

# Übernahme von Daten aus Fachanwendungen – Schnittstellen, Erhaltungsformen, Nutzung

Kai Naumann

Dieser Beitrag soll Fragen aus der Praxis beantworten und vor allem zur selbständigen Archivierung von Daten anregen. Im ersten Teil wird eine Übersicht über die bisher beim Landesarchiv Baden-Württemberg übernommenen Daten gegeben und die grobe Klassifikation in technischer Hinsicht dargestellt. Der zweite Teil zeigt die einzelnen Schritte, die das Landesarchiv bei der Archivierung von Daten aus einer Fachanwendung vollzieht. Der dritte Teil referiert die noch recht dünnen Erfahrungen mit der Erhaltung und Nutzung von Fachanwendungen.

In den neunziger und den frühen Nuller Jahren haben sich die deutschen Archivarinnen und Archivare vor allem mit einer einzelnen Form elektronischer Unterlagen befasst: den Dokumenten-Management-Systemen und den notwendigen Weichenstellungen bei deren Einführung in bestimmten Verwaltungszweigen. Nackte textuelle Daten in Form von Tabellen waren aus dieser Perspektive ein bloßes Hilfsmittel zum Erreichen höherer Ziele, nämlich der Darstellung der elektronischen Akte mit Vermerken und Verfügungen als komplexes Digitales Objekt. Eher beiläufig erschienen die Berichte von Michael Wettengel und Ulf Rathje über die Sicherung von Datenbeständen aus den Rechenzentren der untergegangenen DDR.<sup>1</sup> Man hätte nun die entsprechende Überlieferung der Bundesrepublik und ihrer Länder bewerten und übernehmen können – aber internationale Bemühungen, die auch in deutschen Kreisen bekannt wurden, wurden selten aufgegriffen.<sup>2</sup>

In den letzten Jahren fand hier in der allgemeinen Diskussion zu Bewertung und Übernahme,<sup>3</sup> aber auch in den Bereichen statistische Mikrodaten<sup>4</sup> und Geodaten<sup>5</sup>

---

<sup>1</sup> Ulf RATHJE, Die personenbezogenen Massendatenspeicher der DDR im Bundesarchiv. In: Potsdamer Bulletin für zeithistorische Studien 19 (1997), S. 40–54; Michael WETTENGEL, Archivierung digitaler Datenbestände aus der DDR nach der Wiedervereinigung. In: Udo SCHÄFER (Hrsg.), Archivierung elektronischer Unterlagen, Stuttgart 1999, S. 223–239; Ulf RATHJE, Daten der Staatlichen Zentralverwaltung für Statistik der DDR im Bundesarchiv – Bewertung, Übernahme, technische Bearbeitung und Benutzung. In: Mitteilungen aus dem Bundesarchiv 11 (2003) H. 3, S. 56–61; Ulf RATHJE u.a., Der Zentrale Kaderdatenspeicher des Ministerrates der DDR: zur Erschließung und Rekonstruktion einer bedeutenden zeithistorischen Quelle durch das Bundesarchiv und die Universität Jena. In: Mitteilungen aus dem Bundesarchiv 12 (2004) H. 1, S. 14–25.

<sup>2</sup> Thomas ZÜRCHER THRIER, Sicherung und Bewertung von Datenbanken im Schweizerischen Bundesarchiv. In: Arbido 18 (2003) H. 3, S. 19–21; René QUILLET, Archivierung einer Datenbank am Beispiel der Jugendanwaltschaft Basel. In: Arbido 18 (2003) H. 3, S. 22–23; Barbara HOEN, Workshop „Long-term Preservation of Databases“ im Schweizerischen Bundesarchiv. In: Der Archivar 56 (2003) H. 4, S. 347–348. Die Ausnahmen waren Christian KEITEL, Erste Erfahrungen mit der Langzeitarchivierung von Datenbanken: ein Werkstattbericht. In: Rainer HERING (Hrsg.), Digitales Verwalten – digitales Archivieren: 8. Tagung des Arbeitskreises „Archivierung von Unterlagen aus Digitalen Systemen“ am 27. und 28. April 2004 im Staatsarchiv Hamburg, Hamburg 2004, S. 71–81. Der Aufsatz ist auch erhältlich unter: <http://www.landearchiv-bw.de/web/46914>.

<sup>3</sup> Jürgen TREFFEISEN, Komplementäre Bewertung konventioneller Akten und elektronischer Daten. In: Archive im digitalen Zeitalter. Überlieferung – Erschließung – Präsentation. 79. Deutscher Ar-

ein Umdenken statt. Ganz dialektisch – These, Antithese, Synthese – entwickelt sich ein neues Paradigma für den Umgang mit massenhaft gleichförmigen Fällen. Altbekannt die in den sechziger Jahren formulierte Forderung, man müsse massenhaft gleichförmige Unterlagen überliefern – die These.<sup>6</sup> Danach die Ernüchterung, man könne mit massenhaften Unterlagen, etwa Personalakten, zu wenig<sup>7</sup> anfangen, weil die Auswertung, die massenhafte Erfassung, so kostenträchtig sei – das war die Antithese.

