

Archivierung von DDR-Daten im Bundesarchiv: ein Rückblick auf zehn Jahre

Rathje, Ulf

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:

Rathje, U. (2003). Archivierung von DDR-Daten im Bundesarchiv: ein Rückblick auf zehn Jahre. *Historical Social Research*, 28(1/2), 57-72. <https://doi.org/10.12759/hsr.28.2003.1/2.57-72>

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>

Archivierung von DDR-Daten im Bundesarchiv – ein Rückblick auf zehn Jahre

*Ulf Rathje**

Abstract: More than ninety percent of the electronic data files repositied at the Bundesarchiv have been taken over from the dissolved central state agencies of the former GDR.

The individual data files are divided into three main groups: personally-related data bases such as the “Zentraler Kaderdatenspeicher” of the Ministers` Council, statistical data collections such as the census data from 1971 and 1981 as well as data collections for specific agency tasks such as the database on incidents at the east-german border. Nowadays data holdings of the former GDR contain about 170 objects. Several data holdings are available as scientific use files and some data bases are analyzed in research projects. Their value for the scientific research of GDR history is much appreciated. Furthermore, a complete conversion and an access to the data files contents will still require many years of work.

1. Datenarchivierung im Bundesarchiv bis 1990

Mehr als neunzig Prozent der im Bundesarchiv archivierten Dateien stammen von zentralen Organen und nachgeordneten Einrichtungen der ehemaligen DDR (vgl. Wettengel 1993a). Ich möchte nachfolgend einen größeren Überblick über die Datenüberlieferung der DDR im Bundesarchiv¹ und Einblicke in meine Arbeit als „Datenarchivar“ geben.

Das Bundesarchiv hat bereits lange vor der Wiedervereinigung, nämlich bereits 1973, die ersten Daten zur Archivierung übernommen, und zwar einen

* Address all communications to Ulf Rathje, c/o Bundesarchiv, Finckensteinallee 63, 12205 Berlin. E-mail: u.rathje@barch.bund.de.

¹ Zum Überblick über die Datenüberlieferung der DDR im Bundesarchiv siehe die früheren Veröffentlichungen: Rathje 1997a, Wettengel 1999 sowie Rathje/Wettengel 1999.

Stichtagsdatenbestand aus dem Ausländerzentralregister² (AZR) des Bundesverwaltungsamts in Köln. Das AZR ist eine ausgesprochen große Datenbank, in der seit 1967 alle Ausländer erfasst werden, die in Deutschland leben oder gelebt haben (derzeit etwa 12 Millionen Datensätze).

Darüberhinaus war das Bundesarchiv in den 1970er Jahren in besonderem Maße an den frühen Forschungsarbeiten aus der historischen Sozialforschung interessiert, die quantifizierende Methoden angewandt haben. Zu den allerersten deutschen Wissenschaftlern, die Archivalien mit quantifizierenden Methoden und der EDV auswerteten, gehörten Heinrich Best³ und Konrad H. Jarausch⁴. Das Bundesarchiv hat die Dateien archiviert, die aus diesen Forschungsarbeiten entstanden sind, und kann sie für Sekundäranalysen zur Verfügung stellen.

2. Erste Datenübernahmen aus der DDR 1991

Mit der Wiedervereinigung wurden auf zentraler Ebene der DDR zahlreiche Behörden und Rechenzentren geschlossen. Die Akten aus diesen Einrichtungen wurden vom Zentralen Staatsarchiv in Potsdam übernommen, das am 3. Oktober 1990 mit dem Bundesarchiv vereinigt wurde. Mit der Archivierung dieser Akten ist heute die Abteilung DDR des Bundesarchivs betraut. Mit der Übernahme der Daten aus den Rechenzentren der ehemaligen DDR wurde in Koblenz 1991 Michael Wettengel beauftragt, der ab sofort quasi als „Einzelkämpfer“ nicht nur die Daten der DDR, sondern auch die aus zivilen und militärischen Einrichtungen der Bundesrepublik archivieren sollte. Erst 1993 kamen zwei Mitarbeiter hinzu (darunter auch der Verfasser) und es konnte ein PC mit

² Bundesarchiv-Bestand B 311 MD (Bundesverwaltungsamt – maschinenlesbare Daten), siehe auch <http://www.bundesverwaltungsamt.de> (unter Aufgaben A-Z: Ausländerzentralregister).

³ Bundesarchiv-Bestand FSg 4 MD (Sammlung zur Geschichte der deutschen Revolution 1848/49 - maschinenlesbare Daten): Heinrich Best wertete 1973 bis 1976 die 3.776 handelspolitischen Petitionen („Eingaben, die zollpolitische Forderungen stellten und/oder auf die Herstellung eines geschlossenen deutschen Binnenmarktes Bezug nahmen“) an die Frankfurter Nationalversammlung für sein Dissertationsvorhaben aus. Dieses Forschungsprojekt war eines der ersten in Deutschland, das Methoden der quantitativen historischen Sozialforschung anwandte. Siehe Best 1980.

⁴ Bundesarchiv-Bestand MD 1 (Dateien aus dem Forschungsprojekt „Soziale Struktur und politische Sozialisation der Studenten im Kaiserreich“ von Konrad H. Jarausch): Mit diesem Forschungsprojekt stellte sich Jarausch 1975 bis 1980 die Aufgabe, den Prozess der gesellschaftlichen Politisierung der akademischen Jugend im Kaiserreich (1871-1914) zu untersuchen, wobei zwei Kernfragen gestellt wurden: zum einen die Frage nach der gesellschaftlichen Herkunft der Studentenschaft, zum anderen die Frage nach der politischen Sozialisierung. Bei dem Forschungsprojekt wurde eine Kombination von quantitativen und qualitativen Verfahren der Historischen Sozialforschung und der Sozialisationsforschung mit Hilfe der EDV angewandt. Siehe Jarausch 1982.

Magnetbandlaufwerken zum ausschließlichen Einsatz für die Sicherungskopie-
rung der Datenbestände beschafft werden (vgl. Wettengel 1993a).

