

# Archivierung digitaler Datenbestände aus der DDR nach der Wiedervereinigung

Von MICHAEL WETTENGEL

Die Sicherung, Erschließung und Nutzbarmachung von Unterlagen der aufgelösten zentralen Staatsorgane der DDR war eine der wichtigsten Aufgaben des Bundesarchivs nach der deutschen Vereinigung. Dabei mußten nicht nur herkömmliche Überlieferungsformen, sondern auch digitale Datenbestände übernommen und archiviert werden. Dieser Aufgabe widmete sich das im August 1991 neu eingerichtete Referat für *maschinenlesbare Dateien* im Bundesarchiv. Sie bildete über Jahre hinweg den wichtigsten Arbeitsschwerpunkt des neuen Referates, das seit Sommer 1993 auch mit Personal und einer Magnetbandstation ausgestattet wurde. Das Kapitel der Sicherung und Übernahme von DDR-Datenbeständen konnte erst zur Jahreswende 1997/98 weitgehend abgeschlossen werden. Gegenwärtig stehen dabei nur noch wenige Datenbestände aus, vor allem die im Besitz der Bundesanstalt für vereinigungsbedingte Sonderaufgaben (BvS)<sup>1</sup> befindlichen DDR-Daten, deren Übergabe aber bereits geregelt ist, und jene DDR-Statistikdateien, die bislang noch nicht vom Statistischen Bundesamt übergeben wurden.

Auch wenn die Erschließung und Aufarbeitung der Datenbestände von Stellen der DDR die Mitarbeiter des Bundesarchivs noch viele Jahre beschäftigen wird, so bietet es sich dennoch an dieser Stelle an, Bilanz zu ziehen und auf die Erfahrungen aus der archivischen Sicherung von DDR-Datenbeständen zurückzublicken. Dabei sollte im Blick behalten werden, daß aufgrund der historischen Umstände besondere Bedingungen bei der Übernahme bestanden. Bestimmte Problemstellungen, Lösungsansätze und Verfahrensweisen bei der Sicherung und Erschließung der Daten lassen sich aber trotzdem verallgemeinern.

## Mengen und Bestände

Bis zum Jahresende 1997 wurden im Bundesarchiv mehr als 1200 Magnetbänder mit etwa 4400 Dateien von zentralen Stellen der DDR als archivwürdig bewertet und archiviert. Diese Dateien können in drei Gruppen unterteilt werden:

### Erhebungen der amtlichen Statistik

Die erste Gruppe umfaßt die elektronisch gespeicherten *Erhebungen der amtlichen Statistik* der DDR.

---

<sup>1</sup> Die BvS ist eine am 1. Januar 1995 gegründete Nachfolgeeinrichtung der Treuhand.

Dazu zählt einer der größten Datenbestände, die 78 Statistiken umfassenden Dateien der *Staatlichen Plankommission* und der *Staatlichen Zentralverwaltung für Statistik*, die insgesamt 2568 Dateien umfassen. Diese wurden dem Bundesarchiv vom Statistischen Bundesamt übergeben, das die Daten der amtlichen Statistik der DDR nach der Vereinigung gesichert hatte. Dabei handelt es sich beispielsweise um Statistiken zur industriellen Produktion, zur Bauproduktion, zur landwirtschaftlichen Produktion, zur Energiewirtschaft, zum Arbeitskräfteeinsatz, zum ökonomischen Nutzen von Investitionen, zum Technikeinsatz sowie zum Personen- und Güterverkehr.

Häufig werden Einwände hinsichtlich der Glaubwürdigkeit der amtlichen DDR-Statistik erhoben. Diese Vorbehalte mögen gegenüber den recht spärlichen veröffentlichten statistischen Ergebnissen durchaus berechtigt sein, nicht aber bei den hier vorliegenden statistischen Erhebungsunterlagen. In diesem Zusammenhang wurde vom Statistischen Bundesamt, das die Statistikdateien zum Zwecke der Rückrechnung auswertete, festgestellt, *daß die Statistik der DDR von den Berichtspflichtigen weitgehend korrekte Angaben erhielt, daß aber in Veröffentlichungen auf Weisung der vorgesetzten Stellen i[n] d[er] R[egel] nur die politisch gewünschten Ergebnisse nachgewiesen werden durften.*<sup>2</sup> So wurden beispielsweise Angaben über Selbstmorde in der DDR vor der Wende nicht bekanntgegeben, und *statistische Ergebnisse der Öffentlichkeit gegenüber oft verfälscht ... oder falsch dargestellt.*<sup>3</sup> Auf den elektronischen Datenträgern befindet sich dagegen das unverfälschte Erhebungsmaterial der amtlichen Statistik der DDR. Allerdings gibt es eine Reihe von Schwachpunkten hinsichtlich der Qualität oder Vollständigkeit der Daten. So wurde beispielsweise der „x-Bereich“ nicht erfaßt. Dazu gehörte unter anderem die Armee, die Polizei, der Strafvollzug, der Zoll und die Staatssicherheit.<sup>4</sup>

### Erhebungen von Fachbehörden

Bei der zweiten Gruppe an Datenbeständen handelt es sich um elektronisch gespeicherte *Erhebungen von Fachbehörden zum Zwecke statistischer Auswertungen*. Diese enthalten in den hier vorliegenden Fällen keine personenidentifizierenden Angaben. Hierzu zählen beispielsweise die folgenden Datenbestände:

- Das *Eingabenprojekt* des Staatsrates der DDR, bei dem die Eingaben der Bürger an den Staatsrat nach einer einheitlichen Systematik erfaßt wurden (Die Dateien umfassen von 1979 bis 1989 zwischen ca. 60 000 und 130 000 Datensätze jährlich.).

<sup>2</sup> Oswald Angermann: Sammlung, Sicherung und Rückrechnung von statistischen Angaben über die ehemalige DDR. In: Allgemeines Statistisches Archiv 78 (1994) S. 340–349, hier S. 342.

<sup>3</sup> Oswald Angermann, wie Anm. 2, S. 342. – Vgl. auch Wolfgang Fritz: Die amtliche Erwerbstätigenstatistik in der DDR. In: Historical Social Research 22 (1997). Heft 3/4. S. 300–357, hier S. 305 ff.

<sup>4</sup> Vgl. dazu ausführlich Wolfgang Fritz, wie Anm. 3, S. 307 ff.; Heike Wirth: Amtliche Bevölkerungserhebungen der DDR als Quelle für sozialstrukturelle Analysen. In: Potsdamer Bulletin für Zeithistorische Studien 10 (1997) S. 25–39, hier S. 27 f.

- Die *Datei Grenzzwischenfälle* des Kommandos der Grenztruppen in Pätz, in der sämtliche Vorkommnisse an den Grenzen der DDR zwischen 1971 und 1990 registriert wurden, allerdings ohne personenbezogene Angaben. Pro Jahr wurden durchschnittlich 35 000 Meldungen aufgenommen.

Diese Projekte dienten vor allem dem fachlichen Berichtswesen. Wie an der *Datei Grenzzwischenfälle* zu ersehen ist, wurden sie aber auch unmittelbar für die Aufgabenerfüllung der jeweiligen Dienststelle genutzt. So sollte die Auswertung der Grenzzwischenfälle eben auch dazu dienen, insbesondere an der innerdeutschen Grenze Schwachstellen zu ermitteln und Häufungen von Fluchtversuchen zu anderen Faktoren – wie beispielsweise Witterungsverhältnissen und Jahreszeiten – in Beziehung zu setzen. Von einer gewissen Zuverlässigkeit der Daten ist daher auszugehen. Die bereichsbezogene Erhebung der Angaben durch die fachlich zuständige Behörde oder Dienststelle dürfte zu ihrer weitgehenden Korrektheit ebenfalls beigetragen haben.

