehung als unvermeidlich hingenommen werden muß: ihre

spezifischen Defekte. Hierfür liegen jedoch noch zu wenig systematische Kenntnisse über diejenigen Einflußfaktoren vor, die die Aussagekraft prozeß-produzierter Daten beeinflussen.

Es wird im folgenden aufgezeigt werden,

- in welcher Weise heute mit prozeß-produzierten Daten gearbeitet wird,
- welchen methodischen Problemen der Indikatorenqualität sich der Sozialforscher gegenüber sieht,
- welche zusätzlichen Informationen er für eine angemessene Nutzung benötigt und
- welche Maßnahmen für eine verbesserte Versorgung der sozialwissenschaftlichen Forschung mit Verwaltungsdaten bzw. Massenakten sinnvoll und notwendig sind.

## 2. Die zunehmende Bedeutung der Analyse prozeß-produzierter Daten für die sozialwissenschaftliche Forschung

Längere Zeit wenig beachtet, dann aber um so deutlicher wahrgenommen, hat sich die akademische sozialwissenschaftliche Forschung in einem ebenfalls noch unbekanntem Zeitraum gewandelt. Noch vor zehn Jahren wäre auf die Frage „Was kennzeichnet methodisch die Sozialforschung?“ die Antwort „Sie ist Umfrageforschung“ ausreichend gewesen. Sie ist in der Tat auch heute (1980) noch sehr stark Umfrageforschung; als Antwort auf die gestellte Frage wäre dennoch diese Antwort nur noch die halbe Wahrheit. Die sozialwissenschaftliche Forschung hat sich von einer ehemals eindeutig durch Umfrageforschung gekennzeichneten Vorgehensweise zu einem Wissenschaftsbereich entwickelt, der zunehmend durch Methodenpluralismus zu charakterisieren ist.

Dieser Methodenpluralismus geht einher mit Verschiebungen in der Zusammensetzung der Datenbasis, die sich heute durch eine stärkere Mischung verschiedener Datentypen kennzeichnen läßt. Neben den durch eigene Befragungen und andere Verfahren selbst erhobenen Daten werden verstärkt fremdproduzierte Daten als Forschungsdaten eingesetzt.

Prozeß-produzierte Daten stellen eine bedeutsame Teilmenge dieses fremdproduzierten Materials dar. Prozeß-produzierte Daten unterscheiden sich von den eigenen Erhebungen der Sozialforschung dadurch, daß sie eben nicht im Rahmen einer Datenerhebung produziert werden, die auf der Grundlage des sozialwissenschaftlichen methodischen Instrumentariums konzipiert, durchgeführt und kontrolliert wurde. Sie sind vielmehr Nebenprodukte von Tätigkeiten, die in einem organisationellen Kontext nach anderen Gesetzmäßigkeiten ablaufen.

Die Gründe für die verstärkte Nutzung fremdproduzierten Materials und insbesondere von prozeß-produzierten Daten sind vielfältig. Zum einen erfolgten Akzentverschiebungen in den Fragestellungen der empirischen Sozialforschung, die nicht mehr nur allein durch die Untersuchung repräsentativer Querschnitte der Gesamtbevölkerung angemessen angegangen werden konnten, sondern durch Untersuchungen spezieller Populationen beantwortet werden mußten (Minoritäten, „soziale Probleme“). Dies war dann jedoch oftmals gleichbedeutend damit, daß Daten von Verwaltungen wichtig wurden. Denn die Subgruppen solcher Untersuchungen werden oftmals durch Verwaltungskategorien wie z. B. Schüler der Schule X, Insasse der Anstalt Y oder Klient von Z bestimmt.

Hinzu kam ein zunehmendes Interesse an der Wirkungsweise von Institutionen in der Gesellschaft, das zur verstärkten Nutzung der Buchführungsdaten dieser Institutionen als Datenbasen für Forschungszwecke führte. Zum anderen wurden zunehmend die Techniken der quantitativen Sozialforschung in früher mehr qualitativ arbeitenden Disziplinen und Wissenschaftsbereichen angewandt, Forschungsrichtungen, für die immer schon – wie z. B. bei der Kriminologie – Daten von Verwaltungen die wichtigsten Unterlagen waren.

Die Möglichkeiten der sozialwissenschaftlichen Forschung, auf große Mengen prozeß-produzierter Daten ohne sehr große Aufbereitungsarbeiten zurückzugreifen, vergrößerten sich zudem in dem Maße, wie die Verwaltungsautomation voranschritt. Hinzu traten methodisch-technische Entwicklungen zur Analyse von Daten, die selbst massenhaft gleichförmige Daten analysierbar werden ließen. Schließlich mußten bei der Methode der direkten Befragung von Personen ein zunehmender Verfall an Sorgfalt in der Beantwortung von Fragen und zunehmend Probleme konstatiert werden, gleichbleibend eine hohe Ausschöpfungsquote zu realisieren; Prozesse, die nach amerikanischen Erfahrungen weniger durch die gestiegene Sensibilisierung in Fragen des Datenschutzes als durch hohe Mobilität und Schwierigkeiten bedingt sind, Personen zu Hause anzutreffen.

Es kann somit eine zunehmende Bedeutung fremdproduzierten Materials für die empirische sozialwissenschaftliche Forschung festgestellt werden. Die Ausdehnung der Staatstätigkeit und als Folge davon der sozialen Buchführung des Institutionengeflechts hat Konsequenzen für die Vielfalt der Daten gehabt, die einer sozialwissenschaftlichen Analyse zugänglich wurden. Trifft diese Angebotslage zugleich auf eine veränderte Nachfrage der Sozialwissenschaften nach unterschiedlichen Datentypen, so erscheint die zunehmende Bedeutung der Analyse fremdproduzierten Materials allerdings nicht mehr überraschend: In Gesellschaften mit zunehmender Bedeutung von Institutionen können Sachverhalte nicht mehr allein bei Individuen erhoben werden.

Diese Bestimmungsgründe wirkten nebeneinander, teilweise auch kumulativ, wie in denjenigen Forschungsbereichen besonders deutlich wurde, deren Objektbereich durch den Wirkungsbereich von formalen Institutionen einge-

grenzt wird, wie z. B. in der Kriminologie/Kriminalsoziologie, Medizinsoziologie, Bildungsforschung und der Rechtssoziologie, und die sich deshalb durch eine besonders häufige Verwendung von prozeß-produzierten Daten kennzeichnen lassen.

### 3. Methodische Probleme bei der Nutzung prozeß-produzierter Daten

Um aufzeigen zu können, welchen methodischen Problemen sich sozialwissenschaftliche Nutzer prozeß-produzierter Daten gegenübersehen, wurden rd. 30 Zeitschriften ausgewählt, in denen empirische Arbeiten mit prozeß-produzierten Daten während der letzten fünf Jahre (1974-1978 inkl.) veröffentlicht wurden. Die einzelnen Beiträge wurden auf die institutionelle Herkunft der Daten, auf die Nutzung als Individual- oder Aggregatdaten und auf die berichteten methodischen Probleme hin untersucht.

Es stand zu vermuten, daß die Nutzung von Individualdaten problembewußter gestaltbar ist als der Rückgriff auf Tabellen und Statistiken. Bei Individualdaten ist es weitaus wahrscheinlicher – und aufgrund der Auswertung der Literatur eindeutig festgestellt worden –, mit den datenproduzierenden Stellen in Kontakt zu kommen. Hieraus folgt dann die Chance, daß die Fragen der Datenqualität in den Gesprächen mit den „Beamten vor Ort“ thematisiert werden.

Die These, daß die Bearbeitung von Mikrodaten das Methodenbewußtsein fördert, wurde überprüft. Grundlage hierzu war ein Katalog von methodischen Problemen, die sich drei Problemgruppen zusortieren lassen:

- 1) Aussagen über den Aktengenerierungsprozeß: Hierunter werden alle Probleme erfaßt, die als Folge von Selektions- und Definitionsprozessen durch das Umfeld der Klienten im Vorfeld des behördlichen Kontaktes oder durch den Aktenproduzenten im Kontakt selbst auftreten. Wichtige Faktoren sind hierbei unter anderem das normativ festgelegte Entscheidungsprogramm, Sachbearbeiterstrategien, das Zusammenwirken mehrerer Organisationen bei der Buchführung und das Informationsverhalten der Klienten selbst.
- 2) Aussagen über die Abweichung zwischen Abbildern und Abgebildeten: In dieser Gruppe werden meßtheoretische Probleme zusammengefaßt, z. B. Aussagen über Dunkelziffern oder über die Differenzen zwischen Sichtweisen der Publika und des Behördenpersonals, über die Gültigkeit und Zuverlässigkeit der Verwaltungsdaten bzw. die Differenzen zwischen den Meßskalen verschiedenster Verwaltungen.
- 3) Methodisch-technische Probleme der Nutzung prozeß-produzierter Daten: Akten sind häufig komplex aufgebaute Zusammenstellungen von

Unterlagen aus verschiedensten Quellen, die zu verschiedenen Zeiten entstanden sind. Auch die Systeme der Aktenhaltung folgen unter Umständen Prinzipien, die Schwierigkeiten bei Stichprobenziehungen mit sich bringen. Neben der komplexen Struktur des Materials und den Problemen bei der Stichprobenziehung werden in dieser Gruppe auch Zugangsprobleme, das Problem der Datenanreicherung bzw. der Datenüberprüfung durch Mehr-Methoden-Ansätze sowie die Fragen zusammengefaßt, die als Folge komparativer Designs auftreten.