Schließlich dämmert aber die Synthese herauf, nämlich die Erkenntnis, dass massenhaft gleichförmige Angaben in Archiven durchaus ihre Berechtigung haben, aber besser in Form von Datentabellen als in Form von Akten überliefert werden sollten.<sup>8</sup>

## I.

Eine Übersicht der im Landesarchiv Baden-Württemberg bislang übernommenen elektronischen Unterlagen<sup>9</sup> zeigt, dass der überwiegende Teil der bisher übernommenen Bestände aus Datentabellen besteht. Ein großer Teil dieser Datentabellen

---

chivtag in Regensburg (Tagungsdokumentation zum Deutschen Archivtag 14), Neustadt a.d. Aisch 2010, S. 193–200; Albrecht ERNST, Christian KEITEL, Elke KOCH, Clemens REHM, Jürgen TREFFEISEN, Überlieferungsbildung bei personenbezogenen Unterlagen. In: *Archivar* 61 (2008) H. 3, S. 275–278; Robert KRETZSCHMAR, Positionen des Arbeitskreises Archivische Bewertung im VdA – Verband deutscher Archivarinnen und Archivare zur archivischen Überlieferungsbildung: Einführung und Textabdruck. In: *Der Archivar* 58 (2005) H. 2, S. 88–94; Eckhard MÖLLER, koop-owl.arch – Die Zusammenarbeit von Archiven und IT-Dienstleistern in Ostwestfalen-Lippe bei der Bewertung von Fachanwendungen. In: *Archivpflege in Westfalen-Lippe* 66 (2007), S. 41–44; Peter WORM, Elektronische Unterlagen der kommunalen Verwaltungen – ein Fall fürs Archiv? In: *Archivpflege in Westfalen-Lippe* 66 (2007), S. 39–41.

<sup>4</sup> Andrea HÄNGER, Rahmenvereinbarung mit den Statistischen Ämtern. In: *Archivar* 61 (2008) H. 2, S. 160–161; Abschlussbericht der ARK-Arbeitsgruppe „Bewertung von Statistikunterlagen“, erhältlich unter <http://www.ark.bundesarchiv.de>; Henning STEINFÜHRER, Voruntersuchungen zur Übernahme elektronischer Unterlagen des Thüringer Landesamtes für Statistik durch die thüringischen Staatsarchive. In: *Neue Konzepte für die archivische Praxis: ausgewählte Transferarbeiten des 37. und 38. Wissenschaftlichen Kurses an der Archivschule Marburg*, hrsg. v. Alexandra LUTZ, 2006, S. 261–294.

<sup>5</sup> Peter SANDNER, Karte – Luftbild – Geoinformationssystem. Archivierung und Präsentation digitaler Topographiedaten der Vermessungsverwaltung. In: *Archive und Öffentlichkeit*. 76. Deutscher Archivtag 2006 in Essen (Tagungsdokumentation zum Deutschen Archivtag 11), Fulda 2007, S. 127–134.

<sup>6</sup> Klaus DÖLL, Die Aufbewahrung sozialwissenschaftlich wichtiger Massendaten, masch. Köln o.J. [1965].

<sup>7</sup> Diese Feststellung wird verständlicherweise ungern öffentlich kundgetan. Eine beachtliche Reihe an manuellen Erfassungsergebnissen existiert natürlich. Nur beispielhaft sei hier genannt: *Deutsche Soldaten 1939–1945: Handbuch einer biographischen Datenbank zu Mannschaften und Unteroffizieren von Heer, Luftwaffe und Waffen-SS*, hrsg. von Christoph RASS und RENÉ ROHRKAMP, Köln 2007. Doch um wieviel besser wäre das Projekt geglückt, hätte ein findiger Archivar schon vor 50 Jahren Lochkartenbestände der Wehrmacht gesichert?

<sup>8</sup> Im aktuellen Heft des *Archivars* wird das Thema eingehend erörtert: Rolf DÄSSLER, Karin SCHWARZ, Archivierung und dauerhafte Nutzung von Datenbankanhalten aus Fachverfahren. In: *Archivar* 63 (2010) H. 1, S. 6–18.

<sup>9</sup> Diese Übersicht steht im Internet und wird regelmäßig aktualisiert: <http://www.landesarchiv-bw.de/web/50811>.

sind statistische Mikrodaten, ein wachsender Teil aber direkte Übernahmen aus Fachanwendungen. Insgesamt (einschließlich der statistischen Mikrodaten) hat das Landesarchiv Baden-Württemberg am 31.12.2009 Daten zu 59,6 Mio. Einzelfällen übernommen. Entspreche jeder Einzelfall einem papiernen Datenblatt, so wären rund 7,5 Kilometer an Regalkapazität zur Lagerung erforderlich.<sup>10</sup>

Fächert man die übernommenen Daten nach ihrem technischen Hintergrund und den Implikationen für Bewertung und Übernahme auf, ergeben sich folgende Typen:

- 1. Laufende Fachanwendung:** Diese Fachanwendungen sind im Echtbetrieb. Für sie mussten also Lösungen gefunden werden, wie und in welcher Form man einen bestimmten Datenbestand zu einem gegebenen Zeitpunkt übernimmt (ADV-Vollzug<sup>11</sup>, LÜVIS<sup>12</sup>, Jüdische Grabsteine<sup>13</sup>).
- 2. Stillgelegte Fachanwendung:** Diese Anwendungen existieren heute nur noch in Gestalt ihrer Nachfolgesysteme, sie wurden jedoch noch auf ihrem alten Stand angetroffen und konnten vor ihrer Abschaltung angezapft werden (PLUS<sup>14</sup>, POVO<sup>15</sup>, Straßendatenbank<sup>16</sup>, Umweltmeldestelle<sup>17</sup>). Hier war die Übernahme dadurch erleichtert, dass bereits eine Schnittstelle für den Export in die Folgeanwendung bestand – das bereits bekannte Prinzip „Bei Umzug Übernahme“.<sup>18</sup>
- 3. Referenz- und Berichtsdatenbank, Fachinformationssystem, Data Warehouse:** Das Umweltinformationssystem (UIS)<sup>19</sup> ist eine Referenzdatenbank, die aus einer Vielzahl von Fachanwendungen Daten bezieht, diese georeferenziert und die Daten dann zum Zweck einer landesweiten Verwendung für Kommunen, Landesbehörden und teilweise auch den Bürger verfügbar macht. Das UIS hat ein Berichtssystem mit eingebauten Exportschnittstellen. Zu den Vorteilen für das Landesarchiv weiter unten mehr.
- 4. Datenbankexporte an das Statistische Landesamt:** Viele der statistischen Mikrodaten können ebenfalls als Überlieferung aus Fachanwendungen verstanden

---

<sup>10</sup> Schätzung nach der bewährten Grundregel für Schriftgut des 20. Jahrhunderts (8000 Blatt je Laufmeter).

<sup>11</sup> Staatsarchiv Sigmaringen, Wü 32/2 T 2 und T 3, <http://www.landearchiv-bw.de/plink/?f=6-671540>.

<sup>12</sup> Staatsarchiv Ludwigsburg, EL 55, <http://www.landearchiv-bw.de/plink/?f=2-1185107>.

<sup>13</sup> Staatsarchiv Ludwigsburg, EL 228 b, <http://www.landearchiv-bw.de/plink/?f=2-1977090>.

<sup>14</sup> Hauptstaatsarchiv Stuttgart, EA 3/156, <http://www.landearchiv-bw.de/plink/?f=1-913658>.

<sup>15</sup> Hauptstaatsarchiv Stuttgart, EA 7/105, <http://www.landearchiv-bw.de/plink/?f=1-854363>.

<sup>16</sup> Staatsarchiv Ludwigsburg, EL 75 V, <http://www.landearchiv-bw.de/plink/?f=2-1023440>.

<sup>17</sup> Hauptstaatsarchiv Stuttgart, EA 16/201, <http://www.landearchiv-bw.de/plink/?f=1-662243>.

<sup>18</sup> Hierzu ausführlich Rolf LANG, Kai NAUMANN, Bei Umzug Übernahme – Bewertung und Ablieferung elektronischer Unterlagen im Rahmen von Systemmigrationen, Vortrag im Rahmen der 12. Tagung des Arbeitskreises Archivierung von Unterlagen aus digitalen Systemen, abgelegt unter [http://www.landearchiv-bw.de/sixcms/media.php/120/50218/Bei\\_Umzug\\_Uebernahme01.pdf](http://www.landearchiv-bw.de/sixcms/media.php/120/50218/Bei_Umzug_Uebernahme01.pdf).

<sup>19</sup> Generallandesarchiv Karlsruhe, 518-1, <http://www.landearchiv-bw.de/plink/?f=4-740417>.

werden,<sup>20</sup> denn heute werden Mikrodaten für die Statistik zunehmend durch Datenbankabfragen aus Inhalten von Fachanwendungen generiert.

## II.

Je nach den spezifischen technischen Eigenschaften gibt es im Ablauf der Bewertung und Übernahme kleine Besonderheiten zu beachten. Grundsätzlich aber werden im Landesarchiv bei Bewertung und Übernahme die folgenden Schritte vollzogen:

### 1. Bewertung

Am Anfang steht die Bewertung im Kontext. Sobald eine Fachanwendung einem Aussonderungsreferenten bekannt wird, ist zunächst ein Abgleich mit der sonstigen Überlieferung der zuständigen Behörde angezeigt. Hierbei werden die vorliegenden Bewertungsmodelle für den jeweiligen Verwaltungszweig aus Baden-Württemberg oder anderen Bundesländern zu Rate gezogen. Da diese jedoch Daten aus Fachanwendungen oft ausdrücklich nicht mit berücksichtigen, ist viel Improvisation gefragt – und ein reger Austausch in der Arbeitsgemeinschaft Überlieferungsbildung (AGÜ) des Landesarchivs. Ein wichtiges Stichwort ist hierbei Komplementarität. Nicht nur die Überlieferungen der Kommunen und Kreise, der Oberbehörden und Ministerien, sondern auch die papiernen und die elektronischen Unterlagen sind so abgestimmt, dass sie sich harmonisch ergänzen.<sup>21</sup>

