

**Burkhard Reiß,
Von EBCDIC nach XML: Das neue Konvertierungsprogramm
des Bundesarchivs zur Migration von Altdaten**

aus:

Digitales Verwalten – Digitales Archivieren

Veröffentlichungen aus dem Staatsarchiv der Freien und Hansestadt
Hamburg, Band 19

Herausgegeben von Rainer Hering und
Udo Schäfer

S. 83-86

Impressum für die Gesamtausgabe

Bibliografische Information Der Deutschen Bibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.ddb.de> abrufbar.

Diese Publikation ist außerdem auf der Website des Verlags Hamburg University Press *open access* verfügbar unter <http://hup.rrz.uni-hamburg.de>.

Die Deutsche Bibliothek hat die Netzpublikation archiviert. Diese ist dauerhaft auf dem Archivserver Der Deutschen Bibliothek verfügbar unter <http://deposit.ddb.de>.

ISBN 3-937816-09-7 (Printausgabe)

ISSN 0436-6638 (Printausgabe)

© 2004 Hamburg University Press, Hamburg

<http://hup.rrz.uni-hamburg.de>

Rechtsträger: Universität Hamburg

Inhalt

Vorwort	9
 Digitale Signatur – Authentizität und Langzeitarchivierung	
Authentizität: Elektronische Signaturen oder Ius Archivi?	13
<i>Udo Schäfer</i>	
Elektronisch signierte Dokumente	33
Anforderungen und Maßnahmen für ihren dauerhaften Erhalt	
<i>Stefanie Fischer-Dieskau</i>	
Vom Posteingang bis in das Archiv	51
Technische und organisatorische Konzepte des ArchiSig-Projekts	
<i>Wolfgang Farnbacher</i>	
Digitale Signatur in der Praxis	67
Elektronischer Rechtsverkehr am Finanzgericht Hamburg	
<i>Jutta Drühmel</i>	
 Berichte und Informationen aus der Praxis	
Erste Erfahrungen mit der Langzeitarchivierung von Datenbanken	71
Ein Werkstattbericht	
<i>Christian Keitel</i>	
Von EBCDIC nach XML: Das neue Konvertierungsprogramm	
des Bundesarchivs zur Migration von Altdaten	83
<i>Burkhard Reiß</i>	
E-Government um jeden Preis?	87
Aktuelle Vorhaben zur Einführung der IT-gestützten Vorgangsbearbeitung und der digitalen Signatur im Freistaat Sachsen	
<i>Andrea Wettmann</i>	

Standardisierung und archivische Bewertung von elektronischen
Geschäftsverwaltungssystemen (GEVER) 95

Werkstattbericht aus dem Schweizerischen Bundesarchiv
Thomas Zürcher Thrier

Elektronische Vorgangsbearbeitung in der Landesverwaltung
Mecklenburg-Vorpommern 105

Entwicklung, Stand, Probleme, Perspektiven
Matthias Manke

Digitale Daten im Unternehmensarchiv in der Historischen
Kommunikation der Volkswagen AG 123
Ulrike Gutzmann

Das System Digitaler Bilderdienst / Bildarchiv
beim Deutschen Bundestag 131
Angela Ullmann

Dokumentenmanagementsysteme (DMS) zwischen Verwaltung und Archiv

Die elektronische Dokumentenverwaltung für Hamburg 143
Heinz Vogel

Dem Informellen einen Rahmen geben 153
Die Einführung des digitalen Dokumentenmanagements unter
besonderer Berücksichtigung der Kategorie des Informellen
in Veränderungsprozessen
Ivy Gumprecht

Change Management und Archive 167
Archivische Aufgaben im Rahmen der Implementierung
von Dokumentenmanagementsystemen
Rainer Hering

Zur Rolle der Archive bei der Erstellung eines Anforderungskatalogs
für ein Dokumentenmanagementsystem 183
Ein Werkstattbericht
Margit Ksoll-Marcon

Dokumentenmanagement bei der Stadtverwaltung Schwabach	191
<i>Wolfgang Dippert</i>	
DMS-Einführung in einer Kommunalverwaltung: Archivische Beteiligung und Erfahrungen	201
<i>Christoph Popp</i>	
Autorinnen- und Autorenverzeichnis	211
Teilnehmende	215

Von EBCDIC nach XML: Das neue Konvertierungsprogramm des Bundesarchivs zur Migration von Altdaten

Burkhard Reiß

Im Rahmen der Sicherung und Verfügbarmachung von DDR-Daten läuft seit dem Jahr 2002 ein gemeinsames Projekt des Bundesarchivs mit dem Institut für Sozialwissenschaftliche Informatik der Universität Koblenz-Landau, vertreten durch Herrn Prof. Dr. Klaus G. Troitzsch.