Datei Grenzzwischenfälle

Die erste Datenbank aus DDR-Provenienz, die 1991 ins Bundesarchiv über-
nommen werden konnte, war die sogenannte „Datei Grenzzwischenfälle“, die
1971 bis 1990 im Rechenzentrum des Kommandos der Grenztruppen in Pätz
bei Berlin entstanden war. Entsprechend den geltenden Vorschriften hatten die
Grenzsoldaten alle Grenzzwischenfälle zu protokollieren. Insgesamt wurden
etwa 650.000 Zwischenfälle mit der EDV erfasst; nicht nur die berüchtigten
Mauerschützenfälle und die Fluchtversuche, sondern alle denkbaren Vor-
kommnisse an der Grenze wurden festgehalten. Der Besuch des amerikani-
schen Präsidenten Ronald Reagan 1986 wurde ebenso erfasst wie jede Schul-
klasse auf einer der Aussichtsplattformen an der Berliner Mauer, jeder Militär-
transport, jedes Manöver in Sichtweite der Grenze usw. Diese Dateien stehen
inzwischen für wissenschaftliche Auswertungen zur Verfügung.

Ende 1991 gelangten zwei weitere große Datenbanken ins Bundesarchiv:
der Datenspeicher Gesellschaftliches Arbeitsvermögen (GAV) und der Zentrale
Kaderdatenspeicher (ZKDS).

Datenspeicher Gesellschaftliches Arbeitsvermögen⁵

Der GAV wurde im Auftrag des Staatssekretariats für Arbeit und Löhne aufge-
baut und enthält Daten über den kompletten beruflichen Werdegang von mehr
als 7,25 Mio. DDR-Bürgern. Die Dekomprimierung, Konvertierung und Gene-
rierung einer anonymisierten Version konnte im April 2002 zu Ende geführt
werden. Das Bundesarchiv hat die anonymisierte Version des GAV dem Insti-
tut für Soziologie an der Universität Jena⁶ zur wissenschaftlichen Auswertung
zur Verfügung gestellt.⁷ Auch für andere Forschungsvorhaben kann der GAV
künftig benutzt werden.

Zentraler Kaderdatenspeicher (ZKDS) des Ministerrats der DDR⁸

Der ZKDS wird teilweise bereits seit 1996 im Rahmen dieses Forschungspro-
jekts ausgewertet.⁹ Er enthält Daten zur Biografie von mehr als 600.000 Füh-

⁵ Bundesarchiv-Bestand DQ 3 MD (Staatssekretariat für Arbeit und Löhne – maschinenles-
bare Daten). Siehe: Dietz/Rudolph 1990, Rathje 1994 sowie Rathje 1996a.

⁶ Siehe <http://www.sozioogie.uni-jena.de>.

⁷ Das erste Forschungsprojekt zur Elitenforschung am Institut für Soziologie der Friedrich-
Schiller-Universität Jena wurde vom Land Thüringen 1996 bis 1999 und von der Deutschen
Forschungsgemeinschaft (DFG) 1999 bis 2001 gefördert. Seit Juli 2001 wird das Projekt
unter der Bezeichnung „Teilprojekt A1: Gesellschaftliche Differenzierungsprozesse in der
DDR“ innerhalb des ebenfalls von der DFG geförderten Sonderforschungsbereichs SFB
580 "Gesellschaftliche Entwicklungen nach dem Systemumbruch. Diskontinuität, Tradition
und Strukturbildung" weitergeführt, siehe auch <http://www.sfb580.uni-jena.de/a1.html>.

⁸ Bundesarchiv-Bestand DC 20 MD (Ministerrat der DDR – maschinenlesbare Daten). Siehe
dazu Rathje 1996b.

rungskadern und Angehörigen der sogenannten Kaderreserve aus Staat, Politik und Wirtschaft der DDR. Die Konvertierung sowie die Generierung einer anonymisierten Version für Forschungszwecke des ZKDS konnten im März 2001 abgeschlossen werden. Auch dieser Datenspeicher kann für wissenschaftliche Benutzungen zur Verfügung gestellt werden.

Mit Ablauf des Jahres 1991 waren im Bundesarchiv 361 Magnetbänder archiviert.

3. Entwicklung der Datenbestände aus der DDR 1992-2001

Besonders in den Jahren 1992, 1995 und 1997 hatte es größere Datenübernahmen gegeben, ab 1998 erfolgten nach der Sichtung, archivischen Bewertung und Kopierung aber auch umfangreiche Kassationen von Magnetbändern. Die vorhandene Raumkapazität für mehr als 100.000 Magnetbänder wurde nie benötigt, der Höchststand lag bei etwa 6.000, derzeit sind noch etwa 400 Magnetbänder vorhanden. Für die Lagerung von Magnetbandkassetten, Digital Audio Tapes (DAT) und CDs werden in Zukunft zwei bis drei Datenträgerschränke auf wenigen Quadratmetern ausreichen.

Den Umfang von Datenbeständen in einem Archiv an der Zahl der Datenträger allein zu messen, wäre irreführend, da die Kapazität der unterschiedlichen Medien ganz erheblich voneinander abweicht. Wir haben daher versucht, die Archivalieneinheit bei Daten neu zu definieren. Ein „Digitales Archivobjekt“ (DAO) ist eine DV-Anwendung, z. B. eine Datenbank, eine Fachstatistik oder eine Forschungsdatei, zu der beliebig viele Dateien gehören können (vgl. Martin-Weber 2002). Derzeit umfassen die gesichteten und als archivwürdig bewerteten Datenbestände der DDR 168 DAOs, die der Bundesrepublik 12, das Verhältnis beträgt also etwa 14:1. Insgesamt sind im Bundesarchiv Daten in 24 Beständen archiviert. Vier dieser Bestände sind staatlichen Provenienzen der Bundesrepublik zuzurechnen, 14 der DDR. Sonderbestände, Nachlässe und Sammlungen gibt es weitere sechs. Von den insgesamt 18 staatlichen Beständen der Bundesrepublik und der DDR stammen zwei aus militärischer Provenienz (vgl. Rathje 2001).

Zu den übernommenen DDR-Datenbeständen gehören zum Beispiel:

*Arbeitskräftedatenspeicher Volksbildung (AKDS)*¹⁰

Der AKDS des Ministeriums für Volksbildung enthält Daten zur Biografie der etwa 330.000 Beschäftigten aus dem gesamten Bereich der Volksbildung. Dazu

⁹ Erste Publikationen aus dem Forschungsprojekt: Hornbostel 1997, Best/Hornbostel 1998 sowie Hornbostel 1999.