### Personenbezogene Massendatenspeicher

Zuletzt seien die *personenbezogenen Massendatenspeicher* genannt.<sup>5</sup> Auch diese Projekte dienten dem fachlichen Berichtswesen sowie der Aufgabenerfüllung der jeweiligen Behörde, etwa im Rahmen der Personalverwaltung. Bei ihnen handelte es sich durchweg um sehr umfangreiche Dateien mit oft Hunderttausenden von Datensätzen. Sie machen derzeit insgesamt etwa ein Drittel aller im Bundesarchiv archivierten Datenbestände aus. Hierzu gehören beispielsweise:

- Der Datenspeicher *Gesellschaftliches Arbeitsvermögen der DDR (GAV)*, der personenbezogene Daten zum Ausbildungs- und Berufsweg von etwa 7,25 Millionen Erwerbstätigen zum Stichtagsdatenbestand vom 31. 12. 1989 enthält. Dies entspricht 75 % aller Beschäftigten der DDR. Der Begriff *gesellschaftliches Arbeitsvermögen* stammt aus der sozialistischen Volkswirtschaft und umfaßt *die Gesamtheit der im gesellschaftlichen Arbeitsprozeß anwendbaren Kenntnisse, Fähigkeiten und Fertigkeiten der arbeitsfähigen Bevölkerung*.<sup>6</sup> Der Datenspeicher diente vor allem der Planung des Arbeitskräfteeinsatzes in der DDR. Mit Hilfe des GAV konnten Erwerbstätige ermittelt werden, die über spezielle Kenntnisse oder Befähigungen verfügten.
- Die *Strafgefangenen- und Verhaftetendateien* des Ministeriums des Innern der DDR, in denen umfangreiche Angaben zu den Häftlingen in der DDR von 1980 bis 1990 gespeichert sind. Dabei dürfte das Projekt NRC von besonderer Bedeutung sein. Zu den im Rahmen dieses Projektes erfaßten Angaben zählen beispielsweise der Festnahmegrund, Familien- und Wohnverhältnisse, Vorstrafen, Zugehörigkeit zu Parteien und Massenorganisationen, Verbindungen ins kapitalistische Ausland, Art des Untersuchungsorgans (zum Beispiel Ministerium für Staatssicherheit oder Ministerium des Innern), Art des verurteilenden Gerichts, Straftat, Strafmaß, Disziplinarmaßnahmen, Entlassungsdatum, Ent-

<sup>5</sup> Vgl. Ulf Rathje: Die personenbezogenen Massendatenspeicher der DDR im Bundesarchiv. In: Potsdamer Bulletin für Zeithistorische Studien 10 (1997) S. 40–54.

<sup>6</sup> Kinze, Knop und Seifert: Volkswirtschaftsplanung. Berlin 1977. S. 75.

lassungsgrund, Maßnahmen zur Wiedereingliederung etc. Mit den Dateien wurden die Verwaltung und der Arbeitseinsatz der Strafgefangenen und Verhafteten, ihre medizinische Versorgung und Fortbildung gesteuert. Zugleich ermöglichte der Datenaustausch mit dem Kriminalregister und der Personendatenbank der DDR eine zielgerichtete, rasche Fahndung auch nach entlassenen Häftlingen.<sup>7</sup>

- Der *Zentrale Kaderdatenspeicher (ZKDS)* des Ministerrates der DDR und verschiedene *Arbeitskräftedatenspeicher der Ministerien (AKDS)*. Der Kaderdatenspeicher enthält je nach Jahrgang und Datei Angaben über 90 000 bis 330 000 „Kader“, d. h. Mitarbeiter zentraler und örtlicher Staatsorgane vom Leiter bis zum technischen Mitarbeiter sowie Leitungspersonal der zentral und örtlich geleiteten Wirtschaft und von Außenhandelsbetrieben. Auch Nomenklaturkader des Ministerrates sind hier gespeichert. Im Datenbestand finden sich nicht nur Angaben über die Ausbildung, den beruflichen Werdegang sowie dienstliche und politische Funktionen der jeweiligen Person, sondern auch ausführliche Informationen über deren familiären Hintergrund.<sup>8</sup> Der ZKDS sollte ein Instrument für die Verwaltung und Planung, den Einsatz und die Kontrolle von Führungspersonal in Staat und Wirtschaft der DDR bilden. Dabei ging es vor allem um eine Entscheidungshilfe für so zentrale Fragen wie der Kaderauslese und -qualifizierung. Der Jenaer Soziologe Stefan Hornbostel spricht von einer „geplanten Elite“, um die Kaderpolitik in der DDR zu beschreiben.<sup>9</sup>

Die hier vorgestellten personenbezogenen Dateien stießen schon früh auf das besondere Interesse der zeitgeschichtlichen und sozialwissenschaftlichen Forschung.<sup>10</sup> Allerdings müssen gerade hier in besonderem Maß datenschutzrechtliche Bestimmungen beachtet werden. Personenbezogene Datenbestände dürfen daher nur in anonymisierter Form zu wissenschaftlichen Zwecken ausgewertet werden.<sup>11</sup> Hinsichtlich der Zuverlässigkeit der Daten in den personenbezogenen Dateien gilt grundsätzlich dasselbe wie für die bereits zu Beginn genannten

<sup>7</sup> Vgl. zum Datenabgleich zwischen dem Projekt NRC und der Personendatenbank der DDR (Projekt NLA) Bundesarchiv (Koblenz) DO 1 MD/3. Dokumentation. Bd. 61 und 66. – Vgl. zu Übergaben von Daten des Projektes NRC an den Datenspeicher Kriminalpolizei (Projekt NIH) Bundesarchiv (Koblenz) DO 1 MD/3. Dokumentation. Bd. 61 und 87.

<sup>8</sup> Vgl. Ulf Rathje: Zentraler Kaderdatenspeicher (ZKDS). Bestand DC 20 MD/1. Vorläufiges Bestandsverzeichnis. Bundesarchiv. Koblenz 1996.

<sup>9</sup> Stefan Hornbostel: Die geplante Elite – erste Ergebnisse aus der Analyse des Kaderdatenspeichers des Ministerrates. In: Potsdamer Bulletin für Zeithistorische Studien 10 (1997) S. 55–69. – Vgl. zur Zielsetzung des Kaderdatenspeichers Bundesarchiv (Berlin-Lichterfelde) DC 20/9006. Gruppe Personalangelegenheiten des Ministerrates in Abwicklung. Vorlage über die Sicherstellung von Unterlagen aus dem Bereich der Personalabteilung des ehemaligen Ministerrates der DDR. 6. November 1990.

<sup>10</sup> Vgl. unter anderem Frido Dietz und Helmut Rudolph: Berufstätigenerhebung und der Datenspeicher „Gesellschaftliches Arbeitsvermögen“. Statistische Grundlagen zu wichtigen Strukturen der Erwerbstätigen in der vormaligen DDR. In: Mitteilungen aus der Arbeitsmarkt- und Berufsforschung 23 (1990) S. 511–518; Stefan Hornbostel, wie Anm. 9, S. 55–69.

<sup>11</sup> § 5 Abs. 2 und 5 BArchG. – Klaus Oldenhage: Bemerkungen zum Bundesarchivgesetz. In: Der Archivar 41 (1988) Sp. 478–498.

