

Archivierung von Registerdaten aus Fachanwendungen

von Martin Kaiser

Datenbanken und Fachanwendungen

Die Archivierung von Datenbanken, bzw. von Daten daraus, ist schon seit einiger Zeit ein häufig diskutiertes Problem der digitalen Archivierung im staatlichen Kontext. Dazu muss sogleich präzisiert werden, dass Datenbanken in staatlichen Verwaltungen in den wenigsten Fällen einfach so, gewissermassen im luftleeren Raum, angetroffen werden. Sie sind vielmehr Bestandteile grösserer Anwendungen. Man spricht dabei oft von datenbankgestützten Anwendungen, in der schweizerischen Terminologie auch von "Fachanwendungen", in Deutschland von "Fachverfahren". Solche Fachanwendungen sind in den staatlichen Verwaltungen weit verbreitet¹ und zudem aus informatikgeschichtlicher Sicht relativ alt. Sie wurden in den Verwaltungen bereits ab den sechziger Jahren eingeführt. Entsprechend gehören solche Daten oft zu den ersten digitalen Angeboten, welche die Archive erhalten – falls die Anwendung nicht einfach ab- oder aufgelöst wird.

Etliche Archive sehen sich seit Kurzem auch mit dem Problem konfrontiert, dass Daten in Fachanwendungen aus Datenschutzgründen innerhalb einer bestimmten Frist gelöscht werden müssen und werden

¹ Bestandesaufnahmen von einigen kleineren Staatsarchiven in der Schweiz haben vor 7 Jahren für die jeweilige Kantonsverwaltung Zahlen im niedrigen dreistelligen Bereich ergeben. Christoph Popp vom Stadtarchiv Mannheim hat zur Grössenordnung des Problems in deutschen Stadtverwaltungen kürzlich ein paar Richtwerte zusammengetragen: Bielefeld 200, Stuttgart 500. Damit ist über die Archivwürdigkeit dieser Fachanwendungen natürlich noch nichts ausgesagt!

und somit für eine spätere Archivierung nicht mehr verfügbar sind². Gründe dafür sind eine ausgeprägtere Sensibilität bezüglich Datenschutzfragen und der Umstand, dass in datenbankgestützten Systemen durch die Datenvernetzung persönlichkeitsrechtlich bedenkliche Datensammlungen zusammengetragen werden können. Im Falle von neuen Sammlungen werden diese in der Regel auch datenschutzrechtlich geprüft und verbindliche Löschrufen festgelegt. Die Daten werden den Archiven in diesen Fällen meist kurzfristig angeboten, für die archivische Bewertung, technische Analyse und die Entwicklung von Übernahmestrategien und Lösungen bleibt wenig Zeit.

Bewertung einer Fachanwendung

Während das technische Vorgehen bei der Archivierung von Datenbanken und dazu adäquate Dateiformate relativ breit diskutiert und erforscht sind, wird die archivische Bewertung von Datenbanken und Fachanwendungen häufig nur implizit diskutiert. Ziel der Bewertung ist es, die zu archivierende Datenmenge zu reduzieren, aber auch, und dieser Aspekt ist bei Fachanwendungen besonders wichtig, die Komplexität des zu übernehmenden Datenbestandes zu reduzieren. Durch verringerte Komplexität reduziert sich der Aufwand für die Beschreibung und Erschließung, und eine sichere Überlieferung gestaltet sich einfacher.

Sehen wir uns die bekannten, in der Archivwelt bereits mehr oder weniger ausführlich diskutierten Lösungswege hier kurz an.

² Ein Beispiel aus der Schweiz sind die kantonal verwalteten Daten der Arbeitslosenversicherung (Versicherungsdaten und Arbeitsmarkt-Massnahmen), wo mit der Einführung einer EDV-Lösung und der datenschutzrechtlichen Beurteilung der damit verbundenen Datensammlung eine Löschung der Daten nach drei Jahren verbindlich geworden ist.

