

Daten der Staatlichen Zentralverwaltung für Statistik der DDR im Bundesarchiv - Bewertung, Übernahme, technische Bearbeitung und Benutzung

Ulf Rathje

1. Die Statistik – Dichtung oder Wahrheit?

„Glaube keiner Statistik, die du nicht selbst gefälscht hast“ – dieser, wohl fälschlich Churchill zugeschriebene Ausspruch, bringt die gängigen Zweifel an der Glaubwürdigkeit statistischer Daten auf den Punkt. Das Bundesarchiv verfügt über Daten aus mehr als 90 Erhebungen der Staatlichen Zentralverwaltung für Statistik der DDR (SZS). Angesichts des hohen Arbeitsaufwandes, diese Daten „für immer“ zu erhalten, muss zunächst die Frage nach ihrer Zuverlässigkeit als zentralem Kriterium für ihre Archivwürdigkeit gestellt werden. Die Glaubwürdigkeit der von der DDR-Regierung publizierten statistischen Zahlen¹ wurde zu Recht immer wieder angezweifelt. Manipulationen sollen im Politbüro des ZK der SED an der Tagesordnung gewesen sein, da die statistischen Zahlen zu Propagandazwecken eingesetzt wurden. Politbüromitglied Günter Mittag wird dabei eine maßgebliche Rolle zugeschrieben.² Aber nicht nur bei den publizierten statistischen Zahlen, sondern auch bei dem statistischen Urmaterial, also den statistischen Daten, wurden immer wieder Zweifel laut.³ Genau hier aber muss differenziert werden. Die Fachleute bei der Staatlichen Zentralverwaltung für Statistik waren selbstverständlich dazu verpflichtet, korrekte Erhebungen durchzuführen, und die Daten wurden nach ihrer Erfassung in mehreren Arbeitsgängen auf auftretende Fehler überprüft. Wie mir mehrere ehemalige Mitarbeiter der Staatlichen Zentralverwaltung für Statistik übereinstimmend versicherten, hätten sie mit einer Verfälschung der Daten im Stadium der Erfassung und Verarbeitung gegen ihre Berufsehre verstoßen.⁴ Korrekte Daten waren für die Arbeit der SZS und der Staatlichen Plankommission unerlässlich. Von Manipulationen an den Urdaten selbst ist daher nicht auszugehen.

2. Die Archivierung der Statistikdaten der DDR im Bundesarchiv

Bis zur Wiedervereinigung wurde bei der Staatlichen Zentralverwaltung für Statistik am Alexanderplatz in Berlin die Erhebung von 223 Statistiken koordiniert und die daraus entstehenden Daten endgültig verarbeitet. Die Erhebungen fanden in den Kreistellen der Staatlichen Zentralverwaltung statt. Die Daten stammen aus den verschiedensten Gebieten des zentralisierten Berichtswesens der Statistik, wobei die Erhebung überwiegend unternehmens- bzw. betriebsbezogen erfolgte. Mit der Wiedervereinigung wurden die

¹ Z. B. Statistisches Jahrbuch für die Deutsche Demokratische Republik 1955-1990, Presseberichte u. v. m. Siehe dazu insbesondere: DDR-Statistik, Sonderreihe Bd 34, Kapitel 8: Die Informations- und Analysetätigkeit (S. 62-74).

² H. Praschek, Besonderheiten der amtlichen Statistik in der ehemaligen DDR, in: Wirtschaft und Statistik, 1/2000, S. 24-29.

³ DDR-Statistik. Grundlagen, Methoden und Organisation, in: Sonderreihe mit Beiträgen für das Gebiet der ehemaligen DDR. Arbeitsunterlage, Bd. 34. Statistisches Bundesamt, Wiesbaden 1999, insbesondere S. 356-394: Zusammenstellung von Meinungsäußerungen, Standpunkten und Einschätzungen zur Statistik und statistischen Ergebnissen aus der DDR.

Staatliche Zentralverwaltung⁵ und das dazugehörige Datenverarbeitungszentrum⁶ am 3. Oktober 1990 aufgelöst und in das „Gemeinsame Statistische Amt der fünf neuen Bundesländer“⁷ umgewandelt, später wurde daraus die Zweigstelle Berlin-Alexanderplatz des Statistischen Bundesamts. Zu den Aufgaben des Gemeinsamen Statistischen Amtes gehörten vor allem die Sicherung und Dokumentierung der DDR-Daten sowie die Vorbereitung der Übergabe an die Statistischen Landesämter in den neuen Bundesländern. Darüberhinaus waren teilweise Rückrechnungsarbeiten an den Daten erforderlich, um sie zumindest partiell an die Daten der Bundesstatistik anzupassen.⁸ Die Ergebnisse der Aufbereitung wurden ab 1993 vom Statistischen Bundesamt in einer eigenen Veröffentlichungsreihe publiziert.⁹ Ende 1994 war die Übergabe der sogenannten „absortierten Länderdateien“ an die Statistischen Landesämter abgeschlossen. Immerhin wurden Daten aus 124 Statistiken, also etwa 70 % aller Erhebungen für die Länder kopiert. Der Betrieb des Rechenzentrums wurde am 15. Dezember 1994 endgültig eingestellt.¹⁰