Datenbestände. Die fachliche Aufgabenerfüllung erforderte auch hier eine möglichst große Korrektheit der Daten. Allerdings gibt es Probleme hinsichtlich der Datenqualität, denn offenbar fiel die laufende zeitnahe Aktualisierung der personenbezogenen Angaben in den zentral geführten Datenspeichern, wie dem GAV und dem ZKDS, besonders schwer. Auch die relative Häufigkeit von Inkonsistenzen und Eingabefehlern mindern oft die Qualität der gespeicherten Daten.<sup>12</sup>

### Weitere Übernahmen

Neben diesen bereits archivierten Datenbeständen hat das Bundesarchiv insgesamt etwa 3600 Magnetbänder mit schätzungsweise 20 800 Dateien übernommen, die noch bewertet werden müssen. Den größten Anteil darunter stellen Daten des Ministeriums des Innern der DDR mit insgesamt fast 3400 Magnetbändern. Dazu zählen vor allem DV-Projekte aus dem Bereich der Deutschen Volkspolizei, wie beispielsweise die Kriminalstatistik, die Fahndungsstatistik sowie Projekte zur Disziplin und zum inneren Zustand der Volkspolizei. Außerdem müssen 1350 Disketten aus DDR-Provenienzen, insbesondere aus dem Bereich Strafvollzug, gesichtet werden. Hinzu kommen noch etwa 100 Magnetbandcassetten aus dem Bereich Außenhandel, die sich derzeit noch bei der BvS befinden und im Laufe des Jahres 1998 bewertet werden sollen. Diese Dateien enthalten überwiegend Angaben über Leistungsrechnungen, Warenlieferungen und Zahlungsverkehr.

### Rechentechnik und Datenträger

Die vom Bundesarchiv übernommenen Datenbestände wurden in der DDR in der Regel in staatlichen Rechenzentren auf Großrechneranlagen verarbeitet. Die dort eingesetzte Rechentechnik unterschied sich nicht grundsätzlich von der im Westen. So waren die in DDR-Rechenzentren üblichen ESER-Rechner (*Einheitliches System Elektronischer Rechentechnik*) Kopien von IBM-Großrechneranlagen.

Als Sicherungsmedien wurden vor allem 9-Spur-Magnetbänder verwendet, deren Speicherdichte bei 800 bis 1600 bpi lag. Diese verursachten bei der Archivierung im Bundesarchiv keine besonderen Schwierigkeiten, da die Magnetbandstation hier auch niedrige Speicherdichten bis zu 800 bpi lesen kann und die Datenträgerkennsätze auf den Magnetbändern (Volume-, Header- und Trailer-Labels), die wichtige Informationen über die gespeicherten Daten enthalten, im wesentlichen dem IBM-Standardschema entsprachen. Im Gegensatz etwa zu Disketten waren die Daten bei Magnetbändern als reine sequentielle Bitstrings abgespeichert. Es sei hier nochmals ausdrücklich betont, daß es keiner besonderen Technik bedurfte, um diese Daten lesen zu können.

Allerdings befanden sich viele dieser Magnetbänder in keinem guten Zustand. Zur physischen Sicherung der Daten mußten die Magnetbänder daher möglichst

<sup>12</sup> Vgl. Stefan Hornbostel, wie Anm. 9, S. 62.

rasch umkopiert werden. Dabei wiesen viele Bänder einen extrem hohen Materialabrieb auf, gelegentlich trat auch klebrige Flüssigkeit zwischen den Bandschichten aus. Datenverluste waren daher bei DDR-Beständen keine Seltenheit, wobei in der Regel nur einzelne Datenblöcke fehlerhaft waren. Es gab aber auch Fälle, bei denen ganze Bänder nicht mehr lesbar waren. In sehr seltenen Fällen löste sich die magnetisierbare Schicht von der Trägerschicht, oder die Konsistenz des Bandes hatte sich so verändert, daß es im Magnetbandlesegerät festklebte und sogar reißen konnte. Der Einsatz von Bandreinigungsgeräten war in solchen Fällen nicht immer erfolgreich.

Die Ursache für diese Schäden konnte oft nicht schlüssig nachgewiesen werden. In der Regel ist hier von einem Ursachenbündel auszugehen. Zunächst muß die schlechte Qualität der verwendeten Datenträger aufgrund von Produktions- und Materialmängeln angeführt werden. Außerdem setzte die Übernahme und Umkopierung der Datenbestände erst mit mehrjähriger Verzögerung ein, so daß die erforderlichen Umkopierperioden überschritten waren. Bis zur Übernahme durch das Bundesarchiv wurden die Datenträger außerdem nicht immer unter günstigen Verhältnissen gelagert. Gerade für die Haltbarkeit von Magnetbändern spielen aber optimale Lagerungsbedingungen mit kontrollierten Temperatur- und Luftfeuchtigkeitswerten eine zentrale Rolle.

Auf die große Bedeutung der ursprünglichen Lagerungsbedingungen und der Umkopierzyklen weist der sehr unterschiedliche Erhaltungszustand der Magnetbänder aus verschiedenen DDR-Beständen hin. So traten bei etwa 10–20 % der Magnetbänder Leseschwierigkeiten auf, die oft aber nur einzelne Blöcke betrafen.<sup>13</sup> In solchen Fällen, in denen Magnetbänder in starkem Maße Schmutz und Witterungseinflüssen ausgesetzt waren, konnte dieser Anteil bei bis zu 40 % liegen. Allerdings hielten sich die tatsächlichen Datenverluste aufgrund von physischen Defekten in Grenzen, da die Daten in der Regel nochmals auf mindestens einer Sicherheitskopie gespeichert waren. Die Dateien konnten daher abgeglichen und fehlende Blöcke oder Bänder zumindest größtenteils kompensiert werden. Trotz widriger Bedingungen haben sich Magnetbänder alles in allem als ein recht haltbarer Datenträger erwiesen, denn selbst bei physischen Beschädigungen der Spule oder des Bandes selbst konnten die Daten in den meisten Fällen größtenteils gerettet werden.

Bei der Übernahme digitaler Datenbestände spielten Magnetplatten und Disketten eine geringere Rolle. Auf Disketten, die im 8-Zoll-Format vorlagen, wurden insbesondere bei Großrechneranlagen meist nur Programme gespeichert oder sie dienten neben den kleinen Magnetbändern dem Datentransfer. Es wurden jedoch auch Disketten im 5¼- oder 8-Zoll-Format übernommen, die zur Sicherung von Daten aus Bürocomputer-Anwendungen der DDR dienten. Dabei konnte es sich beispielsweise um Textdokumente oder um Daten aus REDABAS-Datenbanken<sup>14</sup> handeln, die auf Robotron-Bürocomputern mit den Betriebssystemen SCP und DCP erstellt wurden. Hier war die Ausgangssituation daher grundsätzlich anders als bei Magnetbändern aus Großrechnerumgebungen. Die Daten auf den

<sup>13</sup> Nicht eingerechnet wurden dabei Leerbänder (Ini-tapes), bei denen nicht immer sicher war, ob diese nicht absichtlich oder versehentlich gelöscht wurden.

<sup>14</sup> REDABAS war eine DDR-Kopie des Datenbankprogramms DBASE. REDABAS 5 entsprach dabei der Version DBASE 4.

Disketten lagen nicht in einem softwareunabhängigen Format vor, und sie waren von einer spezifischen Hardware abhängig. Um diese Disketten mit ihren veralteten physischen und logischen Formaten daher noch lesen zu können, waren alte Robotron-Bürocomputer mit entsprechenden Laufwerken erforderlich. Auf diesen Geräten konnten die Disketten gesichtet und die Dateien gegebenenfalls in ein softwareunabhängiges Format konvertiert werden.