In den Arbeiten, in denen auf methodische Probleme Bezug genommen wird, werden somit sowohl das Verhalten der Klienten als auch das der Behörden als Faktoren verstanden, die die Datenentstehung beeinflussen. Somit stellt nicht nur eine Instanzenforschung das notwendige Wissen für eine Datenkunde prozeß-produzierter Daten bereit. Ebenso wichtig hierfür wird die am Klienten orientierte Sozialforschung zum Thema Bürger und Verwaltung.

Trotz dieser relativ hohen Prozentsätze, die Problembewußtsein indizieren, bleibt – bezogen auf die Grundgesamtheit aller erfaßten Aufsätze – die Berücksichtigung der Entstehungsbedingungen prozeß-produzierter Daten bei der Nutzung gering. Zwar ist für verschiedene Forschungsbereiche ein unterschiedliches Problembewußtsein zu konstatieren. Wichtiger jedoch ist der Befund, daß die Berichterstattung keinen Standards folgt, die Vergleiche über die einzelnen Arbeiten hinweg gestatten würden. Abhängig vom Vorwissen des Forschers und seinen Fragestellungen werden – wenn überhaupt – unsystematisch Einzelaspekte hervorgehoben, die die Entstehung des Materials kennzeichnen. Während sich reifere Disziplinen u. a. dadurch auszeichnen, daß Wiederholungen von Bekanntem durch Verweise auf andere Veröffentlichungen („Klassiker“) vermieden werden, werden die Probleme bei der Nutzung prozeß-produzierter Daten immer wieder neu „entdeckt“ und diskutiert.

Die bislang fehlenden Standards zur Beschreibung von prozeß-produzierten Daten erschweren noch eine systematische Kennzeichnung des Materials, das als Datenbasis Eingang in eine größer werdende Zahl von Forschungsarbeiten gefunden hat. Die Einzelbefunde in einer Vielzahl von Veröffentlichungen stellen jedoch das Rohmaterial dar, das Hinweise darauf geben kann, in welcher Weise Beschreibungen prozeß-produzierter Daten angelegt werden sollten. Darüber hinaus gibt es für einzelne Forschungsbereiche einige wenige Versuche, die bisherigen Befunde systematisch zusammenzustellen. Auch in einer Vielzahl von Experteninterviews ist die Notwendigkeit dieser Systematisierung immer wieder betont worden. Die Erfolgchancen für solche Unternehmungen wurden zum Teil optimistisch eingeschätzt. Es wurde jedoch immer wieder darauf hingewiesen, daß solche Beschreibungen umfassender angelegt sein müßten, als dies für die Umfrageforschung heute schon geschieht. Während bei einer Umfrage eine Vielzahl von Faktoren kontrollierbar ist, kann der Nutzer von Vollzugsdaten auf die Produktionsbedingungen keinen Einfluß nehmen. Die Heterogenität der Entstehungskontexte – von der nahezu unver-

bindlichen „Befragung“ im Kontakt mit dem Einwohnermeldeamt bis zu den Protokollen von Verhören im Bereich der Justiz – macht es notwendig, Zusatzinformationen zu einer Reihe von höchst unterschiedlichen Aspekten bereitzustellen.

Im folgenden soll ein Beitrag zu einer Systematisierung bislang vorliegenden Wissens geleistet werden. Standards für die Beschreibung von prozeß-produzierten Daten und ihre Anwendung in der Forschung stellen sicher, daß Einzelbefunde eingeordnet werden können, sich die Chancen für eine Kumulation des Wissens über die Entstehungsbedingungen prozeß-produzierter Daten erhöhen und mit prozeß-produzierten Daten trotz ihrer Fehlerhaftigkeit ohne Fehler gearbeitet werden kann.

#### 4. Ein Bezugsrahmen

Nach wie vor ist die Methode der Befragung das am meisten eingesetzte Instrument der empirischen Sozialforschung. Auf den ersten Blick erscheint das Interview zur Abbildung alltäglichen Verhaltens als die adäquate Methode der Datensammlung. Die großen Ähnlichkeiten zwischen Alltagsgesprächen und Interviewsituationen legen zumindest diese Vermutung nahe. Diese vorwissenschaftlichen Überlegungen haben sich aber nur zum Teil bewährt. Heute liegen nämlich eine fast unübersehbare Zahl von methodologisch orientierten Untersuchungen vor, die deutlich machen, daß „Interviews eine sehr künstliche Vorgehensweise unter Benutzung äußerlicher Übereinstimmung zu alltäglichem Verhalten sind“. Inzwischen ist aber eine Kunstlehre des Interviews entwickelt worden. Hier werden die bisherigen empirischen Befunde zusammengestellt, die aufzeigen, in welcher Weise sich Frageformulierungen, verschiedene Frageformen und die Stellung der Frage im Interview auf die Gültigkeit der Antworten auswirken. Der Interviewer bzw. der Befragte als Fehlerquellen werden in einer Lehre vom Interviewen bzw. vom Befragten erfaßt. Die Probleme mit „schwierigen Fragen“ und die Empfindlichkeit des Befragten gegenüber bestimmten Themen werden im Hinblick auf die Zuverlässigkeit und Gültigkeit von Interviewdaten untersucht. Diese Kunstlehre des Interviews ist wichtiges Instrumentarium bei der Vorbereitung einer Befragung und der Entwicklung des Fragebogens.

Diese unzweifelhaften Fortschritte in der Entwicklung einer Methode mit eingebauten Kontrollmechanismen haben jedoch lange Zeit die Aufnahme vergleichbarer Entwicklungen bei denjenigen Methoden eher behindert, die für die Analyse nicht selbst erhobenen Materials und zur Abbildung anderer Aspekte der gesellschaftlichen Wirklichkeit gedacht waren. Das erneute Interesse an fremdproduzierten Daten – erneut, weil dies einmal sogar die einzige Datenbasis der europäischen Soziologie des 19. und frühen 20. Jahrhunderts

war – geht aber heute damit einher, zumindest einen vergleichbaren Stand der Beherrschbarkeit von Daten zu erreichen, wie er für selbstproduzierte Daten inzwischen erreicht werden konnte.

Auch in Verwaltungen finden täglich Tausende von „Befragungen“ statt, hier jedoch nicht mittels eines von Sozialwissenschaftlern entwickelten Fragebogens, sondern auf der Grundlage von Formularen, Vordrucken und Rechtsvorschriften. Je nach Art des Verwaltungskontaktes entstehen diese Datensammlungen jedoch in äußerst verschiedenen Kontexten. Während der Bürger bei der Erfassung von Daten im Einwohnermeldeamt als relativ Unbeteiligter agiert, der eher einer lästigen Pflicht nachkommt, ist er bei der Antragstellung auf Sozialhilfe auf die staatliche Leistung für seinen Lebensunterhalt angewiesen. Die polizeiliche Erfassung von Tathergängen vor Ort folgt anderen Regelmäßigkeiten als die Urteilsfindung von Gerichten, die Informationen auf der Basis von Indizien und komplizierten Beweisketten erfassen und auswerten. Auch die komplexen Diagnosen von Klinikärzten unterscheiden sich sehr stark von den Aufzeichnungen in Krankenblättern niedergelassener Ärzte, die in noch sehr starkem Maße die Krankheitsbeschreibungen in der Sprache des Patienten abbilden.

Schon diese wenigen Beispiele machen die Variabilität der Faktoren deutlich, die bei der Entstehung von prozeß-produzierten Daten wirksam sind. Während in der Befragungssituation der Interviewer als Partner, Interpret und Ermittler in einem vorher genau bestimmten Rahmen auftritt, der sich durch eine neutrale und unpersönliche Erfassung von Sachverhalten kennzeichnen läßt, finden sich die Klienten je nach Verwaltungsstelle in sehr unterschiedlichen „Befragungssituationen“ wieder.