Ein erster Ansatz besteht darin, die Bewertung auf der Ebene ganzer Datenbanken bzw. ganzer Fachanwendungen anzusiedeln. Dabei soll die Fachanwendung mit der ganzen Datenbank archiviert oder kasziert werden. Dieser Ansatz beruht auf der Überlegung, dass eine Fachanwendung eine bestimmte Funktion innerhalb des staatlichen Handelns erfüllt und als solche bewertet werden kann. Die Archivierung verfolgt dann das Ziel, möglichst alle Aspekte zu erhalten. Das geschieht durch Archivieren der gesamten Datenbank zusammen mit Benutzerhandbuch, *Screenshots*, Beispiellisten etc. aus der betreffenden Anwendung. Der Vorteil ist hier, dass sich eine vertiefte technische und inhaltliche Analyse der Fachanwendung erübrigt. Diese Vorgehen drängt sich in der Regel bei der Ablösung oder Migration einer Fachanwendung auf, weil dann die gesamte mit der Fachanwendung verbundene Datenbank integral übernommen werden kann.

Ein zweiter Ansatz geht davon aus, dass eine Fachanwendung dem Primärnutzer nicht als Datenbank mit normalisierten und optimierten Tabellen gegenübertritt, sondern als ein Werkzeug, das Entitäten der realen Welt abbildet. Ziel der Archivierung von Fachapplikationen muss es deshalb sein, nutzerspezifische Sichten auf die Daten zu erhalten. Das könnte wohl am einfachsten durch direktes Aufbewahren von statischen Ergebnislisten, Tabellen und Dokumente geschehen, die aus der Fachapplikation erzeugt wurden. Dabei entsteht aber viel Redundanz, und eine wie auch immer geartete Neuordnung der Information bzw. auch eine statistische Auswertung der Daten ist nicht mehr möglich. Deshalb erscheint es sinnvoller, die unter diesen statischen Repräsentationen von Nutzersichten liegende Datenbank-Sicht (*View*) zu archivieren, und zwar als Tabellen oder als Relation von Tabellen. Dabei ist es möglich, Bewertungsentscheide bezüglich Zeitraum und *Sampling* technisch zu realisieren.

Ein dritter Ansatz fasst datenbankbasierte Anwendungen als Hilfsmittel für die Verwaltungsarbeit auf. Der Einsatz dieser Hilfsmittel findet in der Regel seinen Niederschlag im Ausstoss von Dokumenten und Reports aus der Anwendung. Dabei kann es sich um Übersichtslisten handeln, aber auch um Datenblätter zu einzelnen Objekten. Viele dieser Dokumente entstehen ganz normal und real im täglichen Verwaltungsprozess, häufig mit dem Ziel, an Kunden oder andere Verwaltungsstellen verschickt zu werden. Deshalb könnte sich eine archivische Bewertung auf diese Dokumente konzentrieren, eine Auswahl aus ihnen als archivwürdig identifizieren und diese im PDF/A-Format archivieren. Die so bewerteten Dokumente können in der Fachanwendung meistens in einem Dossier³-Zusammenhang betrachtet und die Archivierung dieser Dossiers als Archivierung von Daten aus einem Geschäftsverwaltungssystem gesehen werden. Bei massenhaft gleichförmigen Serienakten könnte eine mit Bedacht durchgeführte PDF-Erstellung sogar die Möglichkeit offenhalten, später aus den Dokumenten wieder Daten zu extrahieren und in Tabellenform zu analysieren.