Magnetplatten lagen in bulgarischen Fabrikaten der Typen EC 5261 (29 MB Kapazität) und EC 5266.01 (100 MB Kapazität) vor. Es gab auch einen neueren Typ, der 200 MB Kapazität besaß. Diese Magnetplatten wurden im Großrechnerbetrieb eingesetzt. Abgespeichert wurden auf ihnen zum einen Systemdaten, also die verwendeten Betriebssysteme, wie beispielsweise SVS und OS/ES, ferner Programmdateien, wobei es sich hier um ganze Programmbibliotheken handeln konnte, und schließlich Projektdaten, also zum Beispiel Auswertungsaufträge oder Programmierprojekte. Magnetplatten wurden in den meisten Fällen nicht als Sicherungsmedien für archivrelevante Datenbestände genutzt. Eine Ausnahme bilden dabei Zwischenstände, die auf Magnetplatten gespeichert wurden, sowie vor allem gegen Ende der 1980er Jahre direktadressierte Datenbankanwendungen. Diese Datenbanken waren in hohem Maße hardwareabhängig. Da die bulgarischen Magnetplatten aufgrund von Produktionsmängeln Deformationen aufweisen konnten, brachte ein Leseversuch das Risiko eines „Headcrashes“, einer schweren Beschädigung des Lesekopfes, mit sich. Deshalb war es früher üblich, Magnetplatten vor einem Leseversuch durch entsprechende Messungen zu überprüfen, um das Schadensrisiko auszuschließen.

Ganz generell sollte aber an dieser Stelle betont werden, daß die technische Lesbarkeit von Daten und Schäden an den Datenträgern nicht die Hauptprobleme bei der Übernahme und Archivierung von DDR-Datenbeständen bildeten. Viel schwerwiegender war dagegen die meist lückenhafte oder fehlende Dokumentation der Datenbestände.

## Zeichendarstellungen und Datenstrukturen

Die Datenbestände aus DDR-Provenienzen bestehen vor allem aus numerischen, hierarchisch strukturierten sequentiellen Dateien, wie sie früher in Großrechner-Umgebungen verarbeitet wurden. Relationale Datenbanken und Textdokumente gab es dagegen nur selten. In den Großrechenanlagen wurden die Daten in der Regel mit Hilfe jeweils speziell programmierter, in hohem Maße hardwareabhängiger Assemblerprogramme verarbeitet. Die Dateien wurden vom Bundesarchiv als *flat files* übernommen, also als Datenstrings ohne Programme oder Softwareelemente.

Da in der alten Computertechnik Arbeitsspeicher und Rechenkapazitäten begrenzt waren, wurden die Informationen vor der Datenerfassung am Rechner zuerst inhaltlich codiert. Dies bedeutet, daß eine Angabe durch eine bestimmte Zahl bzw. eine Zahlenkombination repräsentiert wurde. So wurde beispielsweise ein bestimmter Ort durch eine sechsstellige Zahl ausgedrückt, die im Gemeindeverzeichnis der DDR zu finden ist. Ohne „Codeliste“ – in diesem Fall das Gemeindeverzeichnis – ist es nicht möglich herauszufinden, welcher Ort sich hinter einer



bestimmten Codeziffer verbirgt. In der Regel waren Klartextdarstellungen von Angaben nur bei Personennamen in personenbezogenen Massendatenspeichern zu finden, die meisten übrigen Angaben waren numerisch codiert.

Die Zeichen wurden, wie bei Großrechneranwendungen üblich, bei Klartextdarstellungen in dem Normcode EBCDIC<sup>15</sup> dargestellt. Häufig wurde auch der mit EBCDIC weitgehend identische russische Code DKOI (Erweiterter binärer Code zur Informationsumwandlung) verwendet, der in der DDR auch „ESER-Kode“ genannt wurde. Neben diesen Normcodes wurde jedoch zur internen Darstellung von Zahlenwerten Verfahren der binären „Packung“ oder Komprimierung angewandt. Eine solche „gepackte“ Darstellung der Ziffern verringert den erforderlichen Speicherplatz erheblich. Sie ist allerdings nicht standardmäßig in Klartext umsetzbar und erfordert eine besondere Programmierung. Typischerweise waren zum Beispiel bei personenbezogenen Massendatenspeichern die Grundpersonalien (vor allem Personennamen und Personenkennziffer) in EBCDIC dargestellt und somit im Klartext lesbar, die übrigen Angaben wurden dagegen binär dargestellt. Auch andere, spezifische Kompressions- und Verdichtungsverfahren wurden in DDR-Dateien angewandt.

Charakteristisch für den Großrechnerbetrieb war die Verwendung variabler Datensatzlängen. Wenn daher in einem Datensatz einzelne Felder nicht belegt waren, so wurden diese Felder nicht durch *Blanks* ausgefüllt, sondern ausgelassen. Feldtrennungen (*Delimiters*) waren bei DDR-Dateien aus Großrechnerumgebungen dagegen unüblich, so daß Feld auf Feld ohne Unterbrechung folgte. Meist bestanden die Datensätze aus einem ersten, festen Teil und einem zweiten, variablen Teil. Blocklängen und Satzlängen waren in der Regel intern abgespeichert, vor allem in den festen Teilen der Datensätze.

Diese hier beschriebenen Datenstrukturen und Codierungsmethoden waren keineswegs nur in der DDR, sondern auch bei westlichen Großrechneranwendungen üblich. Sie warfen vor allem deshalb besondere Schwierigkeiten auf, weil die Datendokumentation in der Regel sehr lückenhaft war. Ohne eine umfassende Dokumentation ist aber ein Verständnis und eine Erschließung solcher Datenbestände nur unter erheblichem Aufwand oder gar nicht möglich. Diese Datendokumentation lag nicht selten in digitaler Form auf Datenträgern, meist aber in Papierform vor und konnte einen erheblichen Umfang besitzen.

Eine vollständige Datendokumentation sollte bei sequentiellen numerischen Dateien und Datenbanken folgende Bestandteile umfassen:<sup>16</sup>

- Komplette listenmäßige Aufstellung über alle Datenträger und die jeweils darauf befindlichen Dateien mit Dateinamen und näherer Erläuterung (zum Beispiel gegebenenfalls Jahrgang, Generation der Datei, Version, bestimmter Auszug etc.).

<sup>15</sup> Extended Binary Coded Decimal Interchange Code, 8-Bit-Code für die Zuordnung von Zeichen zu Bitkombinationen. Das druckbare Zeichen „A“ entspricht in EBCDIC beispielsweise der Bitkombination „11000001“.

<sup>16</sup> Vgl. dazu Grundsätze für Datenübermittlung und Datenträgeraustausch (Datenübermittlungs-Grundsätze) der Koordinierungs- und Beratungsstelle der Bundesregierung für Informationstechnik in der Bundesverwaltung (Schriftenreihe der KBS 36). Bonn 1997.



- Technische Angaben zu den Datenträgern hinsichtlich des Typs (zum Beispiel 9-Spur-Magnetband), gegebenenfalls Speicherdichte (bei Magnetbändern in bpi), gegebenenfalls Formate.
- Länge der Datensätze und Datensatzformat (fest, variabel).
- Dateiumfang (Anzahl der Datensätze, Umfang in Megabyte) je Datei.
- Speicherungsform (zum Beispiel sequentiell, indexsequentiell).
- Detaillierter Datensatzaufbau (Datensatzstruktur) und Sortierung.
- Feldbezeichnungen mit relativer Position im Datensatz, Feldlänge, Feldformat, Muß-/Kannfelder, Häufigkeiten von Feldern im Satz.
- Interne Zeichendarstellung (zum Beispiel EBCDIC, binär, ASCII), gegebenenfalls eingesetzte Komprimierungsverfahren.
- Bei Verwendung von Kennsätzen: Kennsatzformate (zum Beispiel IBM-Standardlabel) und Abfolge der Kennsätze.
- Bei geblockten Daten: Angaben zur Blockung.
- Bei Verwendung von Codierungen: Vollständige Codelisten.
- Bei indexsequentieller Speicherungsform: Dateischlüssel, Position, Länge und Format.

Vor allem vollständige Codelisten, der Datensatzaufbau, die Angaben über die Datenfelder und den Umfang der Datei sowie eine ausreichende Beschreibung der Zeichendarstellung und eingesetzter Komprimierungsverfahren müssen als Minimaldokumentation vorhanden sein.