Datenproduktion und Sachverhaltsermittlung in Verwaltungsstellen durch Vordrucke bzw. Formulare und Datensammlungen der empirischen Sozialforschung mittels Fragebogen sind daher nur bedingt miteinander vergleichbar. Auf vier wesentliche Unterscheidungsfaktoren sei hier nur kurz hingewiesen:

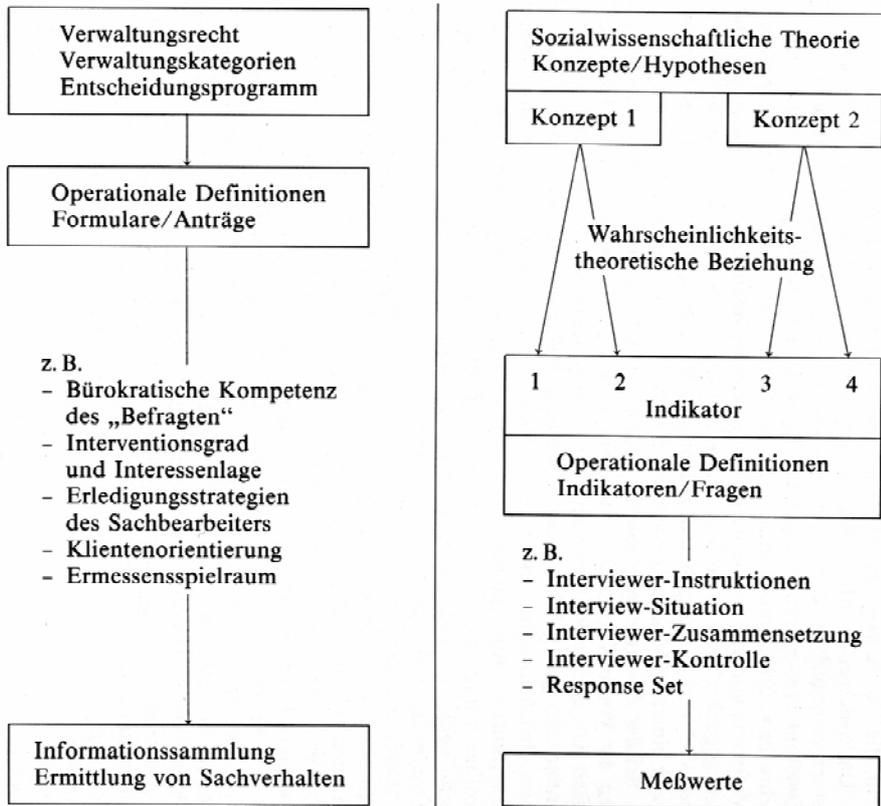
- 1) Bei einer vorgegebenen Fragestellung werden in der Umfrageforschung durch Pretests all die Faktoren bestimmt und dann auch im Erhebungsinstrument erfaßt, die für die Befragten in dieser Sache relevant sind. Die für Verwaltungsentscheidungen notwendigen Sachverhalte sind jedoch vorab festgelegt, folgen dem Prinzip der Kategorien- und nicht der Fallgerechtigkeit und sind daher indifferent gegenüber den im Alltag des Klienten bedeutsamen Faktoren.
- 2) Diese Indifferenz spiegelt sich auch in der Art und Weise wider, in der die Fragen an das Individuum gestellt werden. In der Lehre von der Frageformulierung wird untersucht, welche Fragen für den Befragten verständlich und zumutbar sind. Die Frageformulierungen sind daher der Alltagssprache angepaßt, im Gegensatz zu den an Rechts- und Verwaltungsbegriffen orientierten Antrags- und Formulartexten, die, erst durch Merkblätter ergänzt, verständlich werden sollen.

- 3) Im Interview werden die Reaktionen des Befragten auf vorgegebene Stimuli erfaßt, und diese Eins-zu-eins Widerspiegelung ist auch Ziel der Datensammlung. Verwaltungen hingegen bearbeiten den vorgegebenen Input weiter. Sie ergänzen Daten anderer Stellen, filtern und verdünnen die Angaben und entscheiden schließlich über den Fall. Prozeß-produzierte Daten werden daher wesentlich von den datensammelnden Stellen mitbeeinflußt.
- 4) Verwaltungsentscheidungen sind für die Betroffenen oftmals von großer Bedeutung. Sie sichern Lebensunterhalt oder Fort- und Weiterbildung. Der Interventionscharakter führt dazu, daß Interessen der Klienten im Kontakt bedeutsam werden, ein Faktor, den die Umfrageforschung vorab zu eliminieren versucht.

Schaubild 1 stellt zusammenfassend die Unterschiedlichkeit von Sachverhaltsermittlung in Verwaltungen und Datensammlungen durch Umfragen dar. In beiden Fällen werden vorab definierte Konzepte – Verwaltungsbegriffe einerseits, sozialwissenschaftliche Konzepte andererseits – in Meßwerte übersetzt. Die Instrumente – sei es das Formular, sei es der Fragebogen – werden jedoch in unterschiedlicher Nähe zur „Lebenswelt“ des Befragten entwickelt. Die Abstände in Schaubild 1 deuten daher auch die unterschiedlichen Distanzen an, die durch Übersetzungen zwischen Konzept und Meßwerten überwunden werden müssen. Bei prozeß-produzierten Daten wird die Übersetzung auf den Klienten überlagert. Die Indifferenz und die Asymmetrie zwischen den beiden Akteuren führen z. B. dazu, daß der Klient gezwungen ist, sich dem Amtsdeutsch anzupassen, und erfordern z. B. eine hohe bürokratische Kompetenz der Klienten. Diese bei der Datenerhebung wirksamen Faktoren führen jedoch dazu, daß die Gültigkeit der tatsächlich erhobenen Meßwerte zum Problem wird. Daten aus Umfragen hingegen belassen das Problem der Übersetzung beim Forscher. Gerade die „Nähe“ der Fragen zur Alltagssprache führt jedoch dazu, daß Indikatoren und Konzepte vom Forscher selbst in Beziehung gesetzt werden müssen, ein Problem, das schon häufig zu wissenschaftsinternen Diskussionen über die Gültigkeit der erhobenen Indikatoren geführt hat.

Bislang wurden Unterschiede in den Meßprozessen verschiedener Datentypen herausgearbeitet. Dabei wurde jeweils von Daten gesprochen – ob Verwaltungsdaten oder sozialwissenschaftliche Daten –, weil in beiden Fällen Informationen zielgerichtet anhand eines vorgegebenen Beobachtungsrasters – gesetzliche Regelungen bzw. sozialwissenschaftliche Theorien – erhoben und aufbereitet werden. Eine Nutzung von Verwaltungsdaten unter einer sozialwissenschaftlichen Perspektive bedeutet jedoch zunächst einmal, daß diese Daten für den Sozialforscher zu Rohmaterial bzw. Quellen werden, die noch in sozialwissenschaftliche Daten transformiert werden müssen. In diesem Sinne findet ein doppelter Transformationsprozeß statt: Realität wird transformiert in Daten der und für Behörden im Interaktionsfeld Bürger und Behörde. Für die sozial-

Schaubild 1:



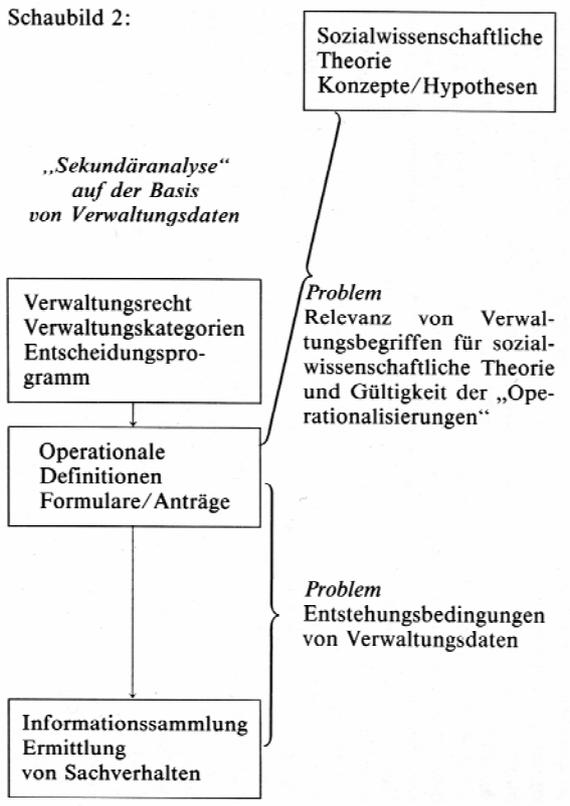
wissenschaftliche Nutzung sind diese dann zunächst einmal Quellen. Zur Umformung dieses Quellenmaterials ist ein zweiter Transformations- bzw. Interpretationsprozeß mittels Zuordnungen von Indikatoren zu Konzepten notwendig.

Primär braucht der Sozialforscher für eine angemessene Nutzung der Quellen bzw. des Materials Kenntnisse darüber, wie sich unterschiedliche Faktoren auf die Sammlung und Aufbereitung des Materials in Behörden ausgewirkt haben. Obgleich für eine solche systematische Zusammenstellung empirischer Befunde über die Entstehung der Quellen auch der Begriff Quellenkunde passen würde, werden wir im folgenden von Datenkunde sprechen und hiermit das Material unter der Perspektive der Umformung von Realität in Behördendaten beleuchten. Grund hierfür ist, daß eine Datenkunde – soll sie allgemein einsetzbar sein – unabhängig von der großen Zahl möglicher sozialwissenschaftlicher Konzepte angelegt sein muß, die Zweitnutzungen zugrunde liegen könnten. Gültigkeit und Zuverlässigkeit von Daten können daher im Rahmen einer Fehlerlehre nur an der durch das Normprogramm bzw. das Datenerhebungsprogramm intendierten Abbildung gemessen werden. Erst wenn die relevanten Störfaktoren bekannt sind, die die vom Datenproduzenten beabsichtigte Erfassung von Sachverhalten bzw. Variablen beeinflussen bzw. verzerren, wird es möglich sein, die Genauigkeit der Meßwerte für weitere Nutzungen abzuschätzen.