Die Fachanwendungen haben in diesem Konzert ihren Part meist darin, den wesentlichen Niederschlag der Bearbeitung von Einzelfällen zu dokumentieren. Datentabellen stehen zwar nicht mehr in ihrem ursprünglichen Bearbeitungskontext in der Fachanwendung. Aber schon Carsten Müller-Boysen konnte sich im Jahr 2000 vorstellen, dass bestimmte Inhalte von Systemen um einen Teil ihrer Bearbeitungsinformationen beraubt werden könnten, wenn dies für künftige Nutzungen als wenig hilfreich eingeschätzt würde.<sup>22</sup>

Aus den Daten einer Fachanwendung lassen sich demnach Einzelauskünfte erteilen. Sie geben also im Falle der Lehrerdatenbank PLUS Antwort auf die Frage: „Senden Sie bitte einen Nachweis, dass ich von 1987 bis 1991 bei der Grundschule XY gearbeitet habe“.

---

<sup>20</sup> Dies gilt nach dem Kenntnisstand des Landesarchivs Baden-Württemberg u.a. für die Einkommensteuerstatistik, die polizeiliche Kriminalstatistik, die Strafvollzugsstatistik und eine ganze Reihe von Umweltstatistiken.

<sup>21</sup> Jürgen TREFFEISEN, wie Anm. 3; Christian KEITEL, *Erweiterte Zeichenverwaltung: Elektronische Archivierung im Landesarchiv Baden-Württemberg* (Vortrag, BSZ-Workshop, 4.7.2005), <http://www.landesarchiv-bw.de/web/46914>.

<sup>22</sup> Carsten MÜLLER-BOYSEN, *Flat file oder virtuelle Behörde? – Was erwartet der Benutzer?* In: Ulrich NIESS (Hrsg.), *Auf der Suche nach archivischen Lösungsstrategien im digitalen Zeitalter. Beiträge zur 4. Jahrestagung des Arbeitskreises „Archivierung von Unterlagen aus digitalen Systemen“ im Stadtarchiv Mannheim 10.–11.4.2000, Mannheim 2001*, S. 63–68, hier S. 67.

Wenn aber alle Datensätze einer Fachanwendung erhalten werden, können sie auch Fragen über die Gesamtheit aller Fälle beantworten, zum Beispiel: „Wieviele Lehrer im Land waren 1988 in der Besoldungsgruppe A 12 an Realschulen angestellt?“. Solche Auswertungen lassen sich über Datentabellen viel besser ausführen als über Akten. Dies gilt übrigens auch für Akten in elektronischer Form, die ja ein komplexes Geflecht aus XML-Dateien und Primärdateien sind.

Sie können aber auch die Auswahl bestimmte Papierunterlagen erleichtern, etwa mit der Antwort auf die Frage: „Wer war die jüngste Schulleiterin Baden-Württembergs in den 90er Jahren des 20. Jahrhunderts?“.

Die gewonnenen Daten stellen also in erster Linie eine bessere Alternative zur massenhaften Übernahme von Personal- und Fallakten dar. Es gibt aber auch drei Nebeneffekte. Zum einen wird die Übernahme von Daten statt Papier erhebliche Lagerkosten einsparen. Zum zweiten entfällt der Erschließungsaufwand, da die Daten bereits tabellarisch vorliegen.<sup>23</sup> Zum dritten erleichtern die Daten die Papieraktenaussonderung: eine Fachanwendung half dem Staatsarchiv Sigmaringen kürzlich, aus mehreren 100 Metern Personalakten die berühmten Fälle zu isolieren und zu erhalten.<sup>24</sup>

Wie man Daten aus Fachanwendungen und Papierunterlagen gemeinsam bewertet, zeigt ein Beispiel aus der Praxis: die Fachanwendungen der Steuerverwaltung bestehen aus einem Grundinformationsdienst (GID) zum Steuerpflichtigen und den jährlich zugeordneten Speicherkonten. Ein Abgleich mit dem Merkmalsvorrat der Mikrodaten der Einkommensteuerstatistik zeigte, dass die Überlieferungsform der Mikrodaten die wesentlichen Merkmale des GID und der Speicherkonten hinreichend abbildet. Daher wurde in Baden-Württemberg der Entschluss gefasst, derzeit keine besondere Exportschnittstelle aus dem GID anzustreben und stattdessen die Mikrodaten beim Statistischen Landesamt zu übernehmen.<sup>25</sup>