Ziel war es, anhand von zwei Diplomarbeiten ein Werkzeug zu entwickeln, mit dessen Hilfe eine Vielzahl der im Bundesarchiv archivierten DDR-Datenbestände in ein oder mehrere Formate konvertiert sowie Anonymisierungen vorgenommen werden können. Die Daten liegen hierbei noch in Formaten der Großrechnerwelt der 1970er und 1980er Jahre vor, in der Regel im EBCDI-Code. Hinzu kommen zusätzliche Komprimierungen, die eine routinemäßige, automatisierte Umsetzung in ein mit modernen Datenbankanwendungen zu verarbeitendes Format unmöglich machen.

Im Vordergrund der Entwicklung standen zwei Zielsetzungen:

Einerseits kam es darauf an, ein einheitliches Format zu erstellen, das für die Langzeitarchivierung geeignet ist und einen mit möglichst geringen Anforderungen verbundenen Zugang zu den Datenbeständen gewährleistet. Hierzu muss das Werkzeug in der Lage sein, die verschiedenen, teilweise in unterschiedlichen Varianten vorliegenden Datenformate innerhalb der DDR-Daten verarbeiten zu können.

Andererseits stellt der große Anteil an personenbezogenen Daten in der digitalen DDR-Überlieferung hohe Anforderungen an die Möglichkeiten des Werkzeugs im Hinblick auf Anonymisierungen unterschiedlichster Art. Um Benutzung überhaupt zu ermöglichen, ist die Anonymisierung vieler Datenbestände unumgänglich.

Als erster Schritt in der Entwicklung des Konvertierungswerkzeugs wurden verschiedene Datenbestände zur Bearbeitung ausgewählt. Sie sollten am Ende des Projekts in einem Format zur Langzeitarchivierung sowie, wo notwendig, in einer anonymisierten Version für die Benutzung vorliegen. Zugleich dienten sie als Anschauungsmaterial für die Entwickler, auf dessen Grundlage ein möglichst umfassendes, vielseitiges Konvertierungswerkzeug erstellt werden sollte. Die ausgewählten Datenbestände mussten somit eine möglichst breite Palette der vorkommenden Datensatzstrukturen abbilden. Ein weiteres Auswahlkriterium war die Bedeutung, die diesen Beständen für die Überlieferung der DDR zukommen dürfte.

Ausgewählt wurden im Einzelnen folgende Datenbestände:

- Ministerrat der DDR Projekt Eingaben der Bürger (Bundesarchiv-Signatur DC 20 MD/02):
In diesem Projekt wurden alle beim Ministerrat der DDR eingehenden Eingaben der Bürger von Juli 1988 bis Oktober 1990 erfasst. Die Daten dienten zur Registrierung und statistischen Auswertung der Eingaben sowie als Grundlage für die Berichterstattung an den Staatsrat. Insgesamt wurden im genannten Zeitraum etwa 60 000 Eingaben erfasst.
- Staatliche Zentralverwaltung für Statistik Volkszählungsdaten (Bundesarchiv-Signatur DE 2 MD/27):
Die Volkszählungsdaten dienten der stichtagsbezogenen Ermittlung der wichtigsten demografischen, sozialen und ökonomischen Merkmale von Einwohnern und Haushalten. In elektronischer Form liegen die Erhebungsdaten der letzten beiden Zählungen von 1971 und 1981 vor. Bei den Daten handelt es sich um Einzeldatensätze je Person und Wohnung. So liegen für das Jahr 1971 Daten zu 16,4 Millionen Personen und sechs Millionen Wohnungen vor.
- Ministerium für Hoch- und Fachschulwesen Projekt PWE Personalwesen der Hochschulen (Bundesarchiv-Signatur DR 2 MD/01):
Das Projekt enthält die Personalstammdaten des wissenschaftlichen und des Leitungspersonals an den Hochschulen der DDR. Erfasst wurden detaillierte personenbezogene Angaben aus den Jahren 1986 bis 1990.
- Ministerium des Innern Projekte NRA, NRB und NRC (Bundesarchiv-Signaturen DO 1 MD/01-03):

In allen Strafvollzugseinrichtungen, Jugendhäusern und Untersuchungsanstalten der DDR wurden monatlich statistische Kennziffern erhoben, getrennt nach Geschlecht sowie nach Straf- und Vollzugsart. Erfasst wurden neben statistischen Angaben unter anderem die Delikte der Inhaftierten sowie personenbezogene Daten.