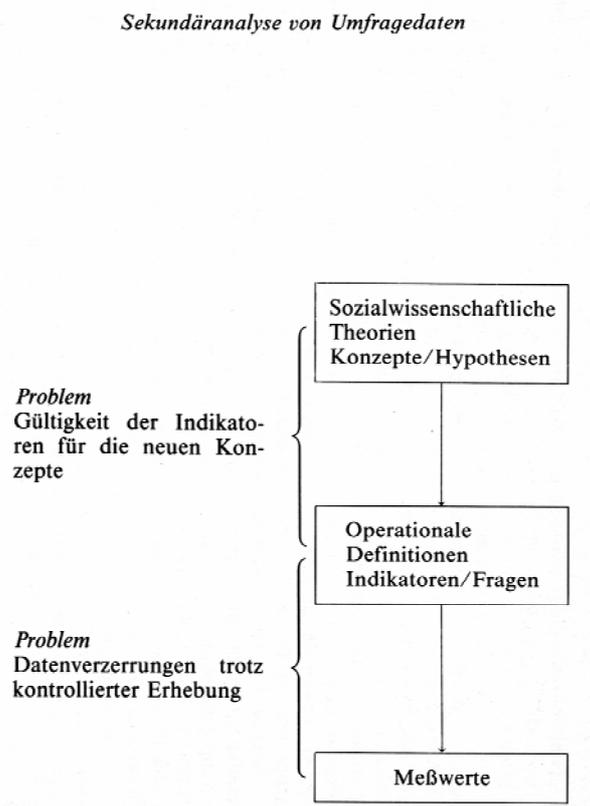
Um vorliegendes Material angemessen für eine Zweitnutzung einsetzen zu können, ist es erforderlich, die Konzepte der ursprünglichen Datenerhebung zu den der Zweitnutzung zugrundeliegenden Konzepten in Beziehung zu setzen. Die Fehlerlehre bzw. Datenkunde ermöglicht in einem ersten Schritt, die Abweichungen der Meßwerte von den im ursprünglichen Konzept intendierten abzuschätzen. Nach dieser möglicherweise sogar quantitativen Korrektur der Meßwerte kann dann aufgrund des Verhältnisses zwischen dem Konzept der Datenerhebung und dem der Zweitnutzung eine zweite, jeweils spezifische Korrektur erfolgen.

Bei der Zweitnutzung von Umfragedaten treten hierbei – trotz aller Möglichkeiten der Vorab-Kontrolle – zwei wesentliche, in Schaubild 2, linke Hälfte, schematisch dargestellte Probleme auf, die immer wieder zu wissenschafts-internen Auseinandersetzungen geführt haben. Einerseits stellt sich das Problem der Gültigkeit der vom Primärforscher erhobenen Indikatoren für die abweichenden Konzepte des Zweitnutzers. Andererseits lassen sich trotz der Möglichkeit, die Datenerhebung kontrollieren und dadurch Verzerrungen möglichst gering halten zu können, systematische Fehler bei der Messung nicht vermeiden. Diese Verzerrungen müssen bei der Zweitnutzung Berücksichtigung finden.

Schaubild 2:



*Sekundäranalyse von Umfragedaten*



Beide Problembereiche vergrößern sich jedoch, wenn die Zweitnutzung auf prozeß-produzierte Daten zurückgreift, also auf Daten, die in einem „fremden“ Bezugsrahmen erhoben wurden.

Hier ist dann dem Zweitnutzer von prozeß-produzierten Daten zunächst einmal unklar, wie die Daten im Vollzug entstanden sind. Werden ihm diese Kenntnisse nicht vermittelt, so besteht die Gefahr, daß verzerrte Daten für evident gehalten werden. Darüber hinaus stellt sich das Problem der Relevanz von Verwaltungsbegriffen für sozialwissenschaftliche oder planerische Fragestellungen und die Gültigkeit der Operationalisierungen in besonders starkem Maße.

Die Nutzung von Verwaltungsdaten bringt demnach eine Reihe besonderer Probleme mit sich: Hierzu zählen insbesondere die oftmals unbekanntenen Entstehungsbedingungen dieses Materials, die aber für eine adäquate Nutzung berücksichtigt werden müssen. Auch sind die vorliegenden prozeß-produzierten Daten häufig nur beschränkt als geeignete Meßwerte für die zu überprüfenden Hypothesen einsetzbar.

Zwei Faktoren führen jedoch dazu, daß die Nutzung dieser Daten für die sozialwissenschaftliche Forschung insgesamt erleichtert wird.

- 1) Zum einen faßt die empirische Sozialforschung Meßwerte nur als Indikatoren auf, die mehr oder weniger scharf einen interessierenden Sachverhalt abbilden. Konzepte und ihre Meßwerte stehen nur in einer wahrscheinlichkeitstheoretischen, oftmals auch nur indirekten Beziehung. Dieser Ansatz ermöglicht es, die in den Buchführungen der Verwaltung „gemessenen“ Sachverhalte zu Indikatoren für andere Sachverhalte umzudeuten, als diejenigen, die für Verwaltungszwecke ursprünglich erhoben wurden. Dies führt zwar wissenschaftsintern häufig zu Auseinandersetzungen über die Gültigkeit, also darüber, ob der Indikator den Sachverhalt begründbar mißt, den der Forscher beabsichtigt zu erfassen, schafft aber auf der anderen Seite einen Spielraum, der den Einsatz dieser Daten erleichtert.
- 2) Zum zweiten ist Sozialforschung, die auf prozeß-produzierte Daten zurückgreift, häufig Instanzen-nah orientiert, d. h. Verwaltungskategorien als solche sind Untersuchungsgegenstand und werden als abhängige oder unabhängige Variablen untransformiert in das Untersuchungsdesign eingeführt, so z. B. in Untersuchungen zu den Selektions- und Definitionsprozessen im Instanzenzug der strafrechtlichen Sozialkontrolle.

## 5. Die Rekonstruktion der Entstehungsbedingungen – zur Entwicklung einer Datenkunde

Aussagen über die Einwirkung unterschiedlichster Faktoren auf die Abbildqualität von Umfragedaten können – da die Datenerhebung vom Sozialforscher selbst konzipiert und durchgeführt wird – bei der Datenerhebung Berücksichtigung finden. Standards für die Zusammensetzung des Interviewerstabs und Aussagen über die Anordnung bestimmter Fragen im Fragebogen sind beispielsweise Kunstregeln, die vorab bedacht und einkalkuliert werden können. Die Entstehungsbedingungen sind hier weitaus kontrollierbarer als bei der Entstehung von Daten in Verwaltungen, die z. T. als reine Nebenprodukte der Abwicklung von Verwaltungsvorgängen anfallen. Die Entstehungsbedingungen von prozeß-produzierten Daten sind daher nur im nachhinein rekonstruierbar.

Wir möchten deshalb im Falle von Umfragedaten von einer *Fehlerlehre* und bei prozeß-produzierten Daten von einer *Datenkunde* sprechen. Mit dem Begriff der Fehlerlehre soll die Möglichkeit der Fehlerreduktion durch modifizierte Instrumente, durch den der Datenkunde die Notwendigkeit der Rekonstruktion des nicht (mehr) beeinflussbaren Datenerhebungsprozesses hervorgehoben werden.

Die Faktoren aufzuzeigen, die bei der Produktion prozeß-produzierter Daten eine bedeutsame Rolle spielen, ist ein wichtiger Schritt in Richtung auf eine Datenkunde für diese Datenbestände. Wird die Beherrschung dieser Faktoren Bestandteil einer Kunstlehre des Umgangs mit prozeß-produzierten Daten, so wird die adäquate sekundäranalytische Nutzung dieser Bestände erst möglich. Die Fehlerlehre für das Interview gibt hierzu jedoch nur wenig Anhaltspunkte. Dies ist insbesondere Folge davon, daß sie sich auf die vorab kontrollierbaren Randbedingungen konzentriert, also nicht auf die Rekonstruktion eines Interviewablaufs abzielt, sich in starkem Maße für die Erfassung von Einstellungen und Motivationen interessiert und als Methode der Datenerhebung eben nicht mit einer so großen Variabilität von Einfluß- und Störfaktoren konfrontiert ist.