Ein vierter Ansatz stellt die historische Betrachtung in den Vordergrund. Die heutigen Fachanwendungen in der Verwaltung sind im Prinzip Weiterentwicklungen von Registersystemen⁴ – von Karteien und

³ Dossier bzw. Geschäft ist in der schweizerischen Verwaltungsterminologie der Fachbegriff für Akte oder Vorgang. Die beiden Begriffe sind gleichwertig, im Unterschied zur Verwendung von Akte und Vorgang in Deutschland als hierarchische Beziehung. Ein Dossier ist immer im hierarchisch aufgebauten Ordnungssystem der Verwaltung oder des Amtes abgelegt. Aus der Position im Ordnungssystem ergibt sich die Dossiernummer, bzw. das Aktenzeichen. Das Ordnungssystem entspricht dem deutschen Aktenplan, wobei dieser Begriff in der Schweiz durchaus auch üblich ist.

⁴ Erste Datenbanken mit grafischer Benutzerschnittstelle haben diesen Zusammenhang auch visuell genutzt, ein Beispiel dazu ist etwa die bekannte Datenbank *Filemaker*. Man beachte die Rollkartei zur Navigation, links auf dem *Screenshot* in der Dissertation von Hillman, D. C. A. (1996). Improved coding and data management for

Registerbüchern. Verweise auf weiterführende Unterlagen/Dossier und Vorgänge werden dabei an der entsprechenden Stelle im Register eingetragen. Damit ist die Registerkarte einerseits Zugriffsmittel zum Dossier und andererseits eine Kurzfassung der in der Regel auf dem Dossierdeckblatt festgehalten Vorgangsinformationen⁵. Wenn Fachanwendungen auf diese Registerfunktion reduziert werden können, dann liegt der Schluss nahe, diese Daten direkt als "archivische Registerdaten", d.h. als Metadaten, in das Archivinformationssystem (AIS) zu übernehmen. Sie würden im AIS dann wiederum auf die archivierten Dossiers (analog oder digital) verweisen.

Im letzten hier betrachteten Fall bewerten wir nur die Registerfunktion der Fachanwendung als archivwürdig⁶ und archivieren dieses Register wiederum als Datenbank. Das Register ist in diesem Fall ein Teil der vollständigen Datenbasis der Fachapplikation, das als eine oder mehrere Datenbank-Views realisiert wird. Es werden bei dieser Lösung in der Regel bereits bestehende Archivreihen, bestehend aus Registerbüchern und Registerkarteien, ungebrochen in die digitale Zeit fortgesetzt. Durch die in der Bewertung erreichte Reduktion der Daten, in der Regel auf wenige Prozente der ursprünglichen Datenbasis, ist es möglich, auch grösste Datenreihen ohne *Sampling* zu übernehmen.

discourse analysis: *A case study in face-to-face and computer-mediated classroom interaction*. Doctoral dissertation, University of Cambridge, Cambridge. Figure 1 <http://www.quahog.org/thesis/comparing.html>

⁵ Nach heutiger Records Management Terminologie sind das die *Workflow*- oder Prozessdaten.

⁶ In diesem Fall geht es spezifisch um den historischen Nachweis bestimmter Verwaltungshandlungen und/oder um Daten für sozialgeschichtliche Studien.

Bewertung und Datenreduktion

Zwei der skizzierten Ansätze kommen nicht in Frage, weil sie entweder die Komplexität nicht reduzieren (Ansatz 1) oder Primärdaten im Archiv unzulässigerweise in Metadaten umwandeln (Ansatz 4). Die anderen Ansätze bilden die theoretische Basis für zwei Beispiele aus dem Beratungsportfolio der KOST, die im folgenden etwas detaillierter betrachtet werden, und zwar auch unter dem Gesichtspunkt, dass es in einer längerfristigen Strategie zur Archivierung von Fachanwendungen nicht nur um das Reduzieren von Datenmenge und Komplexität der Datenbasis einer Fachanwendung gehen kann, sondern dass auch aus ökonomischen Gründen das Ziel sein muss, fachanwendungsübergreifend die Sicht auf die Datenbasis zu vereinheitlichen. So können wir einerseits weiterhin kontinuierliche Archivreihen über bestimmte, im Prinzip unveränderte Verwaltungstätigkeiten bei wechselnden technischen Lösungen pflegen, verhindern andererseits aber auch, im Laufe der Zeit mit einer Vielzahl von unterschiedlichen Archivlösungen vom Typ „Fachanwendung A archiviert in der Archivlösung X“ konfrontiert zu sein.