¹⁰ Bundesarchiv-Bestand DR 2 MD (Ministerium für Volksbildung – maschinenlesbare Daten). Siehe: Rathje 1995.

gehörten die Kindergärten, Horte, Tagesstätten, Heime, Sporteinrichtungen und die allgemeinbildenden Schulen. Eine in ihrem Merkmalsumfang reduzierte anonymisierte Version wurde dem Forschungsprojekt in Jena bereits zur Verfügung gestellt. In Kürze soll jedoch die weitere Bearbeitung der Originaldaten erfolgen. Auch hier stehen Dekomprimierungs-, Konvertierungs- und Anonymisierungsarbeiten an. Auch für andere Forschungsprojekte wird diese Datenbank künftig zur Verfügung gestellt werden können.

Häftlingsdateien¹¹

Aus demselben Rechenzentrum wie das Zentrale Einwohnerregister stammen die Strafgefangenen- und Verhaftetendateien. Im Auftrag der Verwaltung Strafvollzug (VSV) des Ministeriums des Innern (Mdi) wurden mehr als 170 Einzelmerkmale zu den über 350.000 DDR-Häftlingen der Jahre 1980-1990 gespeichert: Angaben zu Festnahme, U-Haft, Strafmaß, Urteil und Entlassung ebenso wie alle denkbaren Informationen über die persönlichen Verhältnisse der Häftlinge bis hin zu ihren sexuellen Neigungen. Es handelt sich um eine hochinteressante Quelle, die bisher noch nicht wissenschaftlich ausgewertet worden ist (vgl. Rathje 1997b und Schröder/Wilke 1997, insbesondere S. 49-61). Vor einer Benutzung steht allerdings noch die Konvertierung der Daten in ein Standardformat und die Generierung einer anonymisierten Version an.

Datei Studienbewerber¹²

Aus dem Rechenzentrum der TU Magdeburg¹³ haben wir die Daten der über 600.000 Studienbewerber der fünf letzten Jahre der DDR übernommen, die bei der Zentralstelle für Studienbewerbungen erfasst worden sind. Diese Dateien sind inzwischen in das Standardformat ASCII csv konvertiert. Eine anonymisierte Version wurde erstellt, so dass diese für wissenschaftliche Auswertungen zur Verfügung gestellt werden kann. Ich stelle mir eine Erforschung der Fragen sehr spannend vor, wer warum einen Studienplatz bekommen hat bzw. warum nicht und wovon das im Detail abhing.

Daten der Außenhandelsbetriebe¹⁴

Auch die Buchführungsdaten der Außenhandelsbetriebe der DDR, die im Auftrag des Ministeriums für Außenhandel in einem eigenen Rechenzentrum verarbeitet wurden, sind im Bundesarchiv archiviert. Vor kurzem musste ich in diesen Daten nach einem Betrag von mehr als sieben Millionen DM recher-

¹¹ Bundesarchiv-Bestand DO 1 MD (Ministerium des Innern – maschinenlesbare Daten), insbesondere DO 1 MD/3 Strafgefangenen- und Verhaftetendateien (Personenerfassung), 1980-1990.

¹² Bundesarchiv-Bestand DR 306 MD (Zentralstelle für Studienbewerbungen – maschinenlesbare Daten).

¹³ Ehemaliges Rechenzentrum der Otto-von-Guericke-Hochschule Magdeburg, heute Universitäts-Rechenzentrum (URZ) der Otto-von-Guericke-Universität (TU) Magdeburg.

¹⁴ Bundesarchiv-Bestand DL 2 MD (Ministerium für Außenhandel und Innerdeutschen Handel – maschinenlesbare Daten).

chieren, die 1990 nach Moskau verschoben worden sind. Die Recherche war relativ schwierig und zeitaufwändig, da die Daten noch nicht kopiert und konvertiert waren, trotzdem war sie erfolgreich. Die Bearbeitung der Buchführungsdaten steht noch aus.

Eingabenprojekt des Staatsrats¹⁵

Die beim Staatsrat eingehenden Eingaben der Bürger wurden in einer Datenbank registriert, wichtige Informationen über den Inhalt der Eingabe, aber auch zur Charakterisierung des Petenten nach einer einheitlichen Systematik erfasst und aufbereitet. Monatlich wurden beim Staatsrat statistische Auswertungen über die Anzahl der Eingaben nach Themengebieten, regionaler Struktur, inhaltlichen Problemen, sozialer Schichtung usw. angefertigt. Die Eingaben der Jahre 1985 bis 1990 sind im Original archiviert. Der Datenbestand umfasst die Jahre 1979 bis 1989 mit insgesamt etwa 1,2 Mio. Datensätzen und eröffnet Auswertungsmöglichkeiten, die weit über die Sichtung der Originale hinausgeht. Quantitative Analysen aus den Informationen zu den Eingaben sind mit vertretbarem Aufwand nur aufgrund der Daten möglich. Personenbezogene Angaben sind nicht enthalten, so dass eine Anonymisierung der Daten nicht erforderlich ist. Eine erste wissenschaftliche Auswertung dieser Daten findet seit 1998 an der Universität Dresden statt (vgl. Elsner 1999).

Projekt „Eingaben der Bürger“ beim Ministerrat¹⁶

Beim Ministerrat wurde erst 1988 ein Projekt zur EDV-Erfassung der Eingaben der Bürger in den Wirkbetrieb übernommen. Von 1988 bis 1990 wurden etwa 60.000 Eingaben in einer REDABAS-Datenbank erfasst. Da die Originaldaten Namen enthalten, habe ich für die Benutzung eine anonymisierte Version erstellt. Eine Verknüpfung und gemeinsame Auswertung der Daten aus den Eingabeprojekten des Staatsrats und des Ministerrats wäre denkbar.