Hinzu sollten noch inhaltliche Projektbeschreibungen kommen, wie zum Beispiel:

- Angaben zu Herkunft, „Historie“ und Entstehungszusammenhang der Dateien,
- die abgebende Stelle und Organisationseinheit,
- die datenführende Stelle (auch beteiligte Stellen),
- der Name des Auftraggebers für die jeweilige DV-Anwendung (gegebenenfalls Organisationseinheit),
- die Rechtsgrundlage und Zielsetzung der DV-Anwendung,
- und ferner Angaben über den Ablauf der Primärerhebung und -erfassung der Daten sowie über die Nutzung und ursprüngliche Verarbeitung der Dateien in der datenführenden Stelle.

Im allgemeinen waren die Rechenzentren in der DDR gehalten, eine vollständige Dokumentation zu ihren DV-Anwendungen zu führen. Oft wurde sogar eine dreifache Fassung der Datendokumentation verlangt. Daß die Dokumentation dennoch lückenhaft war, lag an den spezifischen Bedingungen der Übernahme nach der Vereinigung. Diese hingen ganz wesentlich davon ab, ob und in welchem Maße sich Bundes- und Landesbehörden nach der Vereinigung um Datenbestände aus der DDR kümmerten. Insbesondere in jenen Fällen, in denen das jeweilige Rechenzentrum nun im Auftrag einer Einrichtung des Bundes oder der neuen Bundesländer weiterarbeitete, konnten gut dokumentierte, vollständige und lesbare Datenbestände übernommen werden. Die Programmierer, die zu Zeiten der DDR die Dateien verarbeitet hatten, standen dabei gelegentlich noch für Fragen zur Verfügung. Eine solche geordnete Übernahme war jedoch nur bei etwa 20 % der bislang im Bundesarchiv archivierten Datenbestände von Stellen der DDR möglich. Etwa genauso hoch war jedoch der Anteil jener Datenbestände, zu

denen keine oder nur eine äußerst lückenhafte Datendokumentation übernommen werden konnte. In vielen Fällen wurden Dokumentationsunterlagen und Projektbeschreibungen nach der „Wende“ aus Unkenntnis ihrer Bedeutung weggeworfen oder ganz bewußt vernichtet. Offenbar wurden Unterlagen aus aufgelösten Rechenzentren auch von dem ehemaligen Personal mitgenommen.

Gut dokumentierte *flat files* sind in der Regel problemlos interpretierbar, wenn nicht besondere Komprimierungsverfahren die Lesbarkeit erschweren bzw. weitere Arbeitsschritte erforderlich machen. Solche Datenbestände hingegen, die nicht oder nicht ausreichend dokumentiert sind, können nur mit großem Aufwand oder überhaupt nicht gelesen werden. Zwar gilt im Bundesarchiv grundsätzlich die Vorschrift, daß Datenbestände ohne ausreichende Datendokumentation kassabel sind.<sup>17</sup> In Ausnahmefällen wurden solche DDR-Datenbestände dennoch archiviert, wenn ihr historischer Wert als besonders bedeutend eingestuft wurde.

Um unzureichend dokumentierte Datenbestände lesen zu können, ist eine Rekonstruktion der Dokumentation oder eine Nachdokumentierung der Daten unerlässlich.<sup>18</sup> Zu diesem Zweck sollten alle verfügbaren Unterlagen zu dem Projekt gesichert werden. Hierzu gehören nicht nur Aufzeichnungen über die Programm- und Datendokumentation im engeren Sinne, sondern auch alle verfügbaren Kontextinformationen zur Konzeption, Erhebung, Verarbeitung und Nutzung der Daten. Zunächst ist zu prüfen, ob sich Bestandteile der Datendokumentation nicht auf den Datenträgern selbst befinden. Daher sollten bei Bandmedien Ausdrücke der Datenträgerkennsätze und der ersten Datenblöcke angefertigt werden. Aus den Datenträgerkennsätzen, die in der Regel dem fest vorgeschriebenen IBM-Standardschema folgen, sind bereits verschiedene Grundangaben über die abgespeicherten Dateien zu entnehmen, die eine zumindest vorläufige Ordnung des Datenbestandes ermöglichen.

Wenn keine oder nur lückenhafte Unterlagen aus dem Rechenzentrum vorhanden sind, können Informationen insbesondere in den Berichten des Rechenzentrums an vorgesetzte Behörden und in deren Konzeptionen und Anweisungen zu finden sein. Eine eingehende Sichtung von Schriftgut ist daher erforderlich. Besonders wichtig sind dabei Angaben zum Datensatzaufbau, zu Feldinhalten und zu den Codelisten, aber auch Projektbeschreibungen. Beziehungen zu anderen Datenbeständen ermöglichen Rückschlüsse auf gemeinsame Datenstrukturen und Codierungen. Der in der DDR praktizierte Datenaustausch zwischen den großen Datenspeichern setzte die Benutzung standardisierter gemeinsamer Codelisten geradezu voraus. Eine zeitaufwendige feldweise Identifizierung des Datensatzaufbaus anhand der vorliegenden Unterlagen schließt sich an. Darüber hinaus ist es vor allem zur Ergänzung noch fehlender Angaben und zur Beschreibung von Komprimierungsverfahren erforderlich, Kontakte zu früheren Programmierern der jeweiligen Rechenzentren zu knüpfen.

---

<sup>17</sup> Anweisung für die archivarische Tätigkeit Nr. 52. Übernahme, Bewertung, Erschließung, Konservierung und Benutzung maschinenlesbarer Dateien. Bundesarchiv. Koblenz 1995. Kap. 3. S. 7.

<sup>18</sup> Vgl. ausführlich an einem konkreten Beispiel Michael Wettengel: Zur Rekonstruktion digitaler Datenbestände aus der DDR nach der Wiedervereinigung. Die Erfahrungen im Bundesarchiv. In: Der Archivar 50 (1997) Sp. 735–748.

In jedem Fall ist eine nachträgliche Rekonstruktion der Dokumentation der Dateien, wenn sie überhaupt gelingt, sehr aufwendig. Sie sollte daher unter normalen Umständen bei der Übernahme elektronischer Datenbestände nicht erforderlich sein. Für eine ausreichende Dokumentation von Datenbeständen müssen vielmehr die abgebenden Behörden sorgen.

## Erschließung und Zugänglichkeit

Bereits frühzeitig setzten die Bemühungen um eine Erschließung der übernommenen Datenbestände ein. Inzwischen sind mindestens 60 % der im Bundesarchiv archivierten Datenbestände von Stellen der DDR erschlossen oder liegen zumindest in ausreichend dokumentierter Form vor.<sup>19</sup> Nicht eingerechnet sind allerdings die noch zu bewertenden DDR-Dateien. Die Erschließung von DDR-Beständen wird daher auch in Zukunft noch eine wichtige Aufgabe bleiben. Zugleich wurde insbesondere die sozialwissenschaftliche Forschung und quantifizierende Geschichtswissenschaft in mehreren Artikeln und Vorträgen auf die im Bundesarchiv befindlichen Datenbestände aufmerksam gemacht.<sup>20</sup> Umgekehrt haben auch Historiker und Sozialwissenschaftler öffentlich auf die Notwendigkeit der Sicherung und Bereitstellung digitaler Datenbestände der DDR hingewiesen, so beispielsweise in der „Kölner Erklärung“ der Deutschen Statistischen Gesellschaft von 1993 und in den Beschlüssen des Deutschen Historikertages in Leipzig 1994.<sup>21</sup>

Damit stellt sich jedoch die Frage der Zugänglichkeit der Daten, die zu den schwierigsten bei der Archivierung digitaler Datenbestände zählt. Eine Online-Auskunftstätigkeit aus den hier vorgestellten umfangreichen Datenbeständen würde einen Datenexport in ein leistungsfähiges Datenbanksystem erfordern, das

<sup>19</sup> Vgl. Ulf Rathje, wie Anm. 8; ders.: Datenspeicher Gesellschaftliches Arbeitsvermögen (GAV). Bestand DQ 3 MD/2. Bestandsverzeichnis (überarbeitete Fassung). Bundesarchiv. Koblenz 1995; ders.: Strafgefangenen- und Verhaftetendateien. Projekt NRB (Statistische Auswertungen über den Arbeitseinsatz Strafgefangener). Bestand DO 1 MD/2. Bestandsverzeichnis. Bundesarchiv. Koblenz 1995; ders.: Strafgefangenen- und Verhaftetendateien. Projekt NRC (Personenerfassung). Bestand DO 1 MD/1. Bestandsverzeichnis. Bundesarchiv. Koblenz 1998.