Die Staatliche Plankommission der DDR erhielt von den wichtigsten Statistiken regelmäßig Kopien. Nach ihrer Auflösung hatte das Bundesarchiv 1991 Statistikdaten auf etwa 720 Magnetbändern von der Plankommission übernehmen können, die es nun dem Statistischen Bundesamt zur Verfügung stellte. Da bei der Plankommission die Aufbewahrungsfristen in der Regel länger waren als bei der Staatlichen Zentralverwaltung, konnten mit diesen Daten die im Statistischen Bundesamt bereits vorhandenen Statistiken weiter ergänzt werden. Im Februar 1995 erhielt das Bundesarchiv neu erstellte Kopien der Statistikdaten der Plankommission, allerdings waren in diesen Kopien die Hilfsmerkmale (v. a. Betriebsnummern) gelöscht worden.¹¹ Durch diese Löschung waren die Daten unbrauchbar geworden, da Informationen nicht mehr verknüpft und zugeordnet werden konnten. Daher stimmte das Bundesministerium des Innern mit Erlass vom 4. Januar 1996¹² grundsätzlich einer Archivierung der Statistikdaten im Bundesarchiv ohne Vornahme von Löschungen zu.

Als 1998 im Bundesarchiv bekannt wurde, dass die Zweigstelle Berlin des Statistischen Bundesamts aufgelöst werden sollte, nahmen wir mit den Statistikern Kontakt auf, in der Befürchtung, dass die Statistikdaten der DDR ungeachtet der Anbietungspflicht vernichtet werden könnten. Bereits im Dezember 1998 begannen die Bewertungsgespräche am Alexanderplatz. 94 Statistiken sowie die sogenannte SPAZ-

⁴ Mündliche Auskünfte während den Besprechungen zur Bewertung der Statistiken Dez. 1998/Jan. 1999.

⁵ Zuletzt unter dem Namen „Statistisches Amt der DDR“.

⁶ VEB DVZ Statistik (Volkseigener Betrieb Datenverarbeitungszentrum Statistik).

⁷ Das Gemeinsame Statistische Amt der fünf neuen Bundesländer (GeStAL) führte entsprechend dem Einigungsvertrag (Anlage I, Kapitel VIII Statistik, § 3) die Aufgaben des Statistischen Amtes der DDR fort.

⁸ O. Angermann, Sammlung, Sicherung und Rückrechnung von statistischen Angaben über die ehemalige DDR, in: Allgemeines Statistisches Archiv, 78/1994, S. 340-349.

⁹ Sonderreihe mit Beiträgen für das Gebiet der ehemaligen DDR. Arbeitsunterlage, Bd. 1-34. Statistisches Bundesamt, Wiesbaden 1993-2000.

¹⁰ I. 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.

¹¹ H. Praschek, Erschließung und Verbleib der Einzeldatenbestände der amtlichen Statistik der DDR, in: Wirtschaft und Statistik, 7/2000, S. 535.

Datenbank wurden vom Bundesarchiv schließlich als archivwürdig bewertet. Zwischen August 1999 und Oktober 2000 konnten die ausgewählten Daten ins Bundesarchiv übernommen und dem Datenbestand DE 2 MD zugeordnet werden.¹³ Die als archivwürdig bewerteten Daten stammen aus allen Bereichen der DDR-Statistik, vor allem aber aus den Bereichen Arbeitskräfte, Bevölkerung (z. B. Daten der Volkszählungen 1971 und 1981), Kultur, Bildung und Wissenschaft, Industrie, Umweltschutz und Landwirtschaft und haben eine Gesamtlaufzeit von 1976 bis 1991.

Auch die Daten aus der Gesundheitsstatistik wurden zusammen mit den anderen Statistikdaten vom Statistischen Bundesamt übernommen, sie stammen jedoch aus einer anderen Provenienz: Für die Gesundheitsstatistik war in der DDR das Ministerium für Gesundheitswesen zuständig, das diese Aufgabe wiederum an das ihm nachgeordnete Institut für Medizinische Statistik und Datenverarbeitung delegiert hatte. Im Bundesarchiv wurde aus diesen Statistiken der Datenbestand DQ 112 MD¹⁴ gebildet. Exemplarisch wären hier zu nennen: Statistiken über Arbeitskräfte des Gesundheitswesens und das Hochschulkaderprojekt, das Einrichtungsregister des Gesundheitswesens oder die Krankenhausstatistiken wie das sogenannte Krankenblatt und die Bettenmeldungen.

Insgesamt übernahm das Bundesarchiv aus der Provenienz Staatliche Zentralverwaltung für Statistik 235 Magnetbandkassetten mit Daten aus 94 statistischen Erhebungen sowie aus der Datenbank SPAZ, aus der Provenienz des Instituts für Medizinische Statistik und Datenverarbeitung zehn Magnetbandkassetten mit Daten aus 16 Gesundheitsstatistiken.