Entscheidet man sich für eine Übernahme, wird zunächst nur eine umrisshafte Definition der zu übernehmenden Inhalte vorhanden sein. Bei unserer Übernahme aus der Fachanwendung ADV-Vollzug etwa hieß diese Vorgabe: „Landesweite Übernahme der Kerndaten zu jedem Inhaftierungsfall im Land. Außerdem möglichst viele Inhalte der Fachanwendung zu denjenigen baden-württembergischen JVAs, von de-

---

<sup>23</sup> Weil die Datenexporte aus Fachanwendungen in der Regel keine Verzeichnung unterhalb der Ebene der Datenbank selbst erfordern, können diese in einem einzelnen digitalen Paket (AIP/Archival Information Package) abgelegt und auch wieder benutzt werden. Die Inhalte eines DMS hingegen müssen in aller Regel in kleine Portionen zerteilt und einzeln im Findmittelsystem nachgewiesen werden. Vgl. ZÜRCHER THRIER, wie Anm. 2.

<sup>24</sup> Ein Beitrag von Franz-Josef ZIWES hierzu erscheint demnächst.

<sup>25</sup> Dies gilt nur für Überlegungen im Zusammenhang mit der Überlieferung von Daten. Bei der Übernahme von Metadaten zu elektronischen Bestandteilen von Steuerakten wird der GID (künftig bundeseinheitliche Anwendung GINSTER) durchaus eine Rolle spielen.

nen nach geltendem Bewertungsmodell<sup>26</sup> größere Mengen Papierakten übernommen werden“.<sup>27</sup> Hier ist es immer sinnvoll, die fachlich zuständige Abteilung innerhalb der Behörde zu beteiligen.

Ebenfalls vorab zu definieren ist die zeitliche Auswahl von Inhalten innerhalb der Fachanwendung. In der Regel wurden bisher alle enthaltenen Fälle an einem gegebenen Zeitpunkt übernommen. Lediglich bei ADV-Vollzug einigte sich das Landesarchiv mit der Justizverwaltung, ganz in Analogie zu den Papierakten zu verfahren, nämlich jährlich nur diejenigen Fälle zu übernehmen, die jeweils im darauffolgenden Jahr zur Löschung anstehen.

## **2. Schnittstellendefinition**

Zu einer Schnittstelle gehört eine Definition, welche Daten in welcher Struktur auf welchem Weg ins Archiv gelangen. Im ersten Schritt ist die Binnenstruktur der Fachanwendung zu analysieren. Sie besteht aus drei Abstraktionsebenen: der Datenhaltung, der Logik und der Darstellung. Diese Aufteilung ermöglicht erst die Geschwindigkeit, mit der Fachanwendungen Erfassung, Recherche und Darstellung abwickeln.

In diesem Abstraktionsgrad liegt aber die archivische Herausforderung. Was als Datenblatt am Bildschirm noch große Ähnlichkeit mit dem Deckblatt einer Fallakte hat, entpuppt sich im Hintergrund, in der relationalen Datenhaltung, als ein Sammelsurium von Feldern aus verschiedenen Tabellen. Beispielsweise liegt der Name einer Person in Tabelle A, ihre Postleitzahl in Tabelle B, ihre Examensnote aber in Tabelle C. Verknüpfendes Element ist eine Personalnummer, die in jeder Tabelle die betreffenden Felder begleitet. Nur durch eine Abfrage, die in der Standardsprache SQL (Structured Query Language) definiert ist, finden die Informationen in der Logik zusammen und werden in der Darstellung übersichtlich aufgebaut. SQL ist ein reines Textformat, dessen Erhaltung leicht fällt und dessen Grundvokabular seit 1979 stabil ist.<sup>28</sup> Ältere Anwendungen aus den Jahren vor 1990 verwenden ein einfaches Datenmodell, das mit 10 bis 20 Tabellen auskommt. Fachanwendungen der Nuller Jahre jedoch verwenden meist hunderte von Tabellen, um den vielfältigen Nutzerwünschen gerecht zu werden.

Es ist eine archivische Bewertungsaufgabe, die wesentlichen Inhalte einer komplexen Fachanwendung auf ein einfaches Datenmodell zu reduzieren. Diese archivi-

---

<sup>26</sup> Barbara HOEN, Konrad KRIMM, Jürgen TREFFEISEN, Erweitertes Auswahlmodell bei Massenakten der Justiz, 1999. Erhältlich beim Landesarchiv, derzeitige URL <http://www.landesar-chiv-bw.de/web/46738>.

<sup>27</sup> Die Rumpfdaten (Wü 32/2 T 2) enthalten Angaben zur Person, zu verwahrenden JVA, zur einweisenden Behörde, zu Beginn und Ende der Inhaftierung. Die Auswahldaten (Wü 32/2 T 3) enthalten den größten Teil der in ADV-Vollzug hinterlegten Informationen zu Inhaftierungsfällen in den sechs Justizvollzugsanstalten Freiburg, Hohenasperg (Sozialtherapie und Vollzugskrankenhaus), Mannheim, Ravensburg, Schwäbisch Gmünd. Vgl. Online-Findmittel unter <http://www.landesar-chiv-bw.de/plink/?f=6-671540>.