<sup>20</sup> Vgl. Ulf Rathje: Der „Zentrale Kaderdatenspeicher“ des Ministerrates der DDR. In: Historical Social Research 21 (1996). Heft 3. S. 137–141; ders.: Der „Datenspeicher Gesellschaftliches Arbeitsvermögen“ der DDR. In: Historical Social Research 21 (1996). Heft 2. S. 113–118; ders.: Die personenbezogenen Massendatenspeicher der DDR im Bundesarchiv. In: Potsdamer Bulletin für Zeithistorische Studien 10 (1997) S. 40–54; ders. und Michael Wettengel: Datenspeicher Gesellschaftliches Arbeitsvermögen der DDR – ein Werkstattbericht. In: Mitteilungen aus dem Bundesarchiv 2 (1994). Heft 3. S. 157–159; Michael Wettengel: Archivierung maschinenlesbarer Datenbestände im Bundesarchiv. In: Historical Social Research 20 (1995). Heft 4. S. 123–127.

<sup>21</sup> Kölner Erklärung der Deutschen Statistischen Gesellschaft zur Sicherung und Dokumentation der statistischen Datenbestände der ehemaligen DDR. In: Historical Social Research 18 (1993). Heft 4. S. 113. – Resolution zur Archivierung maschinenlesbarer Quellen. Verabschiedet von der Mitgliederversammlung während des 40. Historikertages in Leipzig am 30. 9. 1994. In: Zeitschrift für Geschichtswissenschaft 42 (1994) S. 999.

gepflegt und kurrent gehalten werden müßte. Dies schließt die permanente Migration der Daten und des Datenbanksystems mit ein. Außerdem bedeuten Online-Auskünfte eine ständige Belegung von Speicherplatz auf einem Rechner. Dies alles würde beträchtliche Kosten verursachen, die in keiner Relation zu den Anfragen stünden. Nur ein einziger Datenbestand liegt daher zusätzlich zu dem archivierten *flat file*-Format in einem Datenbankformat vor. Prinzipiell werden daher Auskünfte über Dateien erteilt, aber nicht aus Dateien.

Im Bundesarchiv wurde als Verfahren festgelegt, daß Benutzungen von Datenbeständen im Regelfall durch die Erstellung von Benutzerkopien der gewünschten Dateien stattfinden.<sup>22</sup> Die Daten werden in der Regel in einem *flat file*-Format übergeben. Auch die erforderlichen Dokumentationsunterlagen werden für die Benutzer kopiert. Der Benutzer verpflichtet sich dabei, die kopierten Daten nur für das genannte Forschungsvorhaben zu benutzen und diese nicht mißbräuchlich zu verwenden. Voraussetzung für die Benutzung von digitalen Datenbeständen ist jedoch, daß diese ausreichend erschlossen sind.

Die sonst übliche Schutzfrist von 30 Jahren wurde bei Beständen von Stellen der DDR generell verkürzt, so daß diese in der Regel frei zugänglich sind. Es gibt aber rechtliche Bestimmungen, die die Benutzbarkeit von Datenbeständen auch von Stellen der DDR einschränken. Bei Daten der amtlichen Statistik zählt hierzu insbesondere das Statistikgeheimnis.<sup>23</sup> Außerdem ist bei personenbezogenen Dateien eine Benutzung nach § 5 Abs. 5 BArchG nur dann zulässig, wenn *eine Beeinträchtigung schutzwürdiger Belange durch angemessene Maßnahmen, insbesondere durch Vorlage anonymisierter Reproduktionen, ausgeschlossen werden kann*. Ferner ist gemäß § 5 Abs. 9 BArchG auszuschließen, daß durch die *Verknüpfung personenbezogener Daten ... schutzwürdige Belange Betroffener* beeinträchtigt werden. Findet eine Benutzung personenbezogener Dateien daher in der Form statt, daß Kopien dieser Dateien übergeben werden, so werden die Daten zuvor anonymisiert. Dabei bedeutet eine faktische Anonymisierung, daß eine Einzelperson nur mit einem unverhältnismaßigem Aufwand an Zeit, Kosten und Arbeitskraft identifizierbar ist.<sup>24</sup> Um solche Anonymisierungen durchführen zu können, muß zuvor der Datensatzaufbau genau bekannt sein. Danach sind die Felder zu identifizieren, die ganz oder teilweise gelöscht werden sollen. Für die Durchführung dieser feldbezogenen Löschungen muß ein spezielles Programm, das auf den Datensatz zugeschnitten ist, erstellt werden. Nach einem „Anonymisierungs-*lauf*“ sind stichprobenweise Kontrollen erforderlich, um den Erfolg zu überprü-

<sup>22</sup> Vgl. Anweisung für die archivarische Tätigkeit Nr. 52, wie Anm. 17. Kap. 7. S. 11 f.; Verordnung über Kosten beim Bundesarchiv (Bundesarchiv-Kostenverordnung – BArch-KostV) vom 29. September 1997. In: Bundesgesetzblatt I 1997. S. 2380. Anlage zu § 2 (Kostenverzeichnis).

<sup>23</sup> Vgl. Wolf Buchmann und Michael Wettengel: Auslegung des Bundesstatistikgesetzes bei der Archivierung von Statistikunterlagen. In: Der Archivar 49 (1996) Sp. 67–74; Udo Schäfer: Die Pflicht zur Anbietetung und Übergabe von Unterlagen in der archivarischen Praxis. In: Robert Kretzschmar (Hg.): Historische Überlieferung aus Verwaltungsunterlagen. Zur Praxis der archivarischen Bewertung in Baden-Württemberg (Werkhefte der Staatlichen Archivverwaltung Baden-Württemberg A 7). Stuttgart 1997. S. 40–46.

<sup>24</sup> Vgl. Heike Wirth: Die faktische Anonymität von Mikrodaten. Ergebnisse und Konsequenzen eines Forschungsprojektes. In: ZUMA Nachrichten 30. Mai 1992. S. 7–42; Ulf Rathje, wie Anm. 5, S. 44.

fen. Selbstverständlich sollte nur eine Benutzerkopie der jeweiligen Datei anonymisiert werden. Etwa 20% der Dateien der personenbezogenen Massendatenspeicher liegen derzeit auch in anonymisierter Form vor. Weitere Anonymisierungen werden in der Regel nur auf Anfrage erstellt. Vor der Übergabe anonymisierter Dateien wird von den Benutzern zusätzlich eine Verpflichtung gefordert, daß Persönlichkeitsrechte und schutzwürdige Belange Dritter beachtet und unrechtmäßige Verknüpfungen unterlassen werden.

Bislang ist die Zahl an Benutzungen digitaler Datenbestände von Stellen der DDR noch niedrig. Dies liegt einerseits daran, daß sich Benutzergruppen für diese Überlieferung interessieren, die herkömmlicherweise keine Archive aufsuchen. Die traditionellen Archivbenutzer sind an digitaler Überlieferung in der Regel nicht interessiert oder können damit nicht umgehen. Digitale Datenbestände werden von großen Teilen der historischen Forschung noch nicht als potentielle Quellenbasis berücksichtigt.