Daten aus dem Rechenzentrum des Ministeriums des Innern

1997 wurde ich in unsere Außenstelle in Dahlewitz-Hoppegarten bei Berlin gerufen, wo bei einem Wasserrohrbruch im Luftschutzkeller die mehr als 3.000 Magnetbänder aus dem sogenannten „Zentralen Einwohnerregister“ (ZER) nass geworden waren. Die Sichtung ergab zum einen, dass die Bänder glücklicherweise keinen Schaden genommen hatten, und zum anderen, dass es sich bei den Daten keineswegs – wie oftmals vermutet – um die zentrale Personendatenbank (PDB) der DDR handelte, sondern um die verschiedensten Daten aus etwa 30 unterschiedlichen EDV-Projekten, die im Rechenzentrum des Ministerium des Innern abgearbeitet worden waren¹⁷. Es waren die Personalda-

¹⁵ Bundesarchiv-Bestand DA 5 MD (Staatsrat der DDR – maschinenlesbare Daten).

¹⁶ Bundesarchiv-Bestand DC 20 MD (Ministerrat der DDR), hier: DC 20 MD/1 Eingaben der Bürger, 1988-1990.

¹⁷ Siehe Landesbeauftragter für den Datenschutz (LfD), 1. Tätigkeitsbericht 1992, Kapitel 5: http://www.brandenburg.de/land/lfdbbg/tb_info/tb1/tb105.htm#5.6.2.

ten der Deutschen Volkspolizei dabei¹⁸, Dateien über alle registrierten Brände in der DDR¹⁹, Daten über die Disziplinarstrafen bei der Deutschen Volkspolizei²⁰, die Kriminalstatistik der DDR²¹ und viele andere. Immer wieder bin ich gefragt worden, wo das Zentrale Einwohnerregister geblieben ist. Nach dem Bericht des Bundesdatenschutzbeauftragten ist es tatsächlich entsprechend dem Einigungsvertrag Anfang 1993 gelöscht worden.²²

Die Wiedervereinigung und somit die Auflösung der Rechenzentren und Behörden liegen inzwischen zwar lange zurück, oftmals wurden DDR-Daten aber noch über Jahre von Bundesbehörden weitergeführt oder ausgewertet. Noch in den letzten beiden Jahren hat das Bundesarchiv Daten aus der DDR übernehmen können.

*Staatliche Zentralverwaltung für Statistik*²³

Das Bundesarchiv hat im Frühjahr 2000 Daten aus 95 verschiedenen Statistiken der Staatlichen Zentralverwaltung vom Statistischen Bundesamt übernommen. Erst 40 Prozent davon sind gesichtet und kopiert. Eine Statistik aus diesem Bestand, nämlich die Todesursachenstatistik der Jahre 1979-1990²⁴, ist bereits in Benutzung beim Max-Planck-Institut für Demografische Forschung in Rostock. Die Nutzbarmachung weiterer Statistikdaten für drei wissenschaftliche Forschungsvorhaben ist derzeit in Vorbereitung. Von besonderem Interesse dürften weiterhin z. B. die Daten der Volkszählungen 1971 und 1981 sein²⁵, die bisher nur in Auszügen vom ZUMA²⁶ in Mannheim ausgewertet werden konnten.

*Gesundheitsstatistiken des Instituts für medizinische Statistik und Datenverarbeitung*²⁷

Die letzten Daten aus diesem 1990 aufgelösten Institut wurden im Oktober 2001 vom Robert-Koch-Institut ins Bundesarchiv übernommen. Zur Hinterlassenschaft dieses Instituts zählen verschiedenste Daten aus der Gesundheitsstati-

¹⁸ Personal-/Kaderbestand der Deutschen Volkspolizei (Datenbank Kader) bis 1990, Projekt NCD, Signatur DO 1 MD/19.

¹⁹ Brandstatistik der DDR 1976-1990, Projekt NQA, Signatur DO1 MD/11.

²⁰ Disziplinarstatistik der Deutschen Volkspolizei 1983-1990, Projekt NUL, Signatur DO 1 MD/5.

²¹ Kriminalstatistik 1976-1990, Projekt NIA, Signatur DO 1 MD/12.

²² Siehe dazu den 14. Tätigkeitsbericht des Bundesbeauftragten für den Datenschutz (BfD) – Berichtszeitraum Anfang 1991 bis Anfang 1993 –, dem Deutschen Bundestag vorgelegt am 27. April 1993, Kapitel 2.2, S. 20-21 (auch als Bundestags-Drucksache 12/4805 veröffentlicht).

²³ Bundesarchiv-Bestand DE 2 MD (Staatliche Zentralverwaltung für Statistik – maschinenlesbare Daten).

²⁴ Signatur DE 2 MD/8-9.

²⁵ Signatur DE 2 MD/27.

²⁶ Zentrum für Umfragen, Methoden und Analysen (ZUMA) in Mannheim, siehe <http://www.gesis.org/zuma/>.

²⁷ Bundesarchiv-Bestand DQ 112 MD (Institut für Medizinische Statistik und Datenverarbeitung – maschinenlesbare Daten).

stik der DDR, z. B. über die Arbeitskräfte und über Hochschulkader im Gesundheitswesen, das Einrichtungsregister des Gesundheitswesens, eine Kennzifferndatenbank Kinderkrippen, verschiedene Krankenhausstatistiken wie das Krankenblatt und die Bettenmeldungen u. a. Die Daten sind erst teilweise gesichtet und kopiert. Wären diese Daten vor fünf Jahren schon im Bundesarchiv gewesen, hätte ich z. B. die Frage eines Benutzers nach Daten über Ärzte in der DDR positiv beantworten können.