<sup>28</sup> <http://en.wikipedia.org/wiki/SQL>.

schen Datenmodelle sollten so gestaltet sein, dass sich ein einzelner, datenbanker-fahrener Nutzer binnen kurzer Zeit in ihnen zurechtfinden sollte.

Für die mit dieser Arbeit im Landesarchiv befassten Archivare gibt es bisher vier Möglichkeiten, ein reduziertes Datenmodell zu erstellen.

- a) Der Archivar stuft den Export eines Berichtssystems oder das Transferformat bei einer Systemmigration als archivfähig ein und verfügt dessen Übernahme ohne Veränderungen.
- b) Der Archivar geht von der Darstellungsebene aus und schlägt eine Auswahl von Tabellen vor, die in einer bestimmten Verknüpfung zu erstellen sind. Ein Mitarbeiter der Behörde programmiert anschließend vor Ort die SQL-Definition, deren Ergebnis geprüft, eventuell nachgearbeitet und abgenommen wird.
- c) Wie in b), aber die SQL-Definition wird von einem Informatiker des Landesarchivs erstellt.
- d) Wie in b), aber der Archivar untersucht selbst die Datenhaltung und die dort existierenden Tabellen und schreibt selbst in SQL eine Definition, die vom hauseigenen Informatiker und von der Behörde geprüft wird.

Bislang steht nicht fest, welche Methode auf Dauer die häufigste sein wird. Das Vorgehen hängt von der Komplexität der Anwendung und vom Entgegenkommen der Behörde ab. Fest steht aber, dass gewisse Kenntnisse in Datenbankwerkzeugen und SQL für spezialisierte Archivare von Vorteil sind. SQL ist ein Werkzeug, um Archivaren den Vermittlungsprozess zwischen Technik und sachbezogener Bewertung erleichtern.

Bei allen Varianten müssen Behörde, Archivare und Informatiker im Archivdienst eng zusammenarbeiten. Nicht jede von der Sache her gebotene Umformung der Daten ist nämlich technisch im Zeit- und Kostenbudget realisierbar, insbesondere bei den Varianten b) bis d). Gelegentlich steht deshalb die Entscheidung an, ob man für ein bestimmtes Datenfeld einen Arbeitstag mehr investiert oder nicht. Informatiker werden gebraucht, weil sie die Technik beherrschen und abschätzen können. Archivare werden gebraucht, weil sie den Wert der Information abschätzen können und die Verantwortung für ein optimales Bewertungsergebnis tragen.

Zur Dokumentation der Schnittstelle wird eine Übersicht aller in der Fachanwendung vorhandenen Datenbanktabellen und ihrer Verknüpfungen erzeugt. Dazu kommen alle greifbaren Handbücher für Nutzer und Entwickler sowie, wenn möglich, Unterlagen aus dem Entwicklungsprozess der Fachanwendung.

### **3. Export und Ingest**

Nach erfolgreicher Definition der Schnittstelle erfolgt zunächst ein Testexport, der auf die gewünschten Eigenschaften hin überprüft wird. Hierbei können sich die im An-

hang aufgezählten Fehler einstellen, auf deren Vorkommen stets aufmerksam zu achten ist.

Die Echtübernahme erfolgt mittels der frei erhältlichen Software IngestList, die dem Arbeitskreis bereits 2009 vorgestellt wurde.<sup>29</sup> Die ursprünglich für Dateisammlungen angelegte Software besitzt inzwischen eine Datenbank-Anbindung. IngestList ist damit in der Lage, eine Verbindung zu einem Datenbankserver herzustellen und die gewünschten Tabellen in das Format CSV (Character separated values) zu exportieren.<sup>30</sup> Die mit IngestList gewonnene Kapsel aus Primärdaten und Metadaten wird im nächsten Schritt in das Erhaltungssystem DIMAG übertragen, auf Vollständigkeit geprüft, in der Verzeichnisstruktur des zuständigen Staatsarchivs verankert und für die Nutzung beschrieben.

Werden die Daten als Exportdateien (Festbreitenformat oder CSV) abgeliefert, so kommt der Übernahmeprozess ohne die Datenbank-Anbindung aus. IngestList setzt in diesem Fall bei den Exportdaten auf den Datenträgern in der Behörde an.

Wenn die Daten aus einem Berichtssystem übernommen werden, kann der Archivar sich die Auswahl selbst definieren und die CSV-Dateien „frei Haus“ auf seinen Arbeitsplatz-PC herunterladen, wo wiederum IngestList zum Einsatz kommt, um den weiteren Lebensweg zu dokumentieren.

Jedem Transferschritt in diesem Geschäftsgang folgt ein Validierungsschritt, der die Richtigkeit der Übertragung sicherstellt. Der gesamte Weg zwischen Fachanwendung und Landesarchiv wird außerdem so gestaltet, dass die Daten gegen Einsichtnahme oder gar Kopieren durch Unbefugte geschützt sind.