Zu den Daten jeder statistischen Erhebung wurde eine sogenannte „Archivierungsakte“ übergeben, die im Rahmen der Sicherungskopierungs- und Dokumentierungsarbeiten vom Gemeinsamen Statistischen Amt bzw. von der Zweigstelle Berlin des Statistischen Bundesamts zusammengestellt worden war. Die Archivierungsakten bilden einen wichtigen Grundstock für die Datendokumentation und somit für die zukünftigen Findmittel. Die Ordnungsarbeiten sind inzwischen weitgehend abgeschlossen.

3. Die Bewertung der Statistikdaten¹⁵

¹² Bundesministerium des Innern, Geschäftszeichen OI II 3 – 142 002/16.

¹³ Bundesarchiv, Staatliche Zentralverwaltung für Statistik – maschinenlesbare Daten, DE 2 MD.

¹⁴ Bundesarchiv, Institut für Medizinische Statistik und Datenverarbeitung – maschinenlesbare Daten, DQ 112 MD.

¹⁵ Die Bewertung der Statistikdaten erfolgte innerhalb des Bundesarchivs in enger Zusammenarbeit mit Gisela Haker, (Dipl.-Archivarin, Referatsleiterin in der Abt. DDR des Bundesarchivs, u. a. zuständig für den Schriftgutbestand DE 2 Staatliche Zentralverwaltung für Statistik), die über viele Jahre Erfahrung mit der Archivierung von Statistikunterlagen verfügt. Sehr fruchtbar waren weiterhin die Gespräche mit Angela Ullmann (1994-2001 Dipl.-Archivarin (FH) im Sächsischen Hauptstaatsarchiv Dresden, u.a. zuständig für die Überlieferung des Statistischen Landesamtes des Freistaates Sachsen), die sich 1998/1999 als im Hauptstaatsarchiv Dresden mit der Bewertung der Schriftgut- und der Datenüberlieferung der DDR-Bezirke Dresden, Karl-Marx-Stadt und Leipzig beschäftigt hat. Die derzeit noch im Statistischen Landesamt in Kamenz liegenden absortierten Länderdateien der drei genannten Bezirke stellen eine Teilmenge des Statistikdatenbestands im Bundesarchiv dar, so dass die Bewertungsentscheidungen gemeinsam diskutiert und teilweise abgestimmt werden konnten. Weiterhin habe ich ganz bewusst auch den Sachverstand der

Christian Keitel¹⁶ hat auf der fünften. Tagung des Arbeitskreises "Archivierung von Unterlagen aus digitalen Systemen" am 5. und 6. März 2001 in München einen umfassenden Überblick über die archivfachliche Literatur zum Thema Bewertung von Statistikdaten gegeben und darüberhinaus die Überlegungen der Stuttgarter Kollegen zur Bewertung von Statistikdaten dargestellt.¹⁷ Die in Baden-Württemberg angewandten Bewertungskriterien bestätigen in vielen Punkten die Praxis der Bewertung von Statistikdaten im Bundesarchiv und ich glaube daraus schließen zu können, dass bei den wenigen Archivaren in Deutschland, die sich mit der Bewertung von Statistikdaten beschäftigen, ein Konsens bezüglich der Bewertungsgrundsätze für diese spezielle Quellengattung besteht.

Einige *grundsätzliche Überlegungen* waren zu Beginn anzustellen, denn eine Reihe von Gesichtspunkten waren für die Bewertung von großer Bedeutung und sind teilweise unabdingbare Voraussetzung für ihre Archivwürdigkeit:

- Statistiken können nur in ihrer originären, elektronischen Form, also als *Daten* archivwürdig sein. Die ausgefüllten Erfassungsformulare dagegen sind kassabel, da eine Benutzung völlig unwirtschaftlich wäre, da die Daten ein zweites Mal erfasst werden müssten.
- Das Datenmaterial soll *Plausibilitätsprüfungen* durchlaufen haben. Die *bereinigten Basisdaten* „stimmen ... in größerem Maße mit den abzubildenden Phänomenen überein“ als das Urmaterial auf Papier.¹⁸
- Nur die originären, bereinigten Erhebungs- bzw. Urdaten sind archivwürdig, nicht aggregierte, also zusammengefasste Daten (z. B. auf Kreis- und Bezirksebene), da jede Aggregation die Auswertungsmöglichkeiten ganz wesentlich einschränkt. Bestimmte Hilfsmerkmale wie z. B. die Betriebsnummer müssen wegen der Verknüpfbarkeit zusammengehöriger Datensätze in den Daten erhalten bleiben. In aggregierten Dateien fehlen diese Hilfsmerkmale in aller Regel bereits.
- Daten sollen in einem festen Zyklus archiviert werden. Wie Klaus Döll bereits 1965 in seinem Gutachten¹⁹ ausführt, sind sogenannte „Datenester“ zu vermeiden. Ein einzelner Jahrgang ist in der Regel keine ausreichende Datenbasis für wissenschaftliche Auswertungen. Nur das Vorhandensein von

Statistiker am Alexanderplatz miteinbezogen, die mit den Daten seit vielen Jahren vertraut waren und ihre Entstehungsgeschichte teilweise über Jahrzehnte mitgestaltet haben.