Gleiches gilt auch für die „Historische Quellenkritik“, eine Disziplin, die wie keine andere darauf angewiesen ist, mit den schriftlichen Aufzeichnungen von Verwaltungen dennoch etwas über Wirklichkeit aussagen zu müssen. Sie beschränkte sich zumeist auf Authentizitätsprüfungen und die Ermittlung der äußeren Entstehungsbedingungen, und sie ist heute noch nicht über den bereits im vorigen Jahrhundert formulierten, wenn auch sicherlich immer noch richtigen Satz hinausgekommen: „Wir besitzen in ihnen (den geschäftlichen Akten) auf alle Fälle unmittelbare und untrügliche Zeugnisse für das, was diejenigen, welche diese Akten abfaßten, wollten, daß ihr Publikum von den Vorgängen und Verhältnissen und Motiven wissen sollte, bzw. was sie selbst davon wußten, aber wir dürfen niemals ohne Prüfung annehmen, daß wir in diesen Akten

untrügliche Zeugnisse der Vorgänge, Verhältnisse und Motive an sich besitzen“ (Ernst Bernheim). Aber gerade für diese Prüfung fehlen die Kriterien, die wesentlich mehr als nur den Einfluß der rechtlichen Rahmenbedingungen beschreibbar werden lassen.

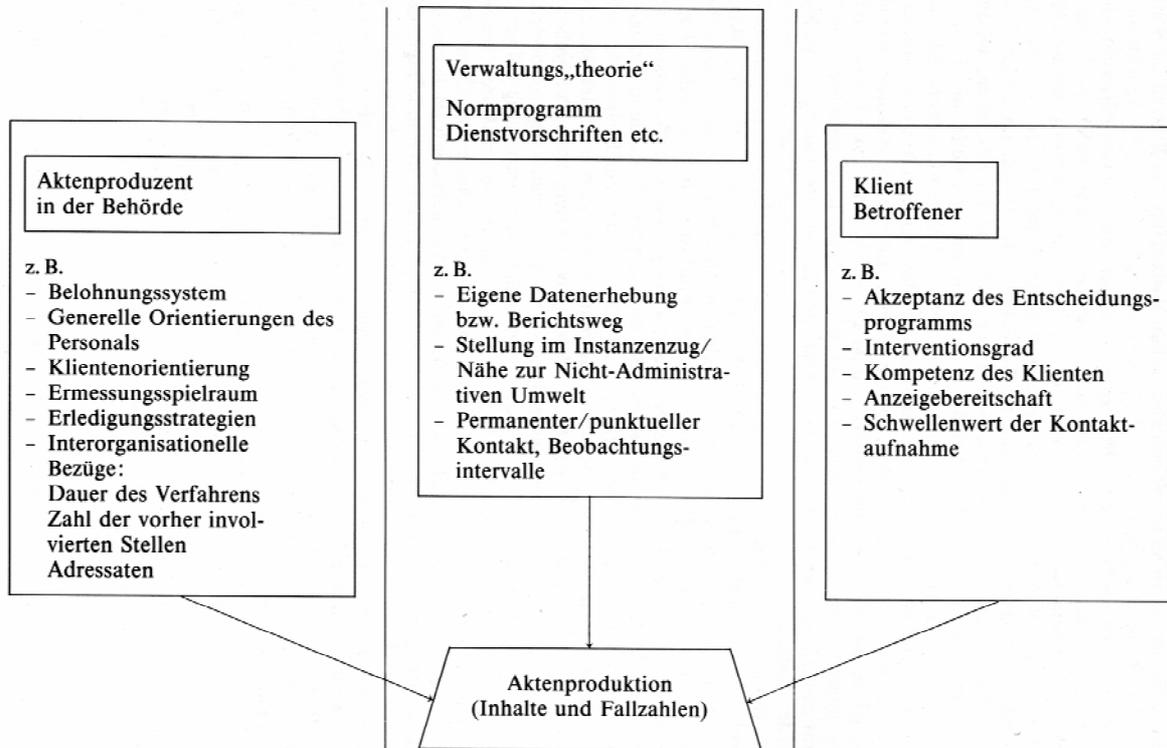
Nicht diskutiert werden sollen im folgenden die Selektionsprozesse, die der Datenproduktion nachgelagert sind. Greift die Sozialforschung auf archivierte Massenakten zurück, so wird sie meist nur einen kleinen Ausschnitt der ursprünglichen Bestände vorfinden, sei es, daß z. B. durch äußere Einflüsse, wie Krieg oder Feuer, oder durch gezielte Eingriffe des Archivars selbst Teile vernichtet bzw. kassiert wurden. Um beispielsweise die Repräsentativität der überlieferten Bestände in Bezug auf die Grundgesamtheit der Personen festlegen zu können, die gemäß eines Verwaltungsprogramms Klienten einer Behörde waren, müssen nicht nur die Selektionsprozesse bei der Datenentstehung – z. B. „Aussonderung“ von Klienten durch Sachbearbeiter oder „freiwillige“ Nicht-Inanspruchnahme – nachvollzogen werden, sondern auch abschätzbar sein, welche Auswahlprozesse nach der Datenentstehung bis zur Datenarchivierung wirksam wurden. Neben eine Datenkunde, die den Datenentstehungsprozeß nachzeichnet, müßte demnach konsequenterweise auch eine Überlieferungskunde treten, die systematisch Informationen zur Vernichtung und Kassation von Massenakten zusammenstellt. Ein wichtiges Ziel unserer Expertenbefragungen und der Literaturanalyse war die Ermittlung von Kenntnissen über die Entstehungsprozesse von Daten in Behörden, und dies sowohl hinsichtlich der Vielfalt von Datenproduzenten als auch einer umfangreichen Berücksichtigung des gesamten Datenerfassungs- und Bearbeitungsprozesses. Auf diese Weise war es möglich, einen Katalog von Einflußgrößen aufzustellen, der die wesentlichen, im Entstehungsprozeß wirksamen Faktoren umfaßt.

Schaubild 3 skizziert die drei Faktorenbündel, die wir zur Klassifizierung der Einflußgrößen gewählt haben. Die Faktoren „Verwaltungstheorie“ und „Aktenproduzent in der Behörde“ umfassen hierbei die Einwirkungen der Verwaltungen selbst, das Faktorenbündel „Klient/Betroffene“ die Gesamtheit möglicher Einflußgrößen, die von den Objekten des Verwaltungshandelns (Klienten, Kunden und Organisationen) her auf die Entstehung von Daten einwirken.

Die Unterscheidung der Einwirkung des organisationellen Kontextes in das Normprogramm und seine weiteren schriftlichen Detaillierungen, bestehend aus den gesetzlichen Grundlagen, Verwaltungsvorschriften und Dienstanweisungen einerseits und in den implizit bleibenden Teil des faktischen Verhaltens von Organisationsmitgliedern andererseits, ermöglicht die Identifikation von administrativ bedingten „Störfaktoren“, die eine Abweichung der ermittelten Sachverhalte von dem ursprünglich intendierten verursachen können.

Mit dieser gleichgewichtigen Darstellung der Einflußfaktoren Normprogramm, faktisches Verwaltungshandeln und Verhaltensweisen des Objektbereichs (Klien-

Schaubild 3:



ten/Betroffene) möchten wir gleichzeitig versuchen, eine vermittelnde Position im Streit um die Aussagekraft von Verwaltungsdaten einzunehmen.

In der Vergangenheit wurde an einzelnen prozeß-produzierten Daten nachgewiesen, dann aber verabsolutiert, wie einer der drei Faktorenbündel sich dominant auf die Aussagefähigkeit von prozeß-produzierten Daten auswirkte. So kam es zu solch einseitigen Festlegungen, prozeß-produzierte Daten eigneten sich „eigentlich“ nur für Instanzenforschungen, nur für die Verwaltungssoziologie. Der „labeling approach“ ist hierfür nur ein Beispiel. Andererseits wurde behauptet, diese Daten seien ideale Beschreibungen von Problemgruppen, Minoritäten und sonstigen Subpopulationen und ihren Motivationen. Die ältere Kriminologie ist hierfür ein Extrembeispiel. Und schließlich wurde auch davon ausgegangen, sie indizierten nur die Umfunktionierung von staatlichen Programmen durch die eigentlich zu beeinflussenden Objekte.

Diese Einseitigkeiten sind heute (1980) schon wieder Vergangenheit. Zunehmend wird die Mehrfachbedingtheit dieser Daten anerkannt und versucht, je nach Fragestellung die jeweils unerwünschten Verzerrungen zu isolieren. Während früher geglaubt wurde, daß einzelne Einflußfaktoren jeweils dominant die Abbildqualität prozeß-produzierter Daten bestimmen, wird heute zunehmend von einem Kontinuum unterschiedlicher Mischungsverhältnisse zwischen den einzelnen Einflußfaktoren ausgegangen. In diesem Kontinuum ist aber dennoch Platz für solche extremen Möglichkeiten, wie: „der Akteninhalt“ ist hauptsächlich Reflex auf Erledigungsstrategien des Behördenpersonals oder „ein getreues Abbild der Intentionen des Normprogramms“.

Diese Offenheit gegenüber der möglichen Koexistenz verschiedenster Einflußfaktoren führte jedoch erst einmal dazu, daß die Unsicherheit bei der Bestimmung einer Vielzahl von Einflußgrößen stieg; die Interpretation von prozeß-produzierten Daten wurde erst einmal schwieriger.