Nach dem Einstellen der Daten werden in DIMAG noch hinzugefügt:

- die SQL-Definition,
- eine Beschreibung der Tabellen, der Felder und ihrer Verknüpfungen (in XML),
- die gesammelte Dokumentation (meist in PDF/A).
- Ein Aufbereitungsbericht in reinem Textformat hält die im konkreten Fall vollzogenen Entscheidungen fest und begründet diese.
- Das in IngestList geborene Protokoll zum Archivobjekt dokumentiert, falls quellenkritische Fragen aufkommen sollten, ausführlich und sekundengenau den Übergang zwischen Fachanwendung und DIMAG.

---

<sup>29</sup> Christian KEITEL, Rolf LANG, Ingest von Fachverfahren im Landesarchiv Baden-Württemberg. Anmerkungen zu Authentizität, Prozessen und Softwareentwicklung. In: Entwicklung in den Bereichen Records Management / Vorarchiv – Übernahme – Langzeitarchivierung. 13. Tagung des Arbeitskreises „Archivierung von Unterlagen aus digitalen Systemen“ am 27./28.4.2009 in St. Gallen, St. Gallen 2009, S. 35–45.

<sup>30</sup> Eine alternative Möglichkeit ist die Nutzung des Werkzeugs SIARD, das die Inhalte einer Datenbank nach XML umwandelt. SIARD ist erhältlich beim Schweizerischen Bundesarchiv <http://www.bar.admin.ch>.



Die Übernahme wird anschließend der Behörde bestätigt, zudem wird ein Termin für die nächste Übernahme verabredet und der Vorgang auf Wiedervorlage gelegt. Im Idealfall wird die SQL-Definition im Datenbankserver der Fachanwendung verankert und steht dort für den nächsten Export bereit.

#### **4. Fortschreibung**

Je nachdem wie häufig die Übernahme stattfinden soll, stehen zu den darauffolgenden Terminen Nacharbeiten an. Nicht selten verändert die Behörde das Datenmodell oder werden von der Herstellerfirma der Fachanwendung neue Module aufgespielt. Wenn die abliefernde Stelle gut mitarbeitet, werden die hinterlegten SQL-Definitionen des Landesarchivs diesen Änderungen angepasst. Andernfalls verliert die Schnittstelle den Anschluss, wird das Archiv vom Fortschritt abgehängt.

Die festgelegte Schnittstelle ist also in jedem Fall zum nächsten Übernahmetermin in einem Testexport wiederum genauestens zu überprüfen. Auch hier zeigt sich, dass spezialisierte Archivare gut beraten sind, etwas von SQL zu verstehen, um kleinere Fehler eventuell selbst auszubessern.

### **III.**

Die Erhaltung der so gewonnenen Daten dürfte sich in den nächsten Jahrzehnten auf die Sicherstellung der Bitfolgen beschränken, da alle verwendeten Formate (CSV, PDF/A, Text, XML) als recht stabil einzustufen sind. Viel spannender ist die Nutzung. Bislang wurden erst einmal Daten aus einer Fachanwendung als Kopie zur Nutzung weitergegeben; zweimal wurden Auskünfte aus einer Datenbank erteilt.

Selbst aufgrund der bisher sehr dünnen Faktenlage ist derzeit festzuhalten:

1. Die Nutzung von Fachanwendungen ist eine Sondernutzung. Datenbanken sind nicht Inhalt der archivarischen Ausbildung. Ein Nutzer in einem Staatsarchiv vor Ort, das provenienzmäßig zuständig ist, kann unter Umständen mangels entsprechend aus- oder fortgebildeter Mitarbeiter nicht wie gewohnt bedient werden.

Das Landesarchiv unterscheidet daher zwischen Nutzung als Sache des zuständigen Staatsarchivs und Bereitstellung als Sache des Querschnittsreferats für elektronische Archivierung beim Staatsarchiv Ludwigsburg. Nutzung umfasst den Erstkontakt, die Klärung des Nutzungsvorhabens, die Prüfung auf Zugangsberechtigung und die Beauskunftung. Bereitstellung umfasst die Klärung der technischen Nutzungsmöglichkeiten und die Umsetzung der Einsichtnahme. In der Praxis wird daher ein Nutzer von archivierten Fachanwendungen teils in Ludwigsburg, teils im zuständigen Staatsarchiv betreut.