4. Technisches Konzept²⁸

Die technische Seite der Archivierung von Daten hat uns als „gelernte Papier-Archivare“ mit völlig neuen Problemen konfrontiert. Im Laufe der Jahre haben wir im Bundesarchiv die Grundsätze des „Technischen Konzepts für die Datenarchivierung im Bundesarchiv“ entwickelt:

4.1 Datenformat

Die zu archivierenden Daten werden im Bundesarchiv grundsätzlich als Flat Files²⁹ softwareunabhängig in dem Standard-Zeichencode ASCII³⁰ gespeichert. Mit diesem bis heute gültigen Grundsatz soll verhindert werden, dass Daten in Zukunft nicht mehr interpretierbar sind, weil es die für ihr Verständnis erforderliche Software nicht mehr gibt oder die alte Software in neuen Systemumgebungen nicht mehr lauffähig ist. Die meisten der bis heute ins Bundesarchiv übernommenen Daten wurden auf Großrechnern erstellt und im Format EBCDIC³¹ gespeichert, in den Rechenzentren wurden jedoch die Regeln des EBCDIC-Standards von IBM selten eingehalten. Es war hingegen üblich, gepackte Formate zu verwenden und mit verschiedensten Komprimierungsverfahren zu arbeiten, um Speicherplatz und Rechenzeit zu sparen. Somit liegen diese Daten in aller Regel nicht in einem archivfähigen Standardformat vor, was eine nachträgliche Bearbeitung erforderlich macht. Das Bundesarchiv hat sich auf das in der PC- und in der UNIX-Welt verbreitete Standardformat ASCII festgelegt und konvertiert die übernommenen Daten entsprechend. Um den Import in Anwendungsprogramme (Datenbankmanagementsysteme wie dBASE, ACCESS, Oracle, Informix oder Statistiksoftware wie SPSS, SAS oder TDA) zu erleichtern, werden die Daten als „comma separated values“ (csv) gespeichert.

²⁸ Siehe ausführlicher dazu Wettengel 1997a sowie Rathje 2002.

²⁹ Flat Files sind als sequentielle Zeichenfolge gespeicherte Daten ohne Software-Elemente.

³⁰ ASCII = American Standard Code for Information Interchange.

³¹ EBCDIC = Extended Binary Coded Decimal Interchange Code.

4.2 Bedingungen für die Eignung von Datenträgern zur Archivierung

Die Eignung der Speichermedien hängt davon ab, ob für sie internationale Standards gelten, ob sie am Markt eine ausgesprochen weite Verbreitung haben, als haltbar gelten und daher auch in anderen Nationalarchiven und Forschungseinrichtungen eingesetzt werden. Mit diesen Grundsätzen soll das Risiko minimiert werden, dass der gewählte Archiv-Datenträger vom Markt verschwindet bzw. überraschend von einem Hersteller nicht mehr produziert wird und nicht mehr gelesen werden kann, weil die Laufwerke nicht mehr verfügbar sind.

Das Magnetband gilt als ein sehr haltbares und zuverlässiges Medium und ist bis heute noch in einigen Nationalarchiven im Einsatz. Auch im Bundesarchiv wurden von 1993 bis 1998 Sicherungskopien (Erstsicherungen) auf Magnetband angefertigt. Heutigen Ansprüchen genügt das Magnetband in der Regel nicht mehr. Seit 1995 werden im Bundesarchiv Sicherungskopien auf Digital Audio Tape (DAT) angefertigt, einem seit etwa 1990 sehr weit verbreiteten Backup-Medium. Das DAT wird zudem in vielen Nationalarchiven als Archivierungsmedium verwendet. Im Bundesarchiv wird die Umstellung auf den neuen Standard DDS-3 (Kapazität 12 GB) derzeit durchgeführt. Die Zweitsicherung erfolgt im Bundesarchiv seit 1998 auf CD-R³². Zum selben Zeitpunkt wurden die Ersicherungen auf Magnetband eingestellt. Die CD dürfte heute einer der am meisten verbreiteten Datenträger am Markt sein und damit vielleicht eine Position erreicht haben, die das Magnetband in den 1970er und 1980er Jahren inne hatte.

4.3 Doppelte Sicherungskopierung auf unterschiedlichen Datenträgern

In den National Archives of Canada, Ottawa, wurden in den 1980er Jahren Daten auf nicht standardisierten 12“-WORMs³³ archiviert. Als diese WORM überraschend schnell vom Markt genommen wurde, mussten diese Daten mit erheblichem finanziellen Aufwand vorzeitig auf andere Datenträger kopiert werden (vgl. Klein 1995). Aufgrund dieser schlechten Erfahrungen wurde in Ottawa eine Arbeitsgruppe eingesetzt, die sich mit der Frage zu befassen hatte, welche Medien für die Langzeitarchivierung geeignet sind³⁴. Eine der zentralen Empfehlungen der Experten war, Daten grundsätzlich doppelt auf zwei ver-

³² Die CD-ROM ist standardisiert durch die Norm ISO 9660 und das sogenannte „Orange Book“ (Spezifikation der Firmen Sony und Philips).

³³ WORM = Write Once Read Many (optische Speicherplatte). Es handelte sich um eine WORM, die nur von einem Hersteller vertrieben wurde und für die es keinen internationalen Standard gab.

³⁴ National Archives of Canada: Report of the Working Group on Conservation Standards and Technologies. Ottawa, 1992.

schiedenen Datenträgern zu archivieren. Auch im Bundesarchiv wird daher jede Datei zweifach und auf zwei verschiedenen Medien gespeichert. Die doppelte Kopierung der Daten auf einen magnetischen (DAT) und einen optischen Datenträger (CD-R) bringt aufgrund der Verschiedenartigkeit der Medien zusätzliche Sicherheit bei Hardwareausfällen und Marktveränderungen.

4.4 Sachgerechte Lagerung

Erst- und Zweitsicherungen sind aus Sicherheitsgründen getrennt voneinander in verschiedenen Archivmagazinen aufzubewahren. Mit der getrennten Lagerung der beiden Archivdatenträger wird das Konzept der doppelten Sicherung konsequent fortgeschrieben (vgl. Wettengel 1997a, insbesondere S. 9). Darüberhinaus sind sie in speziellen Datenträgerschränken abzulegen. Die Räumlichkeiten sollen klimatisiert sein, was besonders für Bandmedien von Bedeutung ist. Vor allem sollen die Klimawerte möglichst konstant gehalten werden: Raumtemperatur 17-20°C, relative Luftfeuchtigkeit 35-45%³⁵.