Sind Verwaltungsprogramme und ihre Definitionen selbstverständlicher Teil der Alltagswelt des Klienten geworden und ist ihr Interventionscharakter gering – so beispielsweise beim Standes- oder Einwohnermeldeamt –, so bestimmt das Normprogramm wesentlich den Abbildcharakter der erhobenen prozeß-produzierten Daten. Individuelle Motivationen und das faktische Verhalten der aktenführenden Stellen spielen dann eine vernachlässigbare Rolle.

Bei der Aufnahme von Anzeigen durch die Streifenpolizei hingegen schlagen die Arbeitsbedingungen und die von dieser Gruppe internalisierten Werte und Einstellungen stark durch. Die These der „Barrieren zur Justiz“ indiziert, in welchem starkem Maße die Verfolgung von Rechtsbrüchen abhängig ist von der Anzeigebereitschaft oder etwa von der Verfügbarkeit von Spezialisten in interpersonellen Netzwerken, also von Eigenschaften, die dem Faktorenbündel Klient/Betroffener zuzuordnen sind.

In anderen Verwaltungskontakten steht die Bedeutung des Verwaltungsprogramms, das Verhalten von Sachbearbeitern und Klienten in einem oftmals unbekanntem Mischungsverhältnis zueinander. Änderungen in Klientenzahlen

sind z. B. dann nicht mehr einfach zu interpretieren. Ist die Erhöhung oder Abnahme von polizeilich erfaßten Kriminalitätszahlen die Folge einer veränderten Verfolgungsintensität der Polizei oder auf einen Wandel in der Meldebereitschaft der Bevölkerung zurückzuführen?

Hieraus sollte schon deutlich geworden sein, daß prozeß-produzierte Daten eine Realität eigener Art abbilden. Akten aus dem Justizbereich wurden z. B. als „die zum Zwecke der Rechtfertigung von Entscheidungen und ihrer rechtlichen Absicherung produzierte Version eines Entscheidungsablaufes“ gekennzeichnet. „Diese Charakterisierung bedeutet nicht zu unterstellen, daß in einer Akte falsche Informationen wiedergegeben sind. Sie bedeutet jedoch, daß ... die aktenmäßige Wiedergabe von Vorgängen selektiv und daß die Darstellungsweise vom Interesse an einer bestimmten Wirkung abhängig ist. Das Prinzip, nach dem die Selektion unter möglichen Informationen vorgenommen wird, und das Interesse, das die Darstellungsweise bestimmt, ist die Absicherung (=Legitimierung) von Entscheidungen. Akten müssen als Dokumente interpretiert werden, die diesem Legitimationszweck dienen“ (Erhard Blankenburg).

Prozeß-produzierte Daten sind daher oftmals mehrdeutig als Ergebnis ihrer Beeinflussung durch einen Satz unterschiedlicher Faktoren. Diese nicht eindeutige Indikatorenfunktion soll noch einmal exemplarisch am Beispiel von Schulberichten aufgezeigt werden, in denen Lehrer das Jugendamt über Erziehungsprobleme und Auffälligkeiten von Kindern informieren. Wird mit dem berufsbezogenen Alltagswissen der Bestand an Kenntnissen, Erfahrungen und Einstellungen, auf die Lehrer zurückgreifen, um schulalltägliche Ereignisse sinnhaft darzustellen und zu erklären, „verstanden“, so könnte zunächst vermutet werden, daß die Schulberichte dieses „alltägliche Hintergrundwissen“ der Lehrer, mit dem sie ihre schulbezogenen Handlungssituationen strukturieren, „widerspiegeln“. Dies ist jedoch nicht der Fall. Das berufsbezogene Alltagswissen geht aus den Schulberichten nicht ungebrochen und nicht unmittelbar hervor und ist bereits in vielfacher Weise gefiltert.

Zumindest drei Verzerrungsfaktoren lassen sich nennen:

- Das Verhältnis des Lehrers zur berichteinholenden Behörde: Lehrer berichten vor allem über die Verhaltensmerkmale, von denen sie annehmen, daß sie für die Urteilsfindung und das Kontrollverhalten der Behörde von besonderer Bedeutung sind.
- Das Verhältnis des Lehrers zum Schüler: Der Grad der Vertrautheit mit dem Schüler und seine Sympathie für ihn entscheiden maßgeblich darüber, welches Bild im Bericht gezeichnet wird.
- Das Verhältnis des Lehrers zum Schulbericht: Als weiterer Verzerrungsfaktor bestimmt die persönliche Einschätzung der Notwendigkeit und Wirksamkeit von Schulberichten die Art und das Ausmaß der Brechung des berufsbezogenen Alltagswissens.

Das Arbeiten mit prozeß-produzierten Daten erfordert somit genaue Kenntnis über die Entstehungsbedingungen der benutzten Datenbestände, über die Sachverhalte, die bei der Datenerhebung systematisch ausgeblendet werden, und über die Möglichkeiten, durch Schätzverfahren oder die Kombination von Daten aus anderen Quellen die Begrenztheit der Daten zu überwinden. Häufig muß zunächst einmal ermittelt werden, über welche Ausschnitte der Realität und welchen Objektbereich die Daten überhaupt Auskunft geben können. Hiermit ist gemeint, daß prozeß-produzierte Daten in einem häufig unbekanntem Mischungsverhältnis Verhaltensweisen der Klienten, das Verwaltungshandeln des Datenproduzenten und den organisationellen bzw. interorganisationellen Kontext, in denen beide Akteure agieren, abbilden.

Am Beispiel der Kriminologie ist aufzeigbar, wie lange eine Disziplin Wissen akkumulieren muß, um zu einem angemessenen Verständnis ihrer Datenbasen zu kommen. Für die klassische Kriminologie, vorwiegend von Juristen betrieben, stellten sich Akten als wirklichkeitsgetreues Abbild der Realität dar. In ihnen sollte sich die Kriminalität – so wie sie in der gesellschaftlichen Realität auftrat – unverzerrt widerspiegeln. Akten wurden nicht als Produkt von Interaktionen und Strategien verstanden.

Diese methodische Bewußtlosigkeit der traditionellen Kriminologie wurde erst in den 60er Jahren überwunden, als eine Kriminologie begründet wurde, die mit sozialwissenschaftlichen Fragestellungen erneut an ihr traditionelles Material herantrat. In der Folge wurde das Entscheidungsverhalten der Instanzen sozialer Kontrolle unter den Aspekten der moralischen Legitimation von Entscheidungen und ihrer Überprüfbarkeit bzw. der realen Erledigungsstrategien thematisiert. Diese Faktoren – so wissen wir heute – gehen entscheidend in die Konstruktion von Akteninhalten ein. Einige Jahre wurde in den Labeling-Ansätzen die Bedeutung des Informellen in organisationellen Kontexten einseitig überbetont.

Heute (1980) stellen sich die Kriminologie und Kriminalsoziologie als Varianten einer Disziplin dar, die sich bewußt geworden ist, wie komplex die Prozesse der Datenentstehung in ihrem Untersuchungsbereich verlaufen. Die Daten ermöglichen daher nur jeweils begrenzt den Rückschluß auf das formelle und informelle Entscheidungsverhalten von Instanzen oder auf die Verhaltensweisen ihrer Klienten. Die Begrenztheit ist dabei überwunden worden durch den Einsatz mehrerer Methoden und mehrerer Datentypen. Erst die Kombination – z. B. von Umfrage- und Aktenanalyse – ermöglichte die Untersuchung von Phänomenen, die nur jeweils ausschnitthaft in den Blickwinkel eines jeden Erhebungsinstruments fallen.

Die Expertengespräche und die Literaturanalyse haben deutlich werden lassen, daß inzwischen eine größere Zahl von Studien vorliegt, die z.T. gar nicht beabsichtigt, – d. h. häufig als „Nebenprodukt“ eines an einer anderen inhaltlichen Fragestellung interessierten Projekts – die Datenentstehungsprozesse in Behörden untersucht haben. Sie ermöglichten es erstmals, einen Katalog von

relevanten Faktoren aufzustellen, der auf der Basis eines repräsentativen und umfassenden Querschnitts von Verwaltungsdaten entwickelt wurde. Diese Zusammenstellung stellt jedoch nur einen Maximalkatalog von wünschbaren Zusatzinformationen dar, die den Datenentstehungsprozeß näher kennzeichnen. Abhängig von der Verwaltungsaufgabe werden jeweils Angaben zu einer größeren Zahl von Einflußfaktoren überflüssig sein, weil sie im jeweiligen Falle von vernachlässigbarer Bedeutung sind.