A. Ullmann, Maschinenlesbare Daten in staatlichen Behörden und Archiven Sachsens, in: Brandenburgische Archive, 13/1999, S. 4-6,

dies.: Der Archivar im Datenschwungel: Intensive Zusammenarbeit zwischen dem Sächsischen Hauptstaatsarchiv Dresden und dem Statistischen Landesamt Kamenz, in: Sächsisches Archivblatt, H. 1/1999, S. 13-14 und H. 2/1999, S. 9-10 sowie

dies.: Die Bewertung der Unterlagen der Staatlichen Zentralverwaltung für Statistik beim Ministerrat der DDR, Bezirksstelle Dresden, in: Sächsisches Archivblatt, H. 2/2000, S. 13-15.

¹⁶ Dipl.-Archivar Dr. Christian Keitel, Referent bei der Landesarchivdirektion Stuttgart.

¹⁷ C. Keitel, „Die archivische Bewertung elektronischer Statistiken“, in: Karl-Ernst Lupprian; 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).

¹⁸ Keitel, Bewertung.

¹⁹ K. Döll, Die Aufbewahrung sozialwissenschaftlich wichtiger Massendaten. Empfehlungen für die Archivierung statistischen und anderen behördlichen Quellenmaterials unter sozialwissenschaftlichen Aspekten, Köln 1965.

Statistikdaten aus einer Reihe von Jahren ermöglicht Längsschnittanalysen. In unserem Bestand haben nur wenige Statistiken eine Laufzeit, die der Gesamtlaufzeit des Datenbestands (1976-1991) annähernd entspricht. Einige Statistiken wurden nicht übernommen, da nur Daten aus einem einzigen Jahr vorlagen.

- Eine der wichtigsten Voraussetzungen für die Archivwürdigkeit ist das Vorhandensein einer vollständigen und aussagekräftigen *Datendokumentation* zu einer Statistik. Diese umfasst mindestens: Datensatzbeschreibungen, Erhebungsformulare, Schlüsselverzeichnisse für codierte Felder, Begriffsdefinitionen, Ausfüllrichtlinien usw. Diese Unterlagen wurden vom Statistischen Bundesamt in den sogenannten „Archivierungsakten“ zusammengestellt. Lücken in diesen Archivierungsakten sind ein Kassationsgrund.

Einige *inhaltliche Überlegungen* spielten ebenfalls eine große Rolle:

- Eine pauschale Komplettarchivierung aller Statistikdaten schied allein schon wegen der großen Menge von Einzelstatistiken mit sehr unterschiedlicher Wertigkeit von vorneherein aus. Vom Inhalt der Fachstatistik her muss mindestens ein gewisser „historischer Wert“ gegeben sein. Ein Maßstab dafür kann z. B. der Umfang des Merkmalskatalogs der Erhebung sein.
- Aus den verschiedenen Gebieten der Statistik wurden die jeweils wichtigsten Einzelstatistiken als archivwürdig bewertet. Zugleich wurden Überlieferungsschwerpunkte gebildet: Als besonders wichtig angesehen wurden Daten zu Arbeitskräften, Bevölkerung, Kultur, Bildung und Wissenschaft, Industrie sowie zum Umweltschutz
- Auch Überlegungen der Bestandsergänzung spielten eine Rolle: Aus dem Bereich der Agrarstatistik wurden eine ganze Reihe von Daten übernommen, die künftig als Ergänzung des leider sehr unvollständigen Datenbestands des Ministeriums für Land-, Forst- und Nahrungsgüterwirtschaft²⁰ dienen werden. Weiterhin waren im Schriftgutbestand Staatliche Zentralverwaltung für Statistik²¹ gerade im Bereich der Wirtschaftsstatistik große Überlieferungslücken zu verzeichnen, so dass wir mit der Bewertung der Datenüberlieferung aus dem Bereich der Wirtschaftsstatistik großzügiger waren, um diese Lücken mit den Daten sinnvoll zu ergänzen.
- Der Bildung von Doppelüberlieferungen soll dagegen kein Vorschub geleistet werden, daher waren Daten aus einigen statistische Erhebungen, die in anderen Statistiken bereits enthalten sind, nicht archivwürdig.
- Statistiken, die in erster Linie zu Zwecken von Agitation und Propaganda erhoben wurden, sind mit Vorbehalt zu betrachten, z. B. war dies bei der Statistik der „Messe der Meister von morgen“ der Fall. Auswertungsmöglichkeiten dürfte es dafür kaum geben und der "historische Wert" dürfte nicht

²⁰ Bundesarchiv, Ministerium für Land- Forst- und Nahrungsgüterwirtschaft – maschinenlesbare Daten, DK 1 MD.

²¹ Bundesarchiv, Staatliche Zentralverwaltung für Statistik, DE 2.

unumstritten sein. Wir gehen davon aus, dass diese Art von Statistiken im Schriftgut ausreichend dokumentiert und überliefert ist.

Auch *technische Gesichtspunkte* sind bei der Bewertung zu berücksichtigen. Die Datenträger selbst (Magnetband-Kassette Typ IBM 3480/3490 compatible) waren unproblematisch. Es war dagegen vor allem die Frage zu stellen, wie die Datenformate aussehen und ob die Daten mit vertretbarem Aufwand in das Standardformat ASCII csv²² konvertierbar sind. Schwierig zu durchschauen ist die besondere Datenstruktur vieler Statistikdateien, die heute die Konvertierung der Daten erschwert und künftig weitere Programmierarbeiten erfordert. Das Problem ist aber mit vertretbarem Aufwand lösbar.