4.5 Regelmäßige Kopierung und Migration der Medien

Um Datenverlusten vorzubeugen, sind die Sicherungskopien regelmäßig umzukopieren.³⁶ Die Umkopierfristen orientieren sich an Herstellerangaben, an Ergebnissen aus Labortests und an Erfahrungen anderer Archive. Aufgrund der ständigen Ablösung der Speichermedien durch neue Typen gilt es, die Entwicklungen auf dem Markt kontinuierlich zu beobachten und zu gegebener Zeit die Kopierung der Daten auf modernere Medien vorzunehmen (Migration). Es bleibt zu abzuwarten, welche Datenträger sich in den nächsten Jahren auf dem Markt durchsetzen. Auch die Erfahrungen anderer Archive sind in die Strategie einzubeziehen. Perspektivisch ist davon auszugehen, dass das DAT durch andere magnetische Speichermedien höherer Kapazität ersetzt wird. Es ist weiterhin damit zu rechnen, dass bei den optischen Medien die DVD-R die CD-R vom Markt verdrängen wird.

5. Technische Probleme³⁷

Wir hatten immer wieder erhebliche Schwierigkeiten mit der Lesbarkeit der alten und oft verschlissenen Magnetbänder. Die Kopierung war daher absolut

³⁵ Bundesarchiv: Anweisung für die archivarisches Tätigkeit (1.1): Übernahme, Bewertung, Erschließung, Konservierung und Benutzung maschinenlesbarer Dateien (Stand 28.08.1995), S. 9, Wettengel 1997a, S. 9.

³⁶ Digital Audio Tapes (DAT) alle fünf, CD-Rs alle zwei Jahre.

³⁷ Zu den technischen Problemen siehe auch: Rathje 1997a, insb. S. 40-41, Wettengel 1997b, insb. Sp. 738-739.

vordringlich, da oftmals Datenverluste bereits eingetreten waren. Glücklicherweise waren aber oft mehrere Kopien einer Datei auf verschiedenen Bändern vorhanden, so dass die Datenverluste in den meisten Fällen später durch sogenannte „Compare-Läufe“ behoben werden konnten. In Gesprächen mit ehemaligen Mitarbeitern von DDR-Rechenzentren habe ich erfahren, dass oft bis zu vier Sicherheitskopien einer Datei auf vier verschiedenen Bändern angefertigt wurden und dass es hin und wieder tatsächlich vorkam, dass auf die vierte Sicherheitskopie zurückgegriffen werden musste, obwohl die Bänder jedes Jahr wieder kopiert wurden.

Wir hatten weiterhin seit 1991 nicht nur Magnetbänder, sondern auch etwa 1.700 Robotron-Disketten übernommen (in den Formaten 8“ und 5¼“), die in DOS-PCs nicht lesbar sind. Alleine die Verfügbarkeit der passenden Diskettenlaufwerke war ein Problem, ganz abgesehen von den spezifischen Besonderheiten (Spuren- und Sektorenaufteilung) der Robotron-Disketten. Es gab nur eine Lösung, nämlich Exemplare derjenigen Rechner zu beschaffen, mit denen diese Disketten beschrieben worden waren. Seit 1996 habe ich daher eine kleine Sammlung von Robotron-PCs verschiedener Typen aufgebaut, um diese Disketten überhaupt sichten und archivisch bewerten zu können. Diese Rechner standen in anderen Behörden entweder zur Verschrottung oder zur Abgabe ans Museum bereit. Damit konnte ein großer Teil der Disketten noch gelesen werden. Nach Abschluss der Arbeiten an diesen Disketten sowie nach Kopierung der Daten auf moderne Speichermedien wird diese kleine Sammlung an ein Computermuseum übergeben werden können.

6. Das Recht auf Persönlichkeitsschutz und die Interessen der Wissenschaft

Auf das schwierige Thema Datenschutz und die daraus folgenden Benutzungsbeschränkungen bei personenbezogenen Datenbeständen im Bundesarchiv möchte ich kurz eingehen: Im Bundesdatenschutzgesetz³⁸ § 1 Abs. 1 heißt es: „Zweck dieses Gesetzes ist es, den einzelnen davor zu schützen, dass er durch den Umgang mit seinen personenbezogenen Daten in seinem Persönlichkeitsrecht beeinträchtigt wird.“ Das Persönlichkeitsrecht gehört zu den höchsten vom Grundgesetz geschützten Werten. Das Bundesverfassungsgericht hat in seinem Volkszählungsurteil am 15. Dez. 1983 entschieden: „Das Grundrecht gewährleistet insoweit die Befugnisse des Einzelnen, grundsätzlich selbst über die Preisgabe und Verwendung seiner persönlichen Daten zu bestimmen.“ Einschränkungen des Rechts auf „informationelle Selbstbestimmung“ bedürfen jeweils der gesetzlichen Grundlage. Die Bestimmungen des Datenschutzes

³⁸ Bundesdatenschutzgesetz (BDSG) vom 20. Dezember 1990 (BGBl. I S. 2954), zuletzt geändert durch Artikel 21 des Gesetzes vom 3. Dezember 2001 (BGBl. I S. 3306).

gelten selbstverständlich auch für das Bundesarchiv. Es hat prinzipiell von der Übernahme an die schutzwürdigen Belange Betroffener zu berücksichtigen, insbesondere bei Unterlagen mit personenbezogenen Daten. So sieht das Bundesarchivgesetz³⁹ nach § 5 Abs. 2 vor, dass „Archivgut des Bundes, das sich auf natürliche Personen bezieht, erst 30 Jahre nach dem Tode des Betroffenen durch Dritte benutzt werden darf. Ist das Todesjahr nicht oder nur mit unvertretbarem Aufwand festzustellen, endet diese Schutzfrist 110 Jahre nach Geburt des Betroffenen.“

Diese Schutzfristen, die den Betroffenen dienen, können allerdings unter bestimmten Voraussetzungen verkürzt werden. So ermöglicht § 5 Abs. 5 BArchG die Verkürzung von Schutzfristen, „wenn die Benutzung für ein wissenschaftliches Forschungsvorhaben oder zur Wahrnehmung berechtigter Belange unerlässlich ist, die im überwiegenden Interesse einer anderen Person oder Stelle liegen, und eine Beeinträchtigung schutzwürdiger Belange durch angemessene Maßnahmen, insbesondere durch Vorlage anonymisierter Reproduktionen, ausgeschlossen werden kann.“ In der Praxis bedeutet dies, dass bei Benutzungen einer Datei in den Einzeldatensätzen bestimmte Felder wie Name, Personenkennzahl, Wohnadresse, Geburtsdatum usw. gelöscht werden müssen, so dass eine Einzelperson nur mit einem unverhältnismäßig hohen Aufwand an Zeit, Kosten und Arbeitskraft identifiziert werden kann. Im Interesse der Forschung sollen auf der anderen Seite jedoch die Daten so wenig wie möglich verändert werden, um das Analysepotential nicht gravierend einzuschränken, so dass es im Einzelfall erforderlich sein wird, einen Kompromiss zwischen Datenschutz und Wissenschaft zu finden, der die Interessen beider Seiten ausreichend befriedigt.⁴⁰