## Katalog von Zusatzinformationen

### *A. Verwaltungsprogramm und -handeln*

1. Gesetzliche Grundlagen, Verwaltungsverordnungen
  - Definition der Verwaltungsbegriffe
  - Funktionen der Datenerhebung
  - Entscheidungsrelevante Faktoren des Verwaltungsprogramms
  - Ermessensspielraum des Verwaltungsprogramms
  - Bescheinigungswesen
  - Standardisierungsgrad
2. Sonstige Unterlagen
  - Dienstanweisungen
  - Schulungsunterlagen (Fallsammlungen)
  - Material von Einspruchsstellen/Revisionen
3. Interaktion zwischen Klienten und informationsammelnden Stellen
  - Instrumente der Datenerhebung: Anträge und Formulare
  - Kontaktformen – Dauer und Ablauf
4. Ausdifferenzierung der Informationsbearbeitung
  - Auslagerung der Datenerhebung auf den Klienten
  - Vorsortierungs-, Erhebungs- und Entscheidungsprozesse
  - Zahl der zuvor involvierten Stellen
5. Systeme der Aktenlagerung
  - Struktur der Aktenhaltung
  - Struktur und Umfang des Einzelmaterials
6. Sachbearbeiterverhalten bzw. -Strategien
  - Nachfragestrategien
  - Hilfestellung für den Klienten
  - Validierung von Angaben
  - Kontaktaufnahme mit anderen Stellen

### *B. Eigenschaften des Klientels*

1. Kompetenz
  - Ratsuche in interpersonalen Netzwerken
  - Bürokratische Kompetenz im Umgang mit Behörden
2. Interessenlage
  - Betroffenheit/Angewiesensein des Klienten
3. Akzeptanz des Entscheidungsprogramms
  - Bereitschaft, fehlerhafte Angaben zu machen
  - Schwellenwerte der Melde- bzw. Anzeigebereitschaft (Dunkelfeld)
  - Ausmaß der Erfassung nicht beabsichtigter Klienten

### *C. Hinweise auf einschlägige Literatur*

Neben wünschbaren, weil häufig relevanten Zusatzinformationen konnten durch die Literaturanalyse und die Expertengespräche auch Hypothesen herausgearbeitet werden, die – zum Teil schon empirisch überprüft – den Einfluß unterschiedlicher Faktoren auf die Aktenproduktion kennzeichnen. In der Folge geben wir beispielhaft eine kurze Zusammenstellung solcher Thesen wieder. Sie sollen demonstrieren, was Teil einer Kunstlehre für prozeß-produzierte Daten werden könnte. Weiter systematisiert und ergänzt, sollten sie Anhaltspunkte und Hinweise geben, die die Arbeit mit prozeß-produzierten Daten erleichtern und einen adäquaten Umgang mit dieser Datenbasis sicherstellen:

- Je wichtiger Informationen für das Entscheidungsprogramm der informationsammelnden Stelle sind, um so größer ist die Genauigkeit der vorliegenden Daten.
- Je stärker Aktenproduzenten bei der Informationssammlung mit schon vollständig vorliegenden Angaben arbeiten können, um so größer ist die Wahrscheinlichkeit, daß fehlerhafte Angaben in den Akten verbleiben.
- Je häufiger Informationen mit anderen Verwaltungsstellen ausgetauscht werden, um so höher ist die Genauigkeit der Informationen.
- Je entscheidender Informationen für die nachfolgenden Instanzen sind, um so sorgfältiger werden sie erhoben.
- Je mehr Stellen im Instanzenzug der Informationssammlung vorgelegt sind, um so mehr setzen sich bei der Datensammlung verwaltungsrechtliche bzw. formalisiertere Sichtweisen durch.
- Je standardisierter das Entscheidungsprogramm festgelegt ist, um so weniger wirkt sich das Informationsverhalten der Aktenproduzenten und des Klienten bzw. Betroffenen auf die Akte aus.
- Je mehr die Informationsaufnahme durch Ausgabe von Formularen und Anträgen außerhalb der Behörde durch den Klienten geleistet wird, um

so stärker werden die in Primärumwelten akkumulierten bürokratischen Kompetenzen und die Verfügbarkeit von spezialisierten Personen die gesammelten Informationen beeinflussen, um so weniger sichtbar werden für Verwaltungen die – auch irrigen – Transformationsprozesse.

- Je höher die bürokratische Kompetenz des Klienten/Betroffenen bei geringem Interventionsgrad der Verwaltungsaufgabe, um so mehr messen Verwaltungen die Sachverhalte, die sie zu messen beabsichtigen.
- Je höher die bürokratische Kompetenz des Klienten bei hohem Interventionsgrad der Verwaltungsaufgabe, um so weniger werden sich ausschließlich organisationelle Faktoren in der Aktenproduktion niederschlagen.

## 6. Konsequenzen für eine angemessene Datenbereitstellung

Prozeß-produzierte Daten mit ihren eigenen Aussagequalitäten, aber auch -limitationen haben in den Bereichen, in denen sie zunehmend eingesetzt werden, ihre Spuren hinterlassen. In der Forschung führte die Mehrdeutigkeit von prozeß-produzierten Daten – in ihnen spiegelt sich oftmals mehr als nur eine Sache wider – zu notwendigen Umorientierungen in methodisch-konzeptioneller und verfahrenstechnischer Hinsicht. Diese Defekte aus der Perspektive der Forschung machen oftmals gerade die Besonderheiten der Verwaltungen aus, die diese Daten produzieren. Prozeß-produzierte Daten werden ja gerade dadurch interessant, daß sie oftmals die Behandlung von Klienten oder Objekten abbilden, Realität immer durch den Filter der Behörden gesehen widerspiegeln, und daß bei prozeß-produzierten Daten oftmals die Aporie gerade in der Untrennbarkeit von Daten und Entstehungsbedingungen liegt.

Sind diese Filterungen dennoch für bestimmte Fragestellungen Störfaktoren und nicht Gegenstand der Forschung, so können diese Daten nicht als solche verändert werden, sondern ihre Behandlung muß eine andere werden. Ihre Mehrdeutigkeit macht es oftmals erforderlich, die eingehenden Faktoren in ihrer Wechselwirkung zu isolieren. Dies kann evtl. durch geeignete Forschungsdesigns geleistet werden. Komparative Ansätze ermöglichen es z. B., Drittfaktoren zu kontrollieren, um so prozeß-produzierte Daten angemessen einsetzen zu können.

Prozeß-produzierte Daten bieten oftmals nur indirekte Messungen der interessierenden Sachverhalte an, die nicht notwendigerweise schlechte „Operationalisierungen“, wohl aber ergänzungsbedürftige Messungen sind. Insofern ergibt sich die Notwendigkeit, prozeß-produzierte Daten mit anderen Daten wie z. B. statistischen Erhebungen zu mischen, um somit annäherungsweise die

interessierenden Sachverhalte zu bestimmen. Hierbei können Angaben in Akten ergänzt oder aber in ihrer Qualität durch Vergleiche abgeschätzt werden.

Für den Zeitraum nach 1945 kann hierzu auf eine große Anzahl verschiedener Datentypen zurückgegriffen werden: Statistische Berichte, Interviews, Daten anderer Verwaltungsstellen. Auch besteht oftmals noch die Möglichkeit, durch das Verfahren der Oral History Daten retrospektiv selbst zu erheben.

Interessanterweise wurde die Mischung von Methoden und Daten selten zur gegenseitigen Korrektur von Defekten in den jeweiligen Datentypen verwandt. Es wurden zumeist die verschiedenen Daten – wie z. B. Akten, Beobachtungsdaten und Befragungsdaten – nicht aufeinander bezogen mit dem Ziel, „einen wahren Wert“ zu ermitteln, sondern die Unterschiedlichkeit der Widerspiegelung von Realität in den verschiedenen Datentypen war selbst Thema. Und tatsächlich stellt sich das Problem, ob eine Wahrheit, eine Zahl mosaikartig zusammengesetzt entstehen kann oder doch nur jeweils Facetten und spezifische Limitationen bestehen bleiben. Dies soll hier nicht weiter verfolgt werden; bemerkenswert ist jedoch, wie es durch Mehrmethoden-Designs möglich wurde, ein weitaus differenzierteres Bild der Wirklichkeit zu zeichnen. Eine solche Vorgehensweise hat dazu den Vorteil, nicht ständig mit der „Aufbesserung“ von Daten beschäftigt zu sein, mit der Ex-post-Konstruktion von Zahlen, die gemäß den Intentionen bei prozeß-produzierten Daten gar nicht zu erstellen beabsichtigt waren.

Solche Nutzungen von prozeß-produzierten Daten führen jedoch nicht zu einer Sozialforschung, die mit prozeß-produzierten Daten eine Datenbasis entdeckt zu haben glaubt, die aussagekräftiger sei als andere. Sie führen vielmehr zu einer Sozialforschung mit einer Vielfalt von Datentypen, die aber jeweils nur begrenzt aussagekräftig sind. Damit vergrößern sich aber insgesamt die Möglichkeiten der empirischen Sozialforschung, Realität differenzierter und gleichzeitig dennoch quantitativ zu analysieren.

Eine komplizierter gewordene sozialwissenschaftliche Forschung bedarf jedoch noch einer verbesserten Versorgung mit Zusatzinformationen, um mit fremdproduzierten Daten angemessen umgehen zu können.