Das Bundesarchiv hat neben den Daten der statistischen Erhebungen sogenannte „sequentielle Auszüge“ der Jahre 1980 bis 1990 aus der *Datenbank SPAZ*²³ übernommen. Es handelt sich dabei nicht um eine Originalquelle, also um die Erhebungs- oder Urdaten selbst, sondern um eine Sekundär-Datenbank, die sich aus vollständigen Zeitreihen von Erhebungsdaten zusammensetzt. Nach Auskunft der Statistiker wurden alle wesentlichen Daten zu allen Betrieben aus den Bereichen Landwirtschaft, Bau und Industrie in dieser Datenbank zusammengestellt. SPAZ diente als Grundlage für die wirtschaftsleitende Tätigkeit der Partei (Frage der Planerfüllung usw.). Nach Angaben der Statistiker ist der Wahrheitsgehalt der Daten unbestritten und die Inhalte sind mit den Urdaten identisch. Die Datenbanklösung wurde Anfang der 1970er Jahre entwickelt, zeitlich in etwa parallel mit der Datenbank STATIS-BUND in Wiesbaden, und ist vermutlich auch technisch vergleichbar.²⁴ Wie Christian Keitel auf der letzten Arbeitskreissitzung in München berichtete²⁵, existiert beim Statistischen Landesamt in Stuttgart die SRDB (Struktur- und Regionaldatenbank). Die dort gespeicherten Informationen stammen zumeist ebenfalls aus den Fachstatistiken. Offensichtlich handelt es sich auch bei der SRDB in Stuttgart um einen vergleichbaren Fall wie bei der SPAZ-Datenbank. Bei der Bewertung der Daten des Statistischen Landesamts Stuttgart durch die Landesarchivdirektion ist für die Zukunft entschieden worden, die SRDB zu archivieren. Im Bundesarchiv sollten ursprünglich nur Fachstatistiken archiviert werden; schließlich wurde die SPAZ-Datenbank zusätzlich übernommen, nicht zuletzt auch deswegen, weil einige Zeitreihen in SPAZ vollständiger sind als in den Fachstatistiken selbst. Ich sehe die Datenbank SPAZ daher als eine sinnvolle und wichtige Ergänzung zu den Einzelstatistiken an.

Eine *Nachbewertung* wird bei einigen Statistiken in Zukunft noch erforderlich sein. Die Dokumentationsunterlagen, vor allem die sogenannten „Archivierungsakten“, konnten noch nicht im Detail auf ihre Vollständigkeit überprüft werden. Es werden ggf. Recherchen im Schriftgutbestand in unserer

²² ASCII (American Standard Code for Information Interchange) csv (comma separated values).

²³ Datenbank SPAZ („Speichern und Auffinden von Zeitreihen“).

²⁴ M. Güttler, Die Datenverarbeitung im statistischen Informationssystem der DDR. Ein Beitrag zur deutschen Vereinigung und zur Geschichte der Informatik, Berlin ²1992 (Technische Universität Berlin: Forschungsberichte des Fachbereichs Informatik, Bericht 1990-36), S. 49-54.

²⁵ Keitel, Bewertung.

Außenstelle Dahlwitz-Hoppegarten bei Berlin durchgeführt werden müssen, in der Hoffnung, nach dem Auffinden weiterer Unterlagen Lücken schließen zu können. Wo endgültig Dokumentationsunterlagen in ausreichendem Umfang nicht zu beschaffen sein werden, werden Statistiken nachbewertet und ggf. kassiert werden müssen.

4. Die Technische Bearbeitung (Kopierung, Konvertierung, Anonymisierung)²⁶

Zur technischen Bearbeitung der Daten gehören im wesentlichen drei Schritte – die Sicherungskopierung, die Konvertierung in ein Standardformat und ggf. die Generierung einer anonymisierten Version für Benutzer.

Die Statistikdaten wurden auf Magnetbandkassetten ins Bundesarchiv übernommen. Die Sicherungskopierung aber erfolgt im Bundesarchiv auf anderen Datenträgern, und zwar zweifach – zum einen auf CD-R²⁷, zum anderen auf Digital Audio Tape. Nachdem zu Beginn der Kopierarbeiten vereinzelt Fehler auf den Datenträgern aufgetreten waren, hat das Statistische Bundesamt Wiesbaden zusätzlich den Duplikatbestand der Magnetbandkassetten mit DDR-Statistikdaten übergeben, so dass jetzt jede Datei doppelt vorliegt. Seither konnten die Kopierarbeiten ohne Beanstandungen weitergeführt werden. Die Statistikdaten aus 38 von 94 Erhebungen (136 von 235 Kassetten) sind inzwischen kopiert, ebenso alle Daten aus der Gesundheitsstatistik, die dem Bestand Institut für Medizinische Statistik und Datenverarbeitung zugeordnet wurden. Die Kopierung soll in der ersten Jahreshälfte 2003 abgeschlossen werden.