Andererseits stellen die meisten Datenbestände aus der DDR auch dann für Benutzer ein Problem dar, wenn sie eine gute Dokumentation besitzen, denn die Datenstruktur und Zeichendarstellung dieser aus Großrechneranlagen stammenden Dateien sowie ihr Umfang erschweren ihre Verarbeitung auf PCs erheblich. Wissenschaftliche Nutzungen der großen Datenspeicher der DDR erfordern daher den Zugang zu geeigneter Rechentechnik sowie vielfach auch Programmierkenntnisse. Für Sozial- und Geisteswissenschaftler ohne solche Kenntnisse läßt sich eine Benutzung daher in der Regel nur durch Hilfe von Fachleuten realisieren, zum Beispiel im Rahmen eines Forschungsprojektes. Bei den Benutzungen von DDR-Datenbeständen handelt es sich daher häufig um Forschungsprojekte, die über Forschungsmittel und Rechentechnik verfügen.

Beispielhaft dafür ist das vom Land Thüringen geförderte Projekt am Institut für Soziologie der Universität Jena über die Eliten in der DDR, das sich ganz besonders auf die Analyse des Zentralen Kaderdatenspeichers stützt. Zu den Ergebnissen des Projektes, die sich bereits jetzt abzeichnen, zählt das Phänomen der „sozialen Schließung“: Die DDR-Elite wies einen wachsenden Grad an Selbstreproduktion auf. Die Chancen von Arbeiterkindern für einen Aufstieg in die Elite war umgekehrt in der Frühzeit der DDR wesentlich höher als in der Endphase.<sup>25</sup>

Das Bundesarchiv verfügt bislang nicht über die Mittel, um Benutzer bei der Auswertung von Datenbeständen in der Weise zu unterstützen, daß es geeignete Rechentechnik zur Verfügung stellen oder analysegerechte Datensätze generieren könnte. Um dennoch den Zugang zu den Datensätzen zu erleichtern, wird die Zusammenarbeit mit der Forschung oder Einrichtungen des Bundes zum Zwecke der Aufbereitung ausgewählter Dateien für Forschungszwecke und der Erstellung von analysegerechten „Forschungsfiles“ gesucht. Außerdem schließt das Bundesarchiv mit Benutzern Vereinbarungen ab, daß nach Abschluß eines Forschungsprojektes eine Kopie der aufbereiteten Dateien an das Bundesarchiv für künftige Benutzer abgegeben wird.

Aus diesen Ausführungen geht hervor, daß die archivische Sicherung einerseits und die Benutzung digitaler Archivbestände andererseits unterschiedliche Konzepte erfordern: Während bei der Sicherung die dauerhafte Aufbewahrung, niedrige Erhaltungskosten, standardisierte Datenträger und Datenformate sowie

<sup>25</sup> Vgl. Stefan Hornbostel, wie Anm. 9, S. 65 f.

die Vermeidung einer Abhängigkeit von bestimmter Hard- und Software im Vordergrund stehen, erfordern Benutzerbedürfnisse eine leichte Zugänglichkeit und Auswertbarkeit der Daten mit Hilfe von Standardsoftware sowie aktuell verbreitete Formate und Datenträger. Die Erfahrung zeigt, daß die Erfüllung beider Aufgaben die eigentliche Herausforderung bei der Archivierung digitaler Datenbestände darstellt. Nach wie vor werden daher Datenbestände zwar als Datenstrings ohne Softwareelemente archiviert, daneben sollen aber auch Benutzerdateien in einer leicht zugänglichen Datenstruktur für Benutzungen bereitgestellt werden. Das Bundesarchiv steht hierbei noch am Anfang.

## Schlußbetrachtung

Trotz aller Besonderheiten lassen sich manche Erfahrungen bei der Übernahme von DDR-Datenbeständen verallgemeinern. Die übernommenen Datenbestände wurden in der Regel in staatlichen Rechenzentren auf Großrechneranlagen verarbeitet, die sich kaum von denen im Westen unterschieden. Dies betraf auch die verwendeten Datenträger und die Verfahren bei der Datenverarbeitung. Daher genügte in der Regel westliche Technik, um die Daten lesen zu können. Die DDR-Datenbestände können somit als typische Daten aus Großrechnerumgebungen angesehen werden, und die Schwierigkeiten, die bei ihrer Archivierung auftraten, dürften auch bei westlichen Großrechner-Daten zu erwarten sein.

Das größte Problem bei der Sicherung der DDR-Datenbestände bildeten nicht, wie ursprünglich angenommen, schadhafte Datenträger oder physisch nicht lesbare Dateien. Als zentrale Herausforderung erwies sich vielmehr die lückenhafte Dokumentation zu den Dateien, deren Inhalte daher nicht mehr verstanden werden konnten. Offenbar ist auch dies keine DDR-spezifische Erfahrung. Die zentrale Bedeutung, die diese Unterlagen für das Verständnis von Datenbeständen und damit für deren Archivierbarkeit besitzen, wird selten hinreichend beachtet, und spätestens nach Abschluß eines DV-Projektes wird seiner Datendokumentation keine Aufmerksamkeit mehr gewidmet. Umso wichtiger ist es, daß sich Archive frühzeitig um die Sicherung solcher Dokumentationsunterlagen und Kontextinformationen bemühen. Die Erfahrungen mit DDR-Dateien haben gezeigt, daß eine nachträgliche Rekonstruktion der Datendokumentation, wenn überhaupt möglich, zumindest sehr aufwendig ist.

Vor allem aber war die Praxis des Umgangs mit Datenbeständen sehr viel lehrreicher als die Fachlektüre. Immer wieder traten im negativen wie im positiven Sinne Phänomene auf, die es eigentlich nicht hätte geben dürfen. So waren fast dreißig Jahre alte Magnetbänder noch fehlerfrei lesbar. Auf Magnetbändern wurden mitten in Dateien *End-of-tape*-Filemarks entdeckt, und es gab eigenwillige Formen der Zeichendarstellung, die auch nach mehrjähriger Praxis Kopfzerbrechen bereiteten. Eine gewisse Skepsis gegenüber Aussagen von Fachleuten oder gegenüber der Fachliteratur ist daher in diesem Bereich durchaus angebracht. Für das Bundesarchiv ist es von großem Vorteil, daß es über eigene Rechentechne verfügt und dadurch eigene Erfahrungen gesammelt werden konnten. Außerdem war es auf diese Weise möglich, sich den besonderen Problemen von DDR-Datenbeständen eingehend zu widmen. In manchen Fällen hätte eine Auftragsvergabe an externe Rechenzentren zu Datenverlusten führen können.



Es ist nicht anzunehmen, daß die hier aufgezeigten Probleme mit Daten aus modernen Systemen abnehmen werden. Die Dateien aus PC-Anwendungen in der DDR, die vom Bundesarchiv übernommen wurden, haben gezeigt, daß diese sehr viel abhängiger von einer spezifischen Software und Hardware waren als solche aus Großrechneranlagen. Auch die verwendeten Disketten waren bereits nach wenigen Jahren veraltet. Die alte Bandtechnologie besaß einen Grad der Standardisierung sowie eine weite und lang andauernde Verbreitung, die die modernen Speichermedien wohl nicht mehr erreichen werden. Außerdem mußte die Erfahrung gemacht werden, daß höhere Speicherkapazitäten auch größere und gravierendere Fehleranfälligkeiten zur Folge hatten. Dies läßt erahnen, welche Schwierigkeiten auf Archive in der Zukunft zukommen werden.