7. Erschließungsarbeiten⁴¹

Es ist Aufgabe des Archivars, die Archivalien - zumeist handelt es sich ja um Akten - in Findhilfsmitteln, also in Findbüchern und Datenbanken, so zu verzeichnen, dass der Benutzer eine Chance hat, das Gesuchte auch zu finden. Die Erschließung von Daten hat aber darüber hinaus noch eine weitere Funktion: Die Daten können nach ihrer Erschließung nicht nur wiedergefunden werden, sondern der Benutzer erhält konkrete Informationen über den Umfang der Daten, die Datenstrukturen, die genauen Inhalte der Daten, auch über Codierungen und Auswertungsmöglichkeiten. Spezifische Unterlagen zu diesen

³⁹ Gesetz über die Sicherung und Nutzung von Archivgut des Bundes (Bundesarchivgesetz – BArchG) vom 6. Januar 1988 (BGBl. I S. 62), zuletzt geändert durch Gesetz zur Änderung des Bundesarchivgesetzes vom 5. Juni 2002 (BGBl. I S. 1782).

⁴⁰ Zum Thema faktische Anonymisierung von Daten siehe Wirth 1992, Müller u. a. 1991, Müller/Blien/Wirth 1995 sowie Müller/Wirth 1995.

⁴¹ Zum Thema Erschließung von Daten siehe insbesondere: Martin-Weber 2002.

Punkten stellt der Archivar in einem Dossier zusammen. Dazu muss er selbst vorher die Dokumentationsunterlagen auswerten, die mit den Daten übergeben worden sind. Die Erfahrung mit den DDR-Daten hat gezeigt, dass diese Dokumentation oftmals nicht mehr vollständig vorhanden sind, was einen zusätzlichen Arbeitsaufwand für den Archivar bedeutet, da er in den Schriftgutbeständen nach Dokumentationsunterlagen recherchieren muss, die von denselben Provenienzstellen stammen. Glücklicherweise waren diese Recherchen in vielen Fällen erfolgreich.

Zum Beispiel stellte die Dokumentation zum bereits mehrfach genannten Zentralen Kaderdatenspeicher ein besonderes Problem dar. Die Dokumentationsunterlagen zum ZKDS wurden von den Mitarbeitern des Rechenzentrums in den Reißwolf gesteckt, als ihr Arbeitsverhältnis gekündigt und das Rechenzentrum privatisiert wurde, mit dem Argument, es würde sich immerhin um ihr persönliches geistiges Eigentum handeln, und das dürfe nicht in fremde Hände gelangen. Tatsächlich dürfte dahinter die Absicht gestanden haben, den Zugang zu den Daten unmöglich zu machen. Der Auftraggeber für den ZKDS war die Abteilung Kader beim Ministerrat und glücklicherweise sind die Akten der Abteilung Kader in Berlin erhalten geblieben, so dass aus dieser Überlieferung eine halbwegs vollständige Ersatzdokumentation gebildet werden konnte. Wenn dies nicht der Fall gewesen wäre, hätte ich die Daten als nicht archivwürdig kassieren müssen, da ihre Inhalte nicht verständlich und interpretierbar gewesen wären (vgl. Wettengel 1997b).

8. Aktuelle Entwicklung und Ausblick

Mehr als zehn Jahre nach der Übernahme der ersten DDR-Daten in das Bundesarchiv sind bei weitem noch nicht alle Probleme gelöst, so dass erst ein Teil der Daten für die wissenschaftliche Benutzung zur Verfügung gestellt werden kann:

- Es sind noch erhebliche Rückstände bei der Sichtung und Sicherungskopierung vorhanden.
- Erst ein Teil der Datenbestände konnten nach ASCII csv konvertiert werden. Dazu sind Programmierarbeiten erforderlich, die das Bundesarchiv außer Haus vergeben muss.
- Erst zu einem Teil der personenbezogenen Daten konnten anonymisierte Versionen für die wissenschaftliche Benutzung generiert werden.
- Eine inhaltliche Erschließung und Dokumentierung der Daten konnte ebenfalls erst für einen Teil der Datenbestände erfolgen.

Hauptgrund für diese Defizite ist das Fehlen von Personal. Darüber hinaus ist das Bundesarchiv im November 2001 in Teilen umstrukturiert worden, u. a. wurde das Referat für die Archivierung von Daten zehn Jahre nach seiner Ein-

richtung aufgelöst. Zum 2. Mai 2002 wurde ein Projekt ins Leben gerufen, mit der Aufgabe, vor allem die inhaltliche Erschließung und technische Bearbeitung der 20 wichtigsten digitalen Archivobjekte innerhalb von zwei Jahren zu Ende zu führen. Die Bearbeitung der übrigen etwa 150 digitalen Archivobjekte muss weiterhin zurückgestellt werden, bis neue Prioritäten zugunsten der Datenarchivierung gesetzt werden und zusätzliches Personal zur Verfügung steht.