Die Daten liegen grundsätzlich im Format EBCDIC²⁸ vor, da sie auf IBM- bzw. ESER-Großrechnern²⁹ erstellt und gespeichert worden sind. Sie enthalten aber teilweise in einzelnen Datenfeldern binär komprimierte Zahlen, die nicht dem EBCDIC-Format entsprechen. Etwa 70% der Dateien sind überdies in einem speziellen Lochkartenformat gespeichert. Die Datenerfassung erfolgte in den 1970er Jahren über Lochkarten, in den 80ern über Datensammelsysteme; die Datenstrukturen wurden aber meistens nicht angepasst.³⁰ Die beispielsweise zu einem Betrieb gespeicherten Informationen sind oftmals in mehreren 80stelligen Datensätzen mit unterschiedlichem Satzaufbau innerhalb einer Datei abgelegt. Eine einzige Datei besteht manchmal aus Datensätzen mit 20 oder 30 verschiedenen Strukturen, die nur durch die in den Daten abgespeicherte Lochkartennummer unterschieden werden können. Dies führt dazu, dass die Daten nicht mit einem handelsüblichen Konverter in den Standard-Code ASCII csv migriert werden können, sondern es

²⁶ U. Rathje, Technisches Konzept für die Datenarchivierung im Bundesarchiv, in: Der Archivar, H. 2/2002, S. 117-120.

²⁷ CD-R (Compact Disk-Recordable, eine einmalig beschreibbare CD).

²⁸ EBCDIC (Extended Binary-Coded Decimal Interchange Code, in den 1960er Jahren entwickelter Normcode für Großrechner, auch als IBM-Format oder IBM-Code bekannt).

²⁹ ESER (Einheitliches System der Elektronischen Rechentechnik der sozialistischen Länder). Es handelte sich um Großrechner (Modelle ESER EC 1040 und EC 1055), die mit den IBM-Systemen 360 und 370 kompatibel waren. Die verschiedenen Komponenten wurden in Bulgarien, der DDR, in Polen, der Sowjetunion und in der Tschechoslowakei produziert.

³⁰ Praschek, Erschließung, S. 536.

müssen spezielle Konvertierungsmodule geschrieben werden. Für diese Programmierarbeiten vergibt das Bundesarchiv Werkverträge im Rahmen des Outsourcing außer Haus.

Das Bundesarchiv ist durch das Bundesarchivgesetz³¹ dazu verpflichtet, bei Unterlagen mit personenbezogenen Inhalten die Vorschriften über die Verarbeitung und Sicherung dieser Unterlagen zu beachten, die für die abgebende Stelle gelten. Die Benutzung ist daher an Bedingungen nach dem Bundesstatistikgesetz (BStatG)³² geknüpft. Nach BStatG kann für wissenschaftliche Forschungsvorhaben eine Benutzungsgenehmigung erteilt werden, wenn bestimmte Voraussetzungen erfüllt sind. Insbesondere dürfen „...Einzelangaben nur übermittelt werden, wenn ... [sie] nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft zugeordnet werden können...“.³³ Daher generiert das Bundesarchiv vor einer wissenschaftlichen Benutzung eine anonymisierte Version³⁴ der Statistikdaten in Abstimmung mit dem Statistischen Bundesamt. Die an dem Forschungsprojekt beteiligten Personen werden vor der Übermittlung der Benutzerkopie der Daten zur Geheimhaltung verpflichtet. Durch organisatorische und technische Maßnahmen ist sicherzustellen, dass nur autorisierte Personen Zugang zu den Daten bekommen. Die Daten müssen sicher verschlossen gelagert werden, so dass ein Missbrauch und die Weitergabe der Daten ausgeschlossen ist. Das Bundesarchiv muss nach Abschluss des Forschungsprojekts auf einer Rückgabe bzw. Löschung der Daten bestehen.

Aus Gründen der Wirtschaftlichkeit entsteht bei den Programmierarbeiten ein einziges Modul für die Konvertierung und die Generierung der anonymisierten Version, so dass diese Arbeiten im Bundesarchiv in einem Arbeitsgang durchgeführt werden können. Im Ergebnis entstehen aus einer Originaldatei zwei neue Dateien: eine im Format ASCII csv, die inhaltlich exakt der Originaldatei entspricht, und eine zweite ASCII-csv-Datei, die eine anonymisierte Version der ersten Datei darstellt.

5. Benutzungen

Die wissenschaftliche Benutzung von DDR-Statistikdaten hat bereits im Herbst 2000 begonnen und es gehen im Bundesarchiv immer wieder Anfragen nach den Statistikdaten der DDR ein.

³¹ 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).

³² Bundesstatistikgesetz (BStatG). Gesetz über die Statistik für Bundeszwecke (BStatG) vom 22. Januar 1987 (BGBl. I, S.462, 565), zuletzt geändert durch Artikel 4 Strafverfahrensänderungsgesetz 1999 (StVÄG 1999) vom 02.08.2000 (BGBl. I, S. 1261).

³³ W. Buchmann, und M. Wettengel, Auslegung des Bundesstatistikgesetzes bei der Archivierung von Statistikerunterlagen, in: Der Archivar, H.1/1996, Sp. 67-74.