- Ministerium des Innern Projekte NIE und NIA Fahndungs- und Kriminalstatistik (Bundesarchiv-Signatur DO 1 ND/04 und 12):

Die Fahndungsstatistik der Hauptabteilung Kriminalpolizei des Ministeriums des Innern enthält ausführliche Angaben zu Personen- und Sachfahndungen aus den Jahren 1975 bis 1990. Eine Jahresdatei enthält Angaben zu 28 000 bis 50 000 Fahndungen.

Die Hauptabteilung Kriminalpolizei des Ministeriums des Innern war ebenfalls für die Kriminalstatistik Teil I zuständig. Sie enthält unter anderem Angaben zu Anzeigen und abgeschlossenen Ermittlungsverfahren. Die Kriminalstatistik Teil II lag im Zuständigkeitsbereich des Generalstaatsanwalts der DDR. Aus diesem Bereich liegen Daten mit so genannten täterbezogenen Informationen vor.

Das aufgrund dieser Datenbestände entwickelte Konvertierungswerkzeug baut sich aus zwei Teilen auf, die im Rahmen der von Prof. Dr. Troitzsch betreuten Diplomarbeiten von Oliver Hartmann und Stephan Eichenlaub entwickelt wurden.¹

Im ersten Teil, dem so genannten Front-End, werden die Daten aus den diversen vorliegenden Formaten in ein standardisiertes, gemeinsames Zwischenformat migriert. Dieses Zwischenformat hat zweierlei Funktion. Einerseits ist es Ausgangspunkt für die Weiterverarbeitung der Daten im zweiten Teil des Konverters, dem so genannten Back-End, das im Anschluss näher erläutert wird. Andererseits findet das Zwischenformat als Langzeitspeicherformat beziehungsweise Archivierungsformat Verwendung. Das Zwischenformat muss somit mehrere Kriterien erfüllen: Es darf

¹ BADDR. Ein Migrationstool-Generator für das Bundesarchiv. Migration und Anonymisierung von Datenbeständen aus der DDR in moderne Datenbank- und Statistikformate. Oliver Hartmann: Teil I: Front-End; Stephan Eichenlaub: Teil II: Back-End. Diplomarbeiten an der Universität Koblenz 2004.

kein proprietäres Format sein, das heißt seine mögliche Weiterverarbeitung darf nicht an bestimmte Programme gebunden sein. Deshalb kam nur ein reines, softwareunabhängiges Textformat in Frage. Weiterhin muss das Zwischenformat die originalen Daten selbstverständlich angemessen wiedergeben. Die Authentizität und Integrität der Daten an sich muss ständig gewährleistet sein. Dies ist sowohl im Hinblick auf die Verwendung des Zwischenformats als Archivierungsformat, als auch für die Weiterverarbeitung der Daten im Back-End eine unabdingbar notwendige Voraussetzung.

Der zweite Teil des Konvertierungswerkzeugs, das so genannte Back-End, übernimmt, wo erforderlich, Anonymisierungen und ermöglicht die Ausgabe der Daten in XML, SQL und in ein Datenformat mit fester Satzlänge für die Weiterverarbeitung mit Statistikprogrammen. Die Weiterverarbeitung im Back-End bietet den großen Vorteil der Verknüpfung der Daten mit den ihnen zugrunde liegenden Codierungen. In der Regel sind in den Ursprungsdaten alle Informationen aus Platzgründen lediglich mit einem Zahlenschlüssel erfasst, der in gesonderten Tabellen, den so genannten Codelisten, festgelegt ist. Die Auflösung der jeweiligen Codierung konnte bisher nur durch Nachschlagen in der betreffenden Codeliste erfolgen. Eine Möglichkeit der automatischen Verknüpfung bestand nicht. Durch die nun mögliche Verknüpfung werden die Daten auch für den technisch weniger versierten Benutzer leichter interpretierbar. Es besteht somit nun für einen größeren Kreis von Benutzern die Möglichkeit, diesen wichtigen Teil der DDR-Überlieferung in ihren Forschungsarbeiten zu berücksichtigen. Die Ausgabe der Daten in XML wiederum wird es dem Bundesarchiv ermöglichen, weitere praktische Erfahrungen im Umgang mit diesem Format zu sammeln. Dies ist auch im Hinblick auf die zukünftige Übernahme elektronischer Akten von nicht zu unterschätzender Bedeutung.