Literatur

- Best, Heinrich (1980): Interessenpolitik und nationale Integration 1848/49. Handelspolitische Konflikte im frühindustriellen Deutschland (Kritische Studien zur Geschichtswissenschaft Bd. 37). Diss., Göttingen.
- Best, Heinrich/Hornbostel, Stefan (1998): Prozeß-produzierte Daten als empirisches Material für eine Soziologie des realen Sozialismus. Das Beispiel des Kaderdatenspeichers des Ministerrates der DDR. In: GESIS et al. (Hg.): Materialien zur Erforschung der DDR-Gesellschaft. Opladen, S. 201-221.
- Dietz, Frido/Rudolph, Helmut (1990): 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, S. 511-518.
- Elsner, Steffen H. (1999): Das „EDV-Eingaben-Projekt“ des Staatsrates der DDR: Stationen einer Odyssee. In: Historical Social Research/Historische Sozialforschung 24, S. 135-149 (sowie Mitteilungen aus dem Bundesarchiv 7, S. 12-19).
- Hornbostel, Stefan (1997): Die geplante Elite - erste Ergebnisse aus der Analyse des Kaderdatenspeichers des Ministerrates. Potsdamer Bulletin für Zeithistorische Studien Nr. 10, S. 55-69.
- Hornbostel, Stefan (1999): Die besten Vertreter der Arbeiterklasse. Kaderpolitik und gesellschaftliche Differenzierungsmuster im Spiegel des zentralen Kaderdatenspeichers des Ministerrates der DDR. In: Stefan Hornbostel (Hg.): Sozialistische Eliten. Horizontale und vertikale Differenzierungsmuster in der DDR (Soziologie der Politik Bd. 2). Opladen, S. 177-209.
- Jarausch, Konrad Hugo (1982): Students, Society, and Politics in Imperial Germany. The Rise of Academic Illiberalism. Princeton.
- Klein, Stefan (1995): Fröhlicher Wildwuchs. In: DER SPIEGEL Nr. 40/1995, S. 228-230.
- Martin-Weber, Bettina (2002): Erschließung und Nutzbarmachung digitaler Unterlagen im Bundesarchiv. In: Karl-Ernst Lupprian (Hg.): Virtuelle Welten im Magazin. Aussonderung, Aufbewahrung, Sicherung und Nutzung. Vorträge der 5. Tagung des Arbeitskreises "Archivierung von Unterlagen aus digitalen Systemen" in München, 5. und 6. März 2001. Im Druck (Stand August 2002).

- Müller, Walter u. a. (1991): Die faktische Anonymität von Mikrodaten, (Forum der Bundesstatistik Bd. 19). Stuttgart.
- Müller, Walter/Blien, Uwe/Wirth, Heike (1995): Identification Risks of Microdata. Evidence from experimental studies. *Sociological Methods & Research* 24, S. 131-157.
- Müller, Walter/Wirth, Heike (1995): Research Needs for European Microdata and Data Confidentiality. International Seminar on Statistical Confidentiality. Proceedings. 28 to 30 November 1994. Luxembourg, S. 235-240.
- Rathje, Ulf (1994): Datenspeicher Gesellschaftliches Arbeitsvermögen der DDR – ein „Werkstattbericht“. In: *Mitteilungen aus dem Bundesarchiv* 2, S. 157-159.
- Rathje, Ulf (1995): Übernahme maschinenlesbarer Daten: Der Arbeitskräftedatenspeicher des Ministeriums für Volksbildung der DDR. In: *Mitteilungen aus dem Bundesarchiv* 3, S. 15.
- Rathje, Ulf (1996a): Der „Datenspeicher Gesellschaftliches Arbeitsvermögen“ der DDR. In: *Historical Social Research / Historische Sozialforschung* 21, S. 113-118.
- Rathje, Ulf (1996b): Der „Zentrale Kaderdatenspeicher“ des Ministerrates der DDR. In: *Historical Social Research / Historische Sozialforschung* 21, S. 137-141.
- Rathje, Ulf (1997a): Die personenbezogenen Massendatenspeicher der DDR im Bundesarchiv. In: *Potsdamer Bulletin für Zeithistorische Studien* Nr. 10, S. 40-47 (sowie in: *Mitteilungen aus dem Bundesarchiv* 5, S. 36-40).
- Rathje, Ulf (1997b): Die Strafgefangenen- und Verhaftetendateien der Verwaltung Strafvollzug des Ministeriums des Innern der DDR. In: *Historical Social Research / Historische Sozialforschung* 22, S. 132-140.
- Rathje, Ulf (2001): Maschinenlesbare Dateien im Bundesarchiv - Datenbestandskatalog. Stand: Oktober 2001. Unveröffentlichtes Typoskript.
- Rathje, Ulf (2002): Technisches Konzept für die Datenarchivierung im Bundesarchiv. In: *Der Archivar* 55, S. 117-120.
- Rathje, Ulf/Wettengel, Michael (1999): Digitale Datenbestände von Behörden und Einrichtungen der DDR im Bundesarchiv. *Historical Social Research/Historische Sozialforschung* 24, S. 70-101.
- Schröder, Wilhelm Heinz/Wilke, Jürgen (1997): Politische Gefangene in der DDR. Eine quantitative Analyse. Wissenschaftliche Expertise für die Enquête-Kommission des Deutschen Bundestags „Überwindung der Folgen der SED-Diktatur im Prozess der deutschen Einheit“. Köln, Zentralarchiv für empirische Sozialforschung.
- Wettengel, Michael (1993a): System zur Archivierung maschinenlesbarer Daten im Bundesarchiv. In: *Mitteilungen aus dem Bundesarchiv* 1, S. 70-72.
- Wettengel, Michael (1993b): Zum Stand der Archivierung maschinenlesbarer Daten im Bundesarchiv. In: *Mitteilungen aus dem Bundesarchiv* 1, S. 21-23.
- Wettengel, Michael (1997a): Technische Infrastruktur für die Archivierung von digitalen Datenbeständen - Anforderungen und Verfahrensweisen. In: *Mitteilungen aus dem Bundesarchiv* 5, S. 8-15.

- Wettengel, Michael (1997b): Zur Rekonstruktion digitaler Datenbestände aus der DDR nach der Wiedervereinigung: Die Erfahrungen im Bundesarchiv. In: *Der Archivar* 50, Sp. 735-748.
- Wettengel, Michael (1999): Archivierung digitaler Datenbestände aus der DDR nach der Wiedervereinigung. In: Udo Schäfer/Nicole Bickhoff (Hg.): *Archivierung elektronischer Unterlagen (Werkhefte der Staatlichen Landesarchivverwaltung Baden-Württemberg, A 13)*. Stuttgart, S. 223-239.
- Wirth, Heike (1992): Die faktische Anonymität von Mikrodaten: Ergebnisse und Konsequenzen eines Forschungsprojektes. *ZUMA-Nachrichten* 30.