

# **Elektronische Archivierung und Nutzung (ELAN)**

## **Eine Projektidee zur Entwicklung von Verfahren für die dauerhafte Archivierung und Nutzung von elektronischen Geschäftsunterlagen aus Behörden und Betrieben**

*Andreas Engel*<sup>1</sup>

Mit der zunehmenden Verbreitung elektronischer Geschäftsunterlagen steigt auch der Bedarf an Verfahrenslösungen für die dauerhafte Aufbewahrung überlieferungswürdiger Unterlagen. Weil aber der organisatorische und technische Aufwand sehr hoch ist, stehen vor allem kleinere und mittelgroße Organisationen (Behörden wie Betriebe) vor dem Problem, entsprechende organisatorisch-technische Infrastrukturen aufzubauen und bereitzustellen. Selbst kleinere Archive dürften Schwierigkeiten haben, diesen Aufwand aus eigener Kraft zu leisten, so dass auch sie auf entsprechende Dienstleistungen für die langfristige Aufbewahrung elektronischer Unterlagen ("Langzeitarchivierung") angewiesen sind.

Es ist seit jeher Aufgabe der Archive, aufbewahrungswerte Unterlagen auf Dauer zu sichern, zu erschließen und für die Nutzung bereitzustellen. Für Behörden und Betriebe, die ihre Unterlagen einem Archiv übergeben, sind praktikable und zuverlässige Verfahren etabliert, um konventionelle Geschäftsunterlagen zu übernehmen und dauerhaft der Benutzung zugänglich zu machen. Es fehlen bisher noch Verfahren, mit denen auch elektronische Unterlagen aller Art - aus Akten oder Datenbanken - dauerhaft aufbewahrt werden können. Darüber hinaus besteht bei Organisationen, die nicht auf die Dienstleistungen eines Archivs zurückgreifen können, ein Bedarf zur langfristigen Aufbewahrung wichtiger elektronischer Geschäftsunterlagen. Hier ist der Sachverstand der Archive gefragt, um auch für diese Anwender zuverlässige Verfahren und spezielle Dienstleistungsangebote zu entwickeln.

In Fortsetzung der Aktivitäten zur Entwicklung von Verfahren zur Aussonderung elektronischer Akten in der Bundesverwaltung<sup>2</sup> plant daher die Forschungsstelle für Verwaltungsinformatik der Universität Koblenz-Landau gemeinsam mit dem Bundesarchiv und Anwendern aus dem öffentlichen und privatwirtschaftlichen Bereich ein Forschungsprojekt, das speziell für kleinere Organisationen Verfahrenslösungen der dauerhaften Aufbewahrung elektronischer Unterlagen untersuchen soll. Für diese Organisationen sollen Modelllösungen entwickelt, in Form von Referenzinstallationen auf ihre Praktikabilität überprüft, erprobt und bewertet werden. Dabei wird sowohl an Lösungen gedacht, die in den jeweiligen Organisationen verwirklicht werden können als auch an die Entwicklung von neuen Dienstleistungsangeboten für solche Organisationen, die den notwendigen technischen und administrativen Aufwand zur Gewährleistung der dauerhaften Aufbewahrung und Nutzung nicht aus eigener Kraft leisten können.

---

<sup>1</sup> Der vorliegende Beitrag baut auf einer Projektskizze auf, die vom Autor in Zusammenarbeit mit den Projektpartnern entwickelt wurde. An der Erstellung der Projektskizze waren beteiligt Frau Bettina Martin-Weber und Dr. Michael Wettengel (Bundesarchiv Koblenz). Das Projekt befindet sich in der Antragsphase. Im vorliegenden Beitrag werden die Projektziele und erste Überlegungen zum Realisierungsansatz erläutert.

<sup>2</sup> Vgl. KBSt (Koordinierungs- und Beratungsstelle der Bundesregierung für Informationstechnik in der Bundesverwaltung) (Hrsg.): Konzept zur Aussonderung elektronischer Akten, Teil 1: Empfehlung des Bundesarchivs zur Aussonderung elektronischer Akten, Teil 2: Erfahrungen zum Aufbau und zur Ablage elektronischer Akten im DOMEA<sup>®</sup>-Projekt (Schriftenreihe der KBSt, Bd. 40), Bonn 1998

Im vorliegenden Beitrag wird schwerpunktmäßig auf das zur prototypischen Entwicklung und Erprobung vorgesehene softwaretechnische Verfahren der dauerhaften Archivierung mit XML (eXtensible Markup Language) eingegangen.