³⁴ Zur näheren Einführung in das Thema faktische Anonymisierung insbesondere: H. Wirth, Die faktische Anonymität von Mikrodaten. Ergebnisse und Konsequenzen eines Forschungsprojektes, in: ZUMA-Nachrichten Nr. 30/1992, S. 7-45,

W. Müller/U. Blien/P. Knoche/H. Wirth, Die faktische Anonymität von Mikrodaten, Statistisches Bundesamt, Stuttgart 1991, (Schriftenreihe Forum der Bundesstatistik, 19),

W. Müller/U. Blien/H. Wirth, Identification Risks of Microdata. Evidence from experimental studies. Sociological Methods & Research 24/1995, S. 131-157 sowie

- Im Rahmen des Forschungsprojekts „Soziokulturelle Unterschiede der Sterblichkeit in Deutschland unter besonderer Berücksichtigung der Todesursachenstatistik“ wertet das Max-Planck-Institut für Demografische Forschung in Rostock³⁵ die Daten der Todesursachenstatistik 1975-1990³⁶ aus.
- Das Institut für Soziologie an der Universität München³⁷ hat im Rahmen des Forschungsprojekts „Geschlechtersegregation in Organisationen“³⁸ mit der wissenschaftlichen Auswertung des Datenspeichers Gesellschaftliches Arbeitsvermögen des Staatssekretariats für Arbeit und Löhne³⁹ begonnen. In den nächsten Monaten wird sich daran eine Benutzung von Statistikdaten anschließen: Gegenstand der Benutzung ist die Einkommensstichprobe 1988.⁴⁰
- Die Einkommensstichprobe soll weiterhin in Kürze für eine Dissertation am Department of Economics an der Universität von Chicago ausgewertet werden. Das Thema der Benutzung lautet: When East meets West: Labor Market Adjustment during the East German Transition.
- Am Institut für Soziologie der Universität Würzburg⁴¹ wird in Kürze im Rahmen des DFG-Projekts „Familiale Arbeitsteilung in den Ländern der Europäischen Union: Länderstudie Deutschland“⁴² eine wissenschaftliche Untersuchung der Zeitbudgeterhebungen 1985 und 1990⁴³ stattfinden.
- Weiterhin finden derzeit erste Vorgespräche mit dem Zentrum für Umfragen, Methoden und Analysen in Mannheim (ZUMA)⁴⁴ statt, in denen die Möglichkeiten einer Aufbereitung der Daten aus den Volkszählungen 1971 und 1981⁴⁵ durch das ZUMA und eine Kooperation mit dem Bundesarchiv verhandelt werden.

6. Ausblick

W. Müller/H. Wirth, Research Needs for European Microdata and Data Confidentiality. International Seminar on Statistical Confidentiality. Proceedings. 28 to 30 November 1994, Eurostat OE, Luxembourg 1995, S. 235-240.

³⁵ http://www.demogr.mpg.de/d_default.htm.

³⁶ Bundesarchiv, Staatliche Zentralverwaltung für Statistik – maschinenlesbare Daten, DE 2 MD/8 (Gestorbene Säuglinge und Totgeborene 1979-1990) und DE 2 MD/9 (Gestorbene 1980-1990).

³⁷ <http://www.soziologie.uni-muenchen.de/>.

³⁸ <http://www.lrz-muenchen.de/~Segregation/>

³⁹ Bundesarchiv, Staatssekretariat für Arbeit und Löhne – maschinenlesbare Daten, DQ 3 MD.

F. Dietz/H. 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,

U. Rathje, Datenspeicher Gesellschaftliches Arbeitsvermögen der DDR – ein „Werkstattbericht“, in: Mitteilungen aus dem Bundesarchiv, H. 3/1994, S. 157-159 sowie

ders.: Der „Datenspeicher Gesellschaftliches Arbeitsvermögen“ der DDR. In: Historical Social Research / Historische Sozialforschung (HSR) No. 2/1996, S. 113-118.

⁴⁰ Bundesarchiv, Staatliche Zentralverwaltung für Statistik – maschinenlesbare Daten, DE 2 MD/38 (Einkommensstichprobe in Arbeiter- und Angestelltenhaushalten 1988).

⁴¹ <http://www.uni-wuerzburg.de/soziologie/index.html>.

⁴² <http://www.uni-wuerzburg.de/soziologie/projekt.html>.

⁴³ Bundesarchiv, Staatliche Zentralverwaltung für Statistik – maschinenlesbare Daten, DE 2 MD/37 (Zeitbudgeterhebungen 1985 und 1990).

⁴⁴ Zentrum für Umfragen, Methoden und Analysen (ZUMA) in Mannheim, siehe <http://www.gesis.org/zuma/>.

⁴⁵ Bundesarchiv, Staatliche Zentralverwaltung für Statistik – maschinenlesbare Daten, DE 2 MD/27 (Volkszählungen 1971 und 1981).