Wir wollen also in einer Archivreihe auch bei sich aus technischen und wirtschaftlichen Gründen fortlaufend verändernden Fachanwendungen über längere Sicht den gleichen Lösungsweg zur Archivierung verwenden und damit eine langfristig stabile Datenbasis für die Archivierung bilden können. Aus ökonomischen Gründen ist es natürlich auch vorteilhaft, wenn wir Fachanwendungen, die unterschiedliche Archivreihen bilden, auf gleiche Weise archivieren können.

Register oder Dossier? Eine Frage der Bewertung

Registerdaten der Gebäudeversicherung

In der Schweiz ist die Gebäudeversicherung, bzw. Brand-Assekuranz kantonal und öffentlich-rechtlich organisiert. Es besteht grösstenteils ein Versicherungszwang. Die meisten Versicherungen wurden bereits Anfangs 19. Jahrhundert gegründet, und in den Archiven ist mit der Überlieferung von Versicherungsbüchern und später Versicherungskarten eine Archivreihe entstanden, die den kantonalen Immobilienbestand⁷ für 200 Jahre vollumfänglich dokumentiert.

Wollishofen Nr. 1

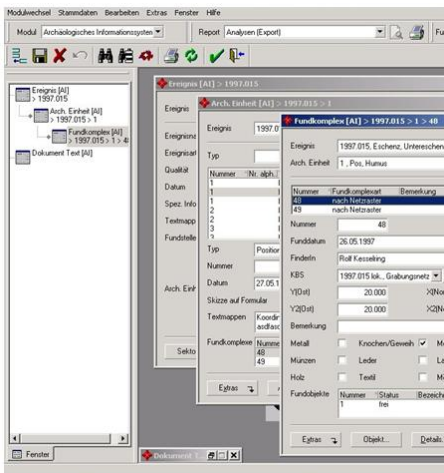
Im Jahr	Namen des Eigentümers.	Bestimmung des Gebäudes.	Quantität						Wohnungswert nach Tabelle.	Zählerwert nach Tabelle.
			Stk.	Pa.	Stk.	Pa.	Stk.	Pa.		
1112	Conr. Luchter	Ein doppel Holzhaus	2	1					1	8000
1118	Filippine v. Wang Rindl	1. "								8000
1122	"	1. "								9000
1181	Schlunstein Ulrich Sauer	1. " (2. " 1. Holzschuppenbau)	1	1					9000	8000
1354	Jacob Staub	1. doppel Wohnhaus 1. Holzschuppenbau	1	1					13200	13068
1355	"	1. doppel Holzschuppenbau	1	1						2000
1415	Scheller	1. " 1. "								24000
1475	"	1. " 1. "								30000
1491	"	1. " 1. "								35000

Brandregister Wollishofen (ZH) 1812 / Mikrofilm

Mit der Einführung der elektronischen Datenverarbeitung in den Versicherungen Mitte der neunziger Jahre des letzten Jahrhunderts

⁷ Die Unterlagen der Gebäudeversicherung erlauben die Entwicklung der Bebauung einzelner Parzellen zu verfolgen, wohingegen das Grundbuch oder Kataster nur die Eigentumsverhältnisse widerspiegelt.

wurde diese Überlieferungsmöglichkeit unterbrochen. Die Daten zu den versicherten Immobilien werden formularorientiert erfasst und datenbanktechnisch gespeichert. Aus den erfassten Daten können Datenblätter zu einzelnen Gebäuden, aber auch Listen und Rechnungen, Mailings und vieles mehr generiert werden. Zu jedem Gebäude wird zusätzlich ein Dossier in Form einer Mappe mit den im Verkehr anfallenden Unterlagen (Rechnungen, Schätzungen, Schadensmeldungen etc.) geführt.