## **1 Projektziele und Beteiligte**

Ziel des Projekts ist die Entwicklung organisatorisch-technischer Verfahren zur dauerhaften Archivierung elektronischer Unterlagen. Im Mittelpunkt sollen prozessgenerierte Unterlagen stehen, wie sie im Verlauf von Geschäfts- und Bearbeitungsprozessen in elektronischer Form entstehen. Dabei kann es sich sowohl um digitale Dokumente, Vorgänge und Akten handeln als auch um Informationen aus Datenbanken. Verfahren zur dauerhaften Archivierung elektronischer Geschäftsunterlagen müssen prinzipiell beide Speicher- und Verarbeitungsformen für Geschäftsunterlagen berücksichtigen. Nicht selten kommt es zudem vor, dass zu ein und demselben Geschäfts- oder Verwaltungsprozess sowohl Informationen in elektronischen Akten als auch in Fachdatenbanken vorliegen. Auch diese Anwendungssituation der Integration unterschiedlicher elektronischer Speicherobjekte und -formate in einer aus fachlicher Sicht als Ganzes zu betrachtenden Überlieferungseinheit ist bei der Aussonderung und dauerhaften Archivierung elektronischer Unterlagen zu berücksichtigen.

Bei der Entwicklung eines Verfahrens zur dauerhaften Archivierung ist der gesamte Lebenszyklus elektronischer Unterlagen zu betrachten und die Auswirkungen organisatorisch-technischer Verfahrensentscheidungen auf alle Phasen zu untersuchen. Da schon das Entstehen überlieferungswürdiger Unterlagen entsprechende organisatorische Regelungen und Praktiken voraussetzt, ist zu klären, welche Anforderungen bereits an die elektronische Bearbeitung zu stellen sind, damit die Voraussetzungen für die Bewertung elektronischer Unterlagen, ihre elektronische Aussonderung, Erschließung, dauerhafte Aufbewahrung und Nutzung gewährleistet sind. Die Entwicklung eines Archivierungsverfahrens kann somit nicht auf einen rein informationstechnischen Lösungsansatz beschränkt bleiben.

Für Organisationen, die nicht auf Archive zurückgreifen können, stellt sich zudem die Frage, unter welchen Bedingungen der Aufbau einer eigenen organisatorisch-technischen Infrastruktur zur dauerhaften Aufbewahrung elektronischer Unterlagen empfohlen werden kann bzw. wann auf externe Dienstleistungsangebote zurückgegriffen werden sollte. Dies impliziert, dass Lösungen für die Entwicklung externer Dienstleistungsangebote zur dauerhaften Aufbewahrung elektronischer Unterlagen untersucht und erprobt werden. Das Projekt soll dazu einen Beitrag leisten, indem es beispielhaft untersucht, welche organisatorischen und technischen Voraussetzungen geschaffen werden müssen, um ein Dienstleistungsangebot für die dauerhafte Archivierung elektronischer Unterlagen zu realisieren.

Die entwickelten Verfahren sollen schließlich in Form von Referenzinstallationen bei den Anwendern installiert und unter Kriterien der Wirtschaftlichkeit, Sicherheit und Vertrauenswürdigkeit erprobt werden. Ergebnisse des Projekts sollen Empfehlungen zum Aufbau einer informationstechnischen Infrastruktur zur elektronischen Archivierung sein mit Aussagen über die zu verwendenden Speicherformate und Konvertierungsstrategien, Hinweise zu organisatorischen Maßnahmen und zur Ausgestaltung von Dienstleistungsangeboten.