Das Bundesarchiv hat mit den Statistikdaten der DDR nicht nur einen umfangreichen Datenbestand übernommen, der von bedeutendem historischem Wert ist, da er die wesentlichen Bereiche der amtlichen Statistik der DDR abbildet, sondern auch eine Verpflichtung, diese Daten sachgerecht zu archivieren und für die Wissenschaft benutzbar zu machen. Die Statistikdaten der DDR sind für die wissenschaftliche Forschung von großem Interesse. Das Bundesarchiv versucht, im Rahmen seiner begrenzten personellen Möglichkeiten den Wünschen der Benutzer weitmöglichst nachzukommen und diese Verpflichtungen zu erfüllen.

7. Literatur

- Angermann, Oswald: Sammlung, Sicherung und Rückrechnung von statistischen Angaben über die ehemalige DDR, in: Allgemeines Statistisches Archiv, 78/1994, S. 340-349
- Barke, Werner (Statistisches Landesamt Baden-Württemberg): Ich glaube nur der Statistik... Was Winston Churchill über Zahlen und die Statistik wirklich sagte und was er gesagt haben soll. Stuttgart 2001
- Buchmann, Wolf und Michael Wettengel, Auslegung des Bundesstatistikgesetzes bei der Archivierung von Statistikerunterlagen, in: Der Archivar, H.1/1996, Sp. 67-74
- Dietz, Frido/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
- Döll, Klaus, Die Aufbewahrung sozialwissenschaftlich wichtiger Massendaten. Empfehlungen für die Archivierung statistischen und anderen behördlichen Quellenmaterials unter sozialwissenschaftlichen Aspekten, Köln 1965
- Güttler, Markus, Die Datenverarbeitung im statistischen Informationssystem der DDR. Ein Beitrag zur deutschen Vereinigung und zur Geschichte der Informatik, Berlin ²1992 (Technische Universität Berlin: Forschungsberichte des Fachbereichs Informatik, Bericht 1990-36)
- Keitel, Christian, Die archivische Bewertung elektronischer Statistiken, in: Karl-Ernst Lupprian; 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)
- Krämer, Walter, Denis Krämer und Götz Trenkler, Das neue Lexikon der populären Irrtümer. 555 weitere Vorurteile, Missverständnisse und Denkfehler von Advent bis Zyniker, Frankfurt am Main 1998
- Müller, W./Blien, U./Knoche, P./Wirth, H., Die faktische Anonymität von Mikrodaten, Stuttgart 1991 (Schriftenreihe Forum der Bundesstatistik, 19)
- Müller, W./Blien, U./Wirth, H., Identification Risks of Microdata. Evidence from experimental studies. Sociological Methods & Research 24 (1995), S. 131-157
- Müller, W./Wirth, H., Research Needs for European Microdata and Data Confidentiality. International Seminar on Statistical Confidentiality. Proceedings. 28 to 30 November 1994, Luxembourg 1995. (Eurostat 0E: 235-240)

- Oertel, Ingrid, 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
- Praschek, Helmut, Erschließung und Verbleib der Einzeldatenbestände der amtlichen Statistik der ehemaligen DDR, in: Wirtschaft und Statistik, 7/2000, S. 533-540
- Praschek, Helmut, Besonderheiten der amtlichen Statistik in der ehemaligen DDR, in: Wirtschaft und Statistik, 1/2000, S. 24-29
- Pressedienst derLandeshauptstadt München, Gute Verwaltung braucht gute Statistiker – 125 Jahre Statistisches Amt, in: Haidhauser Nachrichten (Dokumentation), 12.01.2001, <http://hn.munich-info.de/muc/2000/01/ru-2742.html>
- Rathje, Ulf, Datenspeicher Gesellschaftliches Arbeitsvermögen der DDR – ein „Werkstattbericht“, in: Mitteilungen aus dem Bundesarchiv H. 3/1994, S. 157-159
- Rathje, Ulf, Der „Datenspeicher Gesellschaftliches Arbeitsvermögen“ der DDR, in: Historical Social Research / Historische Sozialforschung (HSR), No. 2/1996, S. 113-118
- Rathje, Ulf, Technisches Konzept für die Datenarchivierung im Bundesarchiv, in: Der Archivar H. 2/2002, S. 117-120
- Sonderreihe mit Beiträgen für das Gebiet der ehemaligen DDR. Arbeitsunterlage, Bd. 1-34. Statistisches Bundesamt, Wiesbaden 1993-2000
- Ullmann, Angela, Maschinenlesbare Daten in staatlichen Behörden und Archiven Sachsens. Brandenburgische Archive, 13/1999, S. 4-6
- Ullmann, Angela, Der Archivar im Datenschlingel: Intensive Zusammenarbeit zwischen dem Sächsischen Hauptstaatsarchiv Dresden und dem Statistischen Landesamt Kamenz. In: Sächsisches Archivblatt, H. 1/1999, S. 13-14 und H. 2/1999, S. 9-10
- Ullmann, Angela, Die Bewertung der Unterlagen der Staatlichen Zentralverwaltung für Statistik beim Ministerrat der DDR, Bezirksstelle Dresden, in: Sächsisches Archivblatt, 2/2000, S. 13-15
- Wirth, Heike, Die faktische Anonymität von Mikrodaten. Ergebnisse und Konsequenzen eines Forschungsprojektes, in: ZUMA-Nachrichten Nr. 30/1992, S. 7-45