Gebäudeblatt

CDL GEMÄSSERBUNG
DES KANTONS LUZERN

Gemeinde: Luzern
Lageadresse: Stadtchliesse 27a
Grossfläche: 728, 730
Umweltzone: Staatsarchiv

Versicherungs-Nr.: 101.2396
Baujahr: 1994
Zweck: 102 Museum, GÄHRN, BÜRO/NE

Eigentümer: Kanton Luzern Dienststelle Immobilien
Stadtchliesse 4, Postfach, 6002 Luzern

Schutzangrund: Revision: 03.10.2007
Schutzungsdatum: Schutzungsgegenstand: Rüstet Peter

Gebäudebeschreibung

Tragwerk

- Stahnerk. Beten
- Mauerwerk
- Massiv
- Holz
- Stahl

Fassaden

- Sichtmauerwerk
- Klinker
- Holz
- Eisen
- Sichtmauer/Beton
- Glas
- Kunststein

Decken

- massiv
- Holzkonstruktion F-30 verkleidet
- Holzkonstruktion sichtbar

Böden

- Beton
- Dielen
- Parkett
- Blech
- Glas
- Klinker

Heizung 1

- Zentralheizung, Wärmepumpe
- Fernheizung
- Zentralheizung
- Wärmepumpe

Heizung 2

- Sonnenkollektoren
- Chemie
- Kachelboden

Bausatz

- massiv
- massiv
- Holzmassiv

Bemerkung

Staatsarchivgebäude mit Natursteinfassaden, 1 Hausverwahrung im DG, AEH für 11 PW im LG

Nützung

Abstellplatz in AEH: ja nein
Bettplatz in Hotelzimmer: ja nein
Betten: ja nein
Betriebsliche Ladenlokale/Restaurants: ja nein
Grossvertriebsstellen in Betrieben: ja nein

Bemerkung und Art der gewöhnlichen Nützung / Lagerort

Bibliothek, Archivräume Staat Luzern

Feuerschutz

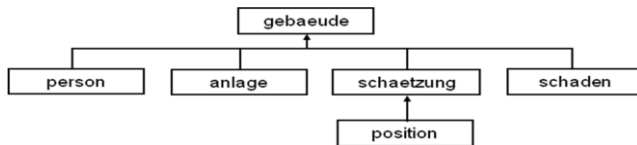
Hydrant: ja nein
Lichtzeichen: ja nein
Feuerpostenringel: ja nein
Distanz zu Gebäude-Nr. (F-F 0-0): 2397, F-F 0-0
Brandzauer: ja nein
Sprinkler: ja nein
Blitzschutz: ja nein
Brandmelder: ja nein
Mittig in nicht-massivem Raum: ja nein

Druckdatum: 24.03.2009 Seite 1 / 3

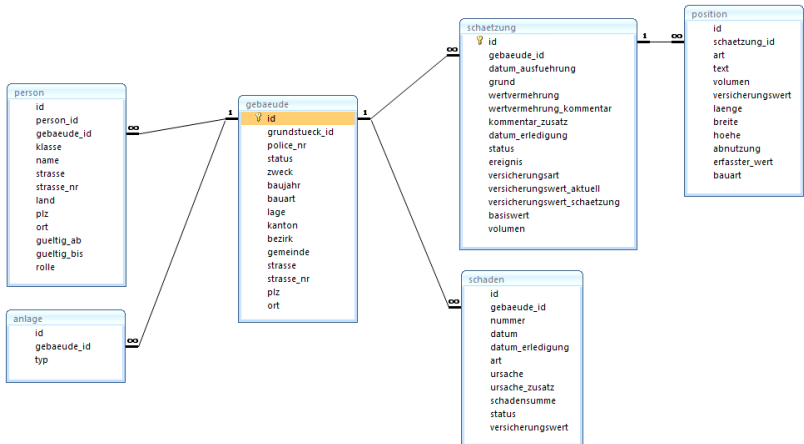
Fachanwendung 1998 und Gebäudeblatt 2009 Kanton Luzern