Die Anwender für das Projekt wurden danach ausgewählt, dass bei allen bereits im nennenswerten Umfang elektronische Unterlagen im Geschäftsprozess entstehen. Weiterhin wurde darauf geachtet, dass ein Bedarf an elektronischer Archivierung besteht, der repräsentativ für eine größere Gruppe von Anwendern ist. Schließlich sollten die Projektanwender möglichst unterschiedliche Anwendungsbereiche bzw. Anforderungsprofile abdecken. Vor diesem Hintergrund haben der Deutsche Industrie- und Handelstag (DIHT) mit einer noch auszuwählenden Industrie- und Handelskammer (IHK) sowie die Stadt Mannheim, vertreten durch das Stadtarchiv Mannheim, ihre Bereitschaft erklärt, als Anwender im Projekt mitzuwirken.

Als Körperschaften des öffentlichen Rechts verwahren IHK Unterlagen aus öffentlich-rechtlicher Aufgabenwahrnehmung, so vor allem Unterlagen über Lehr- und Ausbildungsabschlüsse sowie Gutachten und Berichte aus der Beratungstätigkeit für Behörden. In dem geplanten Projekt soll am Beispiel der dauerhaften Archivierung der Berufsausbildungsverträge ein Modell für ein externes Dienstleistungsangebot entwickelt und erprobt werden, bei dem das Bundesarchiv als Dienstleistungsanbieter auftritt.

Als Pilotpartner aus dem kommunalen Bereich hat sich die Stadt Mannheim mit dem Stadtarchiv Mannheim zur Mitarbeit als Anwendungspartner bereit erklärt. Das Stadtarchiv Mannheim nimmt für die Stadt sowohl die Aufgabe eines Zwischen- und Endarchivs wahr und erbringt in dieser Funktion eine archivische Dienstleistung für städtische Institutionen. Als mögliche Anwendungsfelder für die Erprobung von elektronischen Archivierungsverfahren sind Einwohnermeldedaten sowie elektronisch aufbereitete Mieterdaten der GBG Mannheimer Wohnungsbaugesellschaft vorgesehen.

## 2 Anforderungen an Verfahren zur dauerhaften Archivierung

Für die dauerhafte Archivierung elektronischer Unterlagen werden derzeit zwei Verfahren diskutiert: das *Emulationsverfahren*, bei dem die dauerhafte Aufbewahrung und Nutzung des elektronischen Archivguts in einer Systemumgebung erfolgt, mit der das Anwendungssystem im Nutzungszusammenhang emuliert wird. Alternativ dazu wird das Verfahren der Archivierung durch *Migration* und Speicherung der Daten in Standardformaten vorgeschlagen. Die Vor- und Nachteile beider Verfahren wurden bereits ausführlich erörtert.<sup>3</sup> Im Ergebnis kann festgehalten werden, „dass die Orientierung an Standardformaten und die Einbeziehung von Migrationsstrategien in Archiven derzeit die tragenden Konzepte für die Archivierung von Unterlagen aus digitalen Systemen darstellen.“ Doch es bleibt festzuhalten, dass es sich bei der Präferenz für die Migrationsstrategie „vor allem um eine theoretische Einsicht handelt, die in der Praxis erst noch realisiert werden muss.“<sup>4</sup>

Die Migrationsstrategie baut auf der Annahme auf, dass auf Dauer weder ein allgemein gültiges Speicherformat verfügbar sein wird, noch eine Systemumgebung bereitgestellt werden kann, mit der einmal archivierte Unterlagen in ihrem ursprünglichen Speicherformat

---

<sup>3</sup> Frank M. Bischoff: Emulation - das Archivierungskonzept der Zukunft? In: Michael Wettengel (Hrsg.): Digitale Herausforderungen für Archive (Materialien aus dem Bundesarchiv, Heft 7), Koblenz 1999, S. 15-23.

<sup>4</sup> Ebenda, S. 16

auf unbestimmte Zeit zugänglich gemacht werden können.<sup>5</sup> In dieser Situation empfiehlt die Migrationsstrategie zwei Lösungsansätze: Zum einen sollen zu archivierende elektronische Unterlagen aus gleichartigen Anwendungszusammenhängen, in denen unterschiedliche Bearbeitungsformate (Anwendungsformate) zur Anwendung kommen, für die dauerhafte Aufbewahrung und Nutzung in ein standardisiertes Archivierungsformat übertragen werden. Auf diese Weise soll sichergestellt werden, dass ein möglichst langfristiger Zugriff auf elektronisches Archivgut mit der vorhandenen Systemtechnik gewährleistet bleibt.