2. Als Verwahrer großer Mengen maschinenlesbarer Daten hat das Landesarchiv wesentlich mehr auf technischen und organisatorischen Datenschutz zu achten als in der Vergangenheit. Die bisherigen Nutzungsfälle waren für den Datenschutz glücklicherweise nicht relevant. Die Einrichtung von Kiosk-PCs, die außer der Einsichtnahme keine Übertragungswege nach außen bieten dürfen, wird aber demnächst erforderlich sein.
3. Die archivrechtliche Beurteilung bestimmter Nutzungsfälle wird einheitlich zu klären sein. Besonders heikel ist die Nutzung von Mikrodaten der Statistik, weil hier das Bundesarchivgesetz eine unverkürzbare Sperrfrist von 60 Jahren vorschreibt, soweit personenbezogene Merkmale enthalten sind.<sup>31</sup> Die Beurteilung, ob eine unverkürzbare Sperrfrist für personenbezogene Unterlagen greift oder nicht, hängt bei Mikrodaten wiederum von dem Grad der Anonymisierung ab. Unverkürzbare Sperrfristen liegen bei Exporten aus Fachanwendungen glücklicherweise selten vor. Gleichwohl ist bereits die einheitliche Ausübung des Ermessens im Rahmen der Sperrfristenverkürzung bei solchen Nutzungsvorhaben eine Herausforderung, da zunächst wenige Musterfälle vorliegen.
4. Das Landesarchiv braucht langfristig eine browserfähige Nutzungsumgebung für Datenbanken für das Intra- und das Internet. Vorbild könnten das britische National Digital Archive of Datasets<sup>32</sup> oder das Access to Archival Databases<sup>33</sup> in den USA sein, die beide seit vielen Jahren im Echtbetrieb operieren. Erste Ansätze hierfür sind im neuesten Prototyp von DIMAG bereits verwirklicht. Auch Logiken, die einzelne Teile einer Datentabelle zur Anonymisierung ausfiltern (z.B. alle Fälle vor 1985, ohne bestimmte Felder), werden erforderlich sein.

Insgesamt steht inzwischen eine ganze Reihe an Werkzeugen und Arbeitsvorlagen bereit. Da Fachanwendungen sehr häufig bei mehreren Verwaltungseinheiten, auch länderübergreifend, im Einsatz sind, ergeben sich vielfältige Möglichkeiten der Zusammenarbeit.

Und nicht zuletzt ist die Übernahme von Fachanwendungen für Archivare und Informatiker eine gute Vorübung für die nächst größere Herausforderung. Nämlich für die Übernahme von DMS-Systemen, die für bestimmte andere Nutzungsmotive auch dringend gebraucht werden<sup>34</sup>. Hinter einem DMS verbirgt sich nämlich nichts anderes als eine komplexe Fachanwendung mit der Fähigkeit, digitale Textdokumente, Bilder oder Audiovisuelles anzuzeigen.

---

<sup>31</sup> Vgl. Wolf BUCHMANN, Michael WETTENGEL, Auslegung des Bundesstatistikgesetzes bei der Archivierung von Statistikunterlagen. In: Der Archivar 49 (1996), H. 1, S. 67–74.

<sup>32</sup> <http://www.ndad.nationalarchives.gov.uk/>.

<sup>33</sup> <http://aad.archives.gov/aad/>.

<sup>34</sup> Nahezu im synonymen Gebrauch sind CMS = Content Management System, ECM = Enterprise Content Management, GEVER = Geschäftsverwaltungssystem, (E)RMS = (Electronic) Records Management System, VBS = Vorgangsbearbeitungssystem.

Es ist also Zeit für uns Archivarinnen und Archivare, uns mit der Übernahme von Daten aus Fachanwendungen vorzuwagen und uns die Welt der Datenbanken zu erschließen.

## Anhang

### Mögliche Fehler beim Export von Daten aus Fachanwendungen

#### Die falschen Inhalte werden exportiert:

Beim Export aus dem Datenbanksystem werden nicht die Daten exportiert, die exportiert werden sollten.

#### Nicht alle Datensätze werden exportiert:

Anzahl der exportierten Datensätze stimmt nicht mit der Quelltable im Datenbanksystem überein.

#### Feldinhalte sind unvollständig:

Feldinhalte, die das normale Maß (je nach Hersteller 1 bis 16 KByte) überschreiten, werden unvollständig oder fehlerhaft exportiert.

#### Die CSV-Syntax wird nicht eingehalten:

Felder, die ein Semikolon oder Zeilentrenner enthalten können, sind nicht mit Anführungszeichen begrenzt.

- Falsch: Zelle1;Zelle;2;Zelle 3
- Richtig: Zelle1;"Zelle;2";Zelle 3

#### Ungültige Zeichen werden in Textdateien exportiert:

Zeichen mit Hex-Werten unter 32 (hex20) mit Ausnahme von Line Feed (hex0A), Carriage Return (hex0D) und Tab (hex09) führen zu Importproblemen.

#### Zeichen werden falsch dargestellt:

- Der Zeichensatz für den Export entspricht nicht dem Zeichensatz in der Datenbank.
- In einer Textdatei werden mehrere Zeichentabellen (Latin1 und UTF-8) gemischt verwendet.

#### Einzelne Zeichen werden nicht exportiert (z.B. Umlaute).

#### Datenbestände sind zueinander inkonsistent:

Datenbank A enthält Angaben zu n Fällen, Datenbank B soll andere Angaben zu den gleichen n Fällen enthalten. Die Anzahl der Datensätze stimmt aber nicht überein. Kommt besonders bei GIS-Anwendungen zwischen Geometrie- und Fachdaten-Komponente vor.