Die Bewertung einer Gruppe von kantonalen Staatsarchiven führt 2005 zum Entscheid, Dossiers in Auswahl und die Registerdaten integral zu übernehmen. Die Bewertung führt aber zweitens dazu, von den in der Fachanwendung gepflegten Registerdaten nur einen Teil, den als

archivwürdig betrachteten, zu archivieren. Daraus ist ein reduziertes Datenmodell für Registerdaten der Gebäudeversicherung entstanden. Es umfasst eine vollständige Gebäudeliste mit Versicherungswert zum Stichtag mit Beziehung zu Besitzer/Verwalter, der periodischen Schätzung in Einzelpositionen und der Liste der Schäden. Nicht enthalten sind alle Informationen zum Rechnungswesen.



Datenmodell der bewerteten Registerdaten



Detailliertes Datenmodell der Gebäudeversicherungs-Registerdaten

Ein Datenexport aus der Fachanwendung erzeugt diese reduzierte archivische Datensicht für die Archivierung in SIARD, einem archivauglichen Format zur Archivierung von strukturierten Daten. SIARD ist das

für Datenbanken empfohlene Archivformat des Schweizerischen Bundesarchivs⁸.

Polizeirapport und Journal

Bei diesem Archivierungsprojekt aus dem KOST Beratungsportfolio geht es um die Archivierung von Polizeirapporten, detaillierten Formularen zu jedem Vorfall und dem Polizei-Journal, dem Hauptbuch, in dem alle Vorfälle aufgezeichnet werden. Provenienz ist die Kantonspolizei. Rapport und Journaleintrag müssen, nach einer Verjährungsfrist, aus rechtlichen und Datenschutzgründen vollständig gelöscht werden⁹. Seit Ende neunziger Jahre werden diese Daten bei den meisten Polizeicorps in Fachanwendungen gehalten. Damit durch die notwendigen Löschläufe nicht wertvolle zeitgeschichtliche Dokumente verloren gehen, muss ein Weg gefunden werden, wie diese Daten dem Staatsarchiv übergeben werden können.

Die eingesetzten Fachanwendungen unterscheiden sich technisch/inhaltlich nicht gross, sind aber in der Regel Eigenentwicklungen. Die Erfassung eines Rapports erfolgt über ein Datenbankformular, der Journaleintrag, als Exzerpt des Rapports, wird aus den Datenbankdaten automatisch erzeugt. Auf Grund der Sensibilität der involvierten (Polizei)Stellen müssen wir leider auf die konkrete Darstellung von

⁸ Seit 2008 stellt das Schweizerische Bundesarchiv die Formatspezifikation und Tools zum Generieren und Lesen von SIARD Dateien kostenlos zur Verfügung. Die KOST bietet seit 2010 ein Open Source Tool zum Konvertieren von CSV Dateien in SIARD an: <http://www.bar.admin.ch/dienstleistungen/00823/00825/index.html?lang=de>
http://kost-ceco.ch/cms/index.php?csv2siard_de

⁹ Diese Verjährungsfristen bewegen sich zwischen 5 und 30 Jahren.

Screenshots der Fachanwendung verzichten, grundsätzlich unterscheidet sich diese aber nicht von der Gebäudeversicherungslösung¹⁰.