### Ausgewählte Literaturhinweise

- Oswald *Angermann*: Sammlung, Sicherung und Rückrechnung von statistischen Angaben über die ehemalige DDR. In: Allgemeines Statistisches Archiv 78 (1994) S. 340–349.
- Wolf *Buchmann* und Michael *Wettengel*: Auslegung des Bundesstatistikgesetzes bei der Archivierung von Statistikerunterlagen. In: Der Archivar 49 (1996) Sp. 67–74.
- Frido *Dietz* und Helmut *Rudolph*: Berufstätigenerhebung und der Datenspeicher „Gesellschaftliches Arbeitsvermögen“. Statistische Grundlagen zu wichtigen Strukturen der Erwerbstätigen in der vormaligen DDR. In: Mitteilungen aus der Arbeitsmarkt- und Berufsforschung 23 (1990) S. 511–518.
- Wolfgang *Fritz*: Die amtliche Erwerbstätigenstatistik in der DDR, in: Historical Social Research 22 (1997). Heft 3/4. S. 300–357.
- Markus *Güttler*: Die Datenverarbeitung im statistischen Informationssystem der DDR. Ein Beitrag zur deutschen Vereinigung und zur Geschichte der Informatik. Bericht 1990–36 (TU Berlin, FB 20 – Informatik). Berlin 1992.
- Markus *Güttler*: Die Grenzen der Kontrolle. Das statistische Informationssystem und das Versagen zentralistischer Planwirtschaft in der DDR. In: R. *Bessel* und R. *Jessen* (Hg.): Die Grenzen der Diktatur. Staat und Gesellschaft in der DDR. Göttingen 1996. S. 253–273.
- Markus *Güttler*: Das statistische Informationssystem der DDR. Ein Beispiel für die Grenzen computergestützter Entscheidungsunterstützung. Bericht 1992–2 (TU Berlin, FB 20 – Informatik). Berlin 1992.
- Stefan *Hornbostel*: Die geplante Elite – erste Ergebnisse aus der Analyse des Kaderdatenspeichers des Ministerrates. In: Potsdamer Bulletin für Zeithistorische Studien 10 (1997) S. 55–69.
- Peter *Hübner*: Begriffliche und statistische Aspekte einer Sozialgeschichte des industriellen Leitungspersonals der DDR. In: Potsdamer Bulletin für Zeithistorische Studien 10 (1997) S. 70–76.
- Paul *Lüttinger* und Heike *Wirth* (Hg.): Amtliche Quellen der DDR und der neuen Bundesländer. Informationsquelle für die Sozialwissenschaften. Tagungsdokumentation. Mannheim 1993.



- Holger *Mühlbauer*: Kontinuitäten und Brüche in der Entwicklung des deutschen Einwohnermeldewesens. Historisch-juristische Untersuchung am Beispiel Berlins (Europäische Hochschulschriften III 635). Frankfurt am Main u. a. O. 1995.
- Ingrid *Oertel*: Abschluß der Sicherung und Archivierung der gespeicherten ehemaligen DDR-Statistiken auf elektronischen Datenträgern und Ausblick auf weitere Archivierungsarbeiten. In: Allgemeines Statistisches Archiv 80 (1996) S. 251–260.
- Ulf *Rathje*: Der „Datenspeicher Gesellschaftliches Arbeitsvermögen“ der DDR. In: Historical Social Research 21 (1996). Heft 2. S. 113–118.
- Ulf *Rathje*: Datenspeicher Gesellschaftliches Arbeitsvermögen (GAV). Anonymisierte Versionen. Bestand DQ 3 MD/1. Bestandsverzeichnis. Bundesarchiv. Koblenz 1995.
- Ulf *Rathje*: Datenspeicher Gesellschaftliches Arbeitsvermögen (GAV). Bestand DQ 3 MD/2. Bestandsverzeichnis (überarbeitete Fassung). Bundesarchiv. Koblenz 1995.
- Ulf *Rathje*: Die personenbezogenen Massendatenspeicher der DDR im Bundesarchiv. In: Potsdamer Bulletin für Zeithistorische Studien 10 (1997) S. 40–54.
- Ulf *Rathje*: Strafgefangenen- und Verhaftetendateien. Projekt NRB (Statistische Auswertungen über den Arbeitseinsatz Strafgefangener). Bestand DO 1 MD/2. Bestandsverzeichnis. Bundesarchiv. Koblenz 1995.
- Ulf *Rathje*: Strafgefangenen- und Verhaftetendateien. Projekt NRC (Personenerfassung). Bestand DO 1 MD/1. Bestandsverzeichnis. Bundesarchiv. Koblenz 1998.
- Ulf *Rathje*: Der „Zentrale Kaderdatenspeicher“ des Ministerrates der DDR. In: Historical Social Research 21 (1996). Heft 3. S. 137–141.
- Ulf *Rathje*: Zentraler Kaderdatenspeicher (ZKDS). Bestand DC 20 MD/1. Vorläufiges Bestandsverzeichnis. Bundesarchiv. Koblenz 1996.
- Sabine *Roß*: Zur Entstehungsgeschichte des „Zentralen Kaderdatenspeichers“ des Ministerrates der DDR. In: Potsdamer Bulletin für Zeithistorische Studien 10 (1997) S. 48–54.
- Jürgen *Salomon*: Probleme der Genauigkeit bei der Massendatenverarbeitung unter besonderer Berücksichtigung der Fehlerbereinigung und der Fortschreibung. Dargestellt am Beispiel des Projektes „Gesellschaftliches Arbeitsvermögen“. Diss. Berlin (Ost) 1981.
- Udo *Schäfer*: Die Pflicht zur Anbietetung und Übergabe von Unterlagen in der archivischen Praxis. In: Robert *Kretzschmar* (Hg.): Historische Überlieferung aus Verwaltungsunterlagen. Zur Praxis der archivischen Bewertung in Baden-Württemberg (Werkhefte der Staatlichen Archivverwaltung Baden-Württemberg A 7). Stuttgart 1997. S. 35–46.
- Bernhard *Schimpl-Neimanns*, Christa *Sutter* und Heike *Wirth*: Abschlußbericht zum Projekt. Mikrodaten der amtlichen Statistik der DDR bis 1990. Bestandsaufnahme und Nutzungsmöglichkeiten für Sekundäranalysen über soziale Ungleichheit. Mannheim 1993.
- Matthias *Wagner*: Das Nomenklatursystem. Hauptinstrument der Kaderpolitik der SED. Potsdam (Manuskriptfassung).
- Michael *Wettengel*: Archivierung maschinenlesbarer Datenbestände im Bundesarchiv. In: Historical Social Research 20 (1995). Heft 4. S. 123–127.

- Michael *Wettengel*: Zum Stand der Archivierung maschinenlesbarer Daten im Bundesarchiv. In: Mitteilungen aus dem Bundesarchiv 1 (1993). Heft 1. S. 21–23.
- Michael *Wettengel*: Zur Rekonstruktion digitaler Datenbestände aus der DDR nach der Wiedervereinigung: Die Erfahrungen im Bundesarchiv. In: Der Archivar 50 (1997) Sp. 735–748.
- Michael *Wettengel* in Zusammenarbeit mit Ulf *Rathje*: Datenspeicher Gesellschaftliches Arbeitsvermögen der DDR – ein Werkstattbericht. In: Mitteilungen aus dem Bundesarchiv 2 (1994). Heft 3. S. 157–159.
- Jürgen *Wilke*: Quellen zur Statistik der DDR und von maschinenlesbaren Mikrodaten (Register). Vorläufiges Verzeichnis. Köln 1994.
- Heike *Wirth*: Amtliche Bevölkerungserhebungen der DDR als Quelle für sozialstrukturelle Analysen. In: Potsdamer Bulletin für Zeithistorische Studien 10 (1997) S. 25–39.
- Heike *Wirth*: Die faktische Anonymität von Mikrodaten. Ergebnisse und Konsequenzen eines Forschungsprojektes. In: ZUMA Nachrichten 30. Mai 1992. S. 7–42.