Da aber gleichzeitig nicht vorausgesetzt werden kann, dass auch Standardformate auf Dauer Bestand haben, kommt als weitere Bedingung an ein Archivierungsformat hinzu, dass es möglichst leicht und verlustfrei in beliebige, noch nicht bekannte Speicherformate konvertiert werden kann. Diese zweite Anforderung schließt ein, dass Archivierungsformate möglichst selbsterklärend in dem Sinn sind, dass sie alle Informationen enthalten, die für eine verlustfreie, automatische Konvertierung notwendig sind. Sie müssen also vollständig spezifiziert und dokumentiert sein, damit die Voraussetzungen für eine programmgestützte, automatische Konvertierung erfüllt sind.

Eine weitere Anforderung an Archivierungsformate ist die Möglichkeit zur Beschreibung und Bildung komplexer Archivobjekte. Unter einem *Archivobjekt* soll ein abstrakter Datentyp verstanden werden, der sämtliche Informationen zu einer Überlieferungseinheit in einem elektronischen Objekt zusammenfasst. So enthalten beispielsweise Archivobjekte:

- Ordnungsinformationen des Archivsystems, in dem das Archivobjekt aufbewahrt wird (z.B. Archivnummer, Datenträger, Speicherformat)
- Informationen zum Ablauf des Verfahrens der Bewertung, Aussonderung und Archivierung (z.B. die Dokumentation des Bewertungsverfahrens)
- Ordnungsinformationen aus dem Anwendungssystem, aus dem Archivobjekte ausgesondert wurden (Ordnungs- und Klassifikationsmerkmale)
- Informationen aus dem Nutzungszusammenhang des Anwendungssystems, in dem Archivobjekte entstanden sind (Bearbeitungs- und Kontextinformationen)
- Informationen über den Sachverhalt, der im Anwendungszusammenhang bearbeitet wurde (Primärinformationen).

Da in elektronischen Systemen diese verschiedenen, zu archivierenden Informationen oft in unterschiedlichen Speicherobjekten abgebildet werden (z.B. in elektronischen Dokumenten, Vorgängen und Akten oder als Abbildung eines Sachverhalts in verschiedenen Tabellen einer Datenbank), kommt es darauf an, getrennt abgespeicherte elektronische Informationen zu einer Überlieferungseinheit zum Zweck der Archivierung in einem elektronischen Archivobjekt zusammenzuführen.

Die zu archivierenden Informationen können darüber hinaus in verschiedenen Datei- und Zeichenformaten vorliegen, so dass ein Archivierungsformat prinzipiell auch verschiedene Formate integrieren muss. Schließlich soll die Nutzung archivierter Informationen unabhängig vom Anwendungssystem möglich sein, um eine langfristige Aufbewahrung in einer Archivierungsumgebung zu garantieren.

---

<sup>5</sup> Vgl. Ivar Fønnes: Methoden zur Langzeiterhaltung elektronischer Informationen, in: Udo Schäfer; Nicole Bickhoff (Hrsgg.): Archivierung elektronischer Unterlagen (Werkhefte der staatlichen Archivverwaltung Baden-Württemberg A 13), Stuttgart 1999, S. 213-222, hier S. 214 f.

### 3 Ein XML-basierter Lösungsansatz zur dauerhaften Archivierung elektronischer Unterlagen

Um die beschriebenen Anforderungen zu erfüllen, wird ein Archivierungsverfahren vorgeschlagen, das auf XML (eXtensible Markup Language) als Beschreibungssprache für Archivobjekte aufbaut. **Abbildung 1** verdeutlicht, über welche Konvertierungsschritte das Zielformat für die elektronische Archivierung erreicht werden kann.

#### 3.1 Anwendungsformate als Ausgangspunkt

Ausgangspunkt des Verfahrens ist ein Datenbestand, der in einem anwendungsspezifischen Speicherformat vorliegt und zur Aussonderung ansteht. Bei Anwendungsdaten handelt es sich entweder um Fachdaten, die häufig in Datenbanksystemen verwaltet werden, oder um elektronische Akten, die jeweils anwendungsspezifische Dateiformate aufweisen.