In diesem Falle wurde von zwei Staatsarchiven, die sich schon mit dem Problem beschäftigt haben, zusammen mit der KOST eine Lösung gewählt, die sich am *Records Management Modell* orientiert. In dieser Sicht besteht ein Aktenplan mit einer sehr einfachen Hierarchie aufgebaut aus Organisationseinheit (Polizeicorps) > Jahr > Monat > Tag > Dossier. Als einziges Dokument liegt ein Polizeirapport im Dossier. Zu einem späteren Zeitpunkt ist geplant, noch weitere Dokumente zum jeweiligen Vorfall (Fotos, Office-Dokumente etc.) im Dossier zu speichern, heute werden weitere Fallunterlagen in einem Papierdossier in einer Handablage verwaltet. Der Aktenplan ist für den Bearbeiter als Baumstruktur links im Erfassungsformular sichtbar. Für die Archivierung wird dieser Aktenplan zusammen mit je einem XML- und PDF-Dokument für den jeweiligen Polizeirapport generiert. Das XML-Dokument wird speziell für die Archivierung generiert, damit eine maschinenlesbare Version der Daten längerfristig verfügbar ist. Das PDF-Dokument entspricht dem ausgedruckten Rapport.

Da es sich um sehr sensible Daten und eine grosse Datenmenge handelt, wird nur ein Sample übernommen. Im Pilotprojekt des Staatsarchivs Zürich werden Journaldaten in einem Sample von zwölf Journaltagen pro Jahr und jeweils ein vollständiger Rapport pro Tag aus der Tagessammlung übernommen.

¹⁰ Erfassungsformular, Polizeirapport und Journalliste entsprechen etwa dem Screenshot *Fachanwendung Gebäudeversicherung 1998* weiter oben.

- Journaldirectory:
 - enthält die 12 Tagesjournale als XML-Files
 - enthält die 12 Tagesjournale als HTM-Files
 - enthält die 12 Tagesjournale als PDF/A-Files (erzeugt in der Jahresendverarbeitung)
- 365 resp. 366 Tagessammlungen mit
 - den Registerdaten als XML-File
 - den Registerdaten als HTM-File
 - den Registerdaten als PDF/A-File (erzeugt in der Jahresendverarbeitung)
 - je ein Directory pro ausgewählte Geschäftsgruppe
 - XML-File der Geschäftsgruppe (Extrakt aus Registerdaten)
 - HTM-File der Geschäftsgruppe
 - PDF-File der Geschäftsgruppe
 - PDF/A pro
 - Rapport (Basis HTML)
 - gescannte Beilagen und Justizakten (Basis TIFF, ev. JPG in TIFF)
 - elektronische Beilagen (Originale und PDF/A soweit Konvertierungstools eingesetzt werden)

Die vereinfachte Ablagestruktur der Jahressammlung

Die Einrückstufen zeigen die hierarchische Ordnung innerhalb der Jahressammlung.

Fazit

Wir haben in den beiden betrachteten Fällen eine Datenreduktion durch einen entsprechenden Bewertungsentscheid erreicht; in einem Fall durch ein zeitliches Sampling, im anderen Fall durch die Selektion weniger Entitäten aus dem gesamten Datenmodell:

	Entitäten					
Zeit						

Matrixdarstellung auf der Zeit / Entitäten-Achse

Wir haben aber auch gesehen, dass die angewendete Abstraktion eine Vereinheitlichung der Archivlösung erlaubt. Im einen Fall durch die

Abbildung der Fachanwendung auf ein *Records Management* Modell, im andern Fall durch eine Reduktion auf das Datenmodell *Registerdaten*. Dass es sich um eine verallgemeinernde Modellierung handelt, wird in beiden Beispielen auch dadurch klar ersichtlich, dass die Modelle ausgetauscht werden können. Sowohl *Polizeirapport* wie auch *Gebäudeversicherung* könnten als Register-Fachanwendung wie auch als *Records Management* System gesehen und so archiviert werden.

Aus heutiger Sicht scheint die Frage, welche Modellierung besser und für die Archivierung zukunftssträchtiger ist, nicht abschliessend beantwortet werden zu können. Vielmehr ist wohl auch diese Frage nach der Sicht auf die Daten eine Frage der Bewertung.