In den Experteninterviews wurde deutlich, daß ein weit verstreutes und jeweils begrenztes Einzelwissen bei den verschiedenen Nutzern von prozeß-produzierten Daten vorliegt. Im Bereich der Verwaltungen und der Statistischen Ämter ist dies zumeist „Wissen in Hinterköpfen“ und somit, wie alles mündlich überliefertes Wissen, ständig der Gefahr ausgesetzt „auszusterben“.

Da es zumeist noch als Kunstfertigkeit und noch nicht als systematisches Wissen ausgebildet ist, sind hier die möglichen Verluste an Kontinuität besonders gravierend. Im Wissenschaftsbereich sind die Bemühungen um mehr Systematik ebenfalls sehr selten über diese Zustandsbeschreibung für den Bereich der Verwaltungen hinausgekommen. Hier existieren nebeneinander Ansätze zur Kodifizierung in Form von Handbüchern als auch eine relativ

große Hilfslosigkeit, die Besonderheiten und Einflußfaktoren in wissenschaftlichen Veröffentlichungen auch nur annähernd vergleichbar und vollständig mitzuteilen. Dem Nutzer bleibt es zur Zeit selbst überlassen, die Bedeutsamkeit der einzelnen Störfaktoren abzuschätzen.

Die zunehmende Nutzung von Verwaltungsdaten durch verschiedenste Benutzergruppen bei gleichzeitig weit verstreutem und jeweils nur begrenztem Wissen um die Abbildqualität prozeß-produzierter Daten macht es erforderlich, das bislang vorliegende Wissen aufzubereiten und zu systematisieren, evtl. Lücken durch Zusatzerhebungen zu schließen und Grundlagenuntersuchungen anzustoßen, in denen methodisch-technische Verfahren des „Rechnens mit Fehlern“ entwickelt werden. Dies nur als ein einmaliges Unternehmen zu begreifen, würde übersehen, in welchem starkem Maße gerade die Kenntnisse über die Abbildqualität von prozeß-produzierten Daten nur Nebenprodukte inhaltlich anders orientierter Forschungsvorhaben sind. Der Schwerpunkt des Interesses der an diesen Untersuchungen beteiligten Wissenschaftler und Planer liegt eben nicht in der Aufbereitung, Systematisierung und Vermittlung von „Abfallprodukten“ ihrer Tätigkeit, sondern in der Analyse und Darstellung ihrer inhaltlich anderen Fragestellungen und Ergebnisse. Die Qualität von prozeß-produzierten Daten selbst ist hier nicht das zentrale Thema.

Nur die kontinuierliche Beobachtung und Auswertung von Forschungsarbeiten, Planungsberichten etc. stellt daher sicher, daß die Kenntnisse über die Entstehung von Daten auch weiterhin für die verbesserte Bereitstellung von Zusatzwissen genutzt werden können.

Die Auswertung vorliegenden Materials, Zusatzerhebungen und das Anstoßen von Grundlagenforschung führen dann zu einer verbesserten Datenversorgung, wenn dieses Wissen über die Aussagequalitäten prozeß-produzierter Daten und ihre Begrenztheit benutzergerecht angeboten wird. Während die Bereitstellung von Literatur in IuD-Systemen heute vornehmlich auf der Basis schriftlich fixierter, teilweise sogar selbst maschinenlesbarer Dokumentationen geschieht, haben unsere Ergebnisse deutlich werden lassen, daß die Vermittlung von Zusatzinformationen über Daten – sollte sie sich auf die Datendokumentation beschränken – den Bedürfnissen der verschiedenen Benutzerkategorien nicht voll Rechnung tragen kann. Die Schriftliche Fixierung müßte dann zum Teil so umfangreich werden, daß sie – einmal abgesehen von dem finanziellen Aufwand – für viele Nutzergruppen unbrauchbar, weil undurchsichtig würde. Neben der Datendokumentation sollten daher auch individuelle Beratung und Schulungsmaßnahmen zur Vermittlung des Zusatzwissens wichtige Bausteine einer verbesserten Infrastruktur für die Datenversorgung sein.

Der Aufbau einer Datendokumentation für prozeß-produzierte Daten setzt auf jeden Fall voraus, daß das Wissen in Hinterköpfen in stärkerem Maße als bisher auch schriftlich fixiert und zur Verfügung gestellt wird. Einige Systeme zur Dokumentation numerischer Daten haben in der Vergangenheit ansatzweise

versucht, wichtige Beschreibungselemente für solche Daten zusammenzufassen.

Neben der Datendokumentation zur Beschreibung einzelner Datensätze können Interpretationshilfen zur Nutzung prozeß-produzierter Daten aber auch in Handbüchern bereitgestellt werden, in denen die Kenntnisse über die Besonderheiten dieses Datentyps schriftlich kodifiziert sind, evtl. sogar in Form von „Datengebrauchsanweisungen“. Die Handbücher enthalten eine Datenkunde, die nicht auf der dokumentarischen Methode basiert, sondern aus einem sozialwissenschaftlichen Begleitwissen besteht. Diese schriftliche Kodifizierung der Besonderheiten von prozeß-produzierten Daten wird durch die Wissenschaft vornehmlich in Form von Handbüchern für den Gebrauch in einzelnen Forschungsbereichen angelegt sein. Die Schätzungen darüber, ob die Forschungslage schon in absehbarer Zeit solche Handbücher ermöglichen würde, schwankten in den untersuchten Disziplinen. Für prozeß-produzierte Daten aus dem Bereich der Justiz und der Polizei werden die Möglichkeiten heute schon als sehr gut eingeschätzt.

Zusammenstellungen und Systematisierungen durch Dokumentationen für einzelne Datensätze oder in Handbüchern, die einzelne Forschungsbereiche abdecken, wären schon ein entscheidender Schritt hin zu einer qualitativ verbesserten Informationsversorgung, die nicht nur Daten, sondern auch Interpretationshilfen bereitstellt.

Beschränken sich die Bemühungen jedoch auf die Datenbeschreibung bzw. auf Handbücher, so kann weder dem noch geringen und wenig weit verbreiteten Wissensstand, dem teilweise geringen Standardisierungsgrad noch den Erfordernissen und den Wünschen der verschiedenen Nutzergruppen Rechnung getragen werden. Hieraus folgt die Notwendigkeit eines benutzernahen Services, der auch individuelle Beratungsleistungen anbietet. Ein Nachweis von Experten, also der Personen, die mit dem Typus prozeß-produzierter Daten längere Zeit gearbeitet haben, sollte das Maß an individueller Beratung ermöglichen, das hier als Folge jeweils spezifischer Interessen und Bedürfnisse der Nutzer angebracht ist.

Die Datendokumentation, die Kodifizierung von Wissen in Handbüchern und die individuelle Beratung für Nutzer werden erleichtert durch Schulungsmaßnahmen, die die Kenntnisse über die Entstehungsbedingungen prozeß-produzierter Daten an die mit diesem Material arbeitenden Wissenschaftler und Planer vermitteln. Eine Datenkunde für prozeß-produzierte Daten sollte beispielsweise zum Bestandteil der Methodenkurse in der empirischen Sozialforschung werden, ebenso wie dies schon seit langem für die Datentypen gilt, die vom Sozialforscher selbst erhoben werden. Die zunehmende Nutzung von fremdproduziertem Material, insbesondere von prozeß-produzierten Daten macht es erforderlich, in den Lehrveranstaltungen auf die Besonderheiten dieses Materials hinzuweisen, um so eine verbesserte Nutzung prozeß-produzierter Daten sicherstellen zu können.

Durch ein solches Nebeneinander von Maßnahmen der Datendokumentation, der Kodifizierung, der Beratung und Schulung sollten sich insgesamt die Arbeitsbedingungen für einen angemessenen und verstärkten Umgang mit prozeß-produzierten Daten verbessern. So fehlerhaft oder unvollkommen Vollzugsdaten auch sein mögen, ihre Nichtbenutzung wäre schädlicher als der vorsichtige Umgang mit unvollkommenen Daten, denn prozeß-produzierte Daten liefern wichtige Informationen über Aspekte gesellschaftlicher Wirklichkeit, die durch andere Datentypen nicht erfaßbar sind.

Ihr besonderer Abbildcharakter hat prozeß-produzierte Daten zu einer bedeutsamen Datenbasis für Sozialforschung und Planung werden lassen. Durch die Verwaltungsautomation und den Aufbau von Datenbanken und Informationssystemen haben sich die Chancen für eine größere Nutzung von prozeß-produzierten Daten erhöht.

Die Qualität der Datenversorgung hängt aber nicht von der Bereitstellung einer großen Menge von Daten ab. Informationssysteme und Datenbanken können sogar zu Systemen der Desinformation werden, wenn nicht gleichzeitig Interpretationshilfen für den angemessenen Umgang mit prozeß-produzierten Daten bereitgestellt werden, sei es in Form von Datendokumentationen, einer Datenkunde, individueller Beratung oder Schulungsmaßnahmen.