Die Bewertung der zu archivierenden Unterlagen geschieht in der Regel auf der Basis der Daten im Anwendungsformat. Die informationstechnische Unterstützung des Bewertungsverfahrens kann dabei entweder im Anwendungszusammenhang erfolgen, möglicherweise sogar mit Modulen des Anwendungssystems, oder extern auf der Basis einer elektronischen Aussonderungsliste, die zum Zweck der Bewertung erstellt wird und alle für die Bewertung relevanten Informationen enthält.<sup>6</sup>

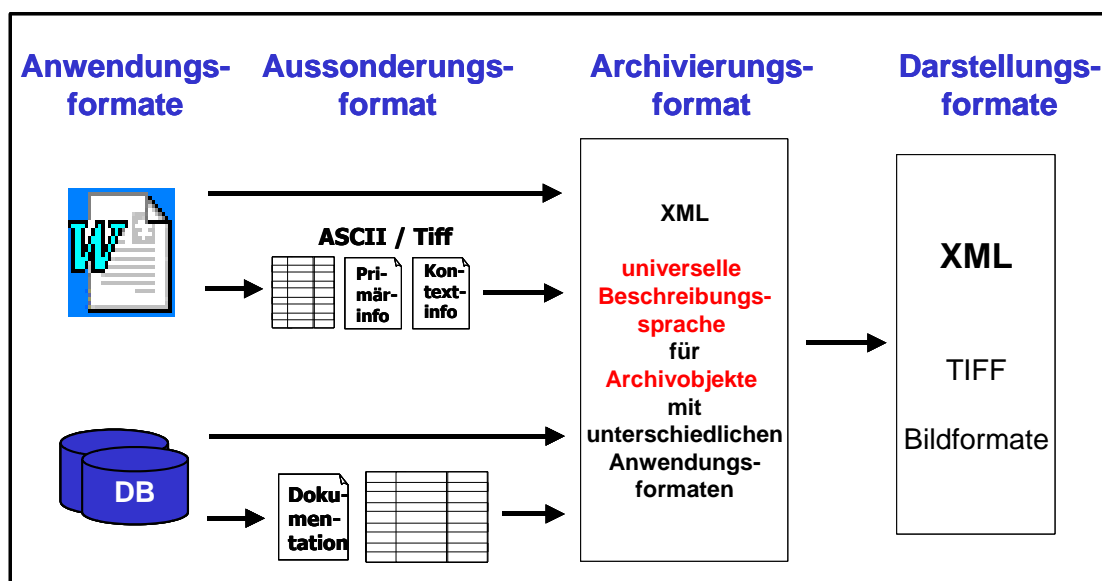


Abbildung 1: Konvertierungsschritte vom Anwendungsformat zum Archivierungs- und Darstellungsformat

#### 3.2 Anwendungsspezifische Aussonderungsformate als Zwischenschritt

Ergebnis der Bewertung ist eine Auswahl von Anwendungsdaten, die in ein Archiv übernommen werden sollen und zu diesem Zweck in ein für die Aussonderung vorgesehenes

<sup>6</sup> Vgl. KBSt: Konzept zur Aussonderung (wie Anm. 2), S. 43. ff

Format übertragen werden müssen (Aussonderungsformat). Dementsprechend sind für das Aussonderungsverfahren Datenstrukturen und Dateiformate festzulegen. In erster Linie sind zu definieren:

- Formate für Ordnungsinformationen (Metainformationen), die aus dem Anwendungssystem übernommen werden sollen und die auch für die Recherche nach archivierten Informationen zur Verfügung stehen sollen;
- Formate für Informationen über den Sachverhalt (Primärinformationen), die in ein Archivierungsformat übertragen und zur archivischen Nutzung bereitgestellt werden sollen.

Die Schwierigkeit bei der Entscheidung für ein Aussonderungsformat besteht darin, dass mit der Aussonderung ein Dateiformat festgelegt werden muss, das auch die Art der späteren Nutzung determiniert. Da die mögliche Nutzung von Archivgut niemals vollständig antizipiert werden kann, kommt es darauf an, für möglichst viele denkbaren Nutzungsformen offen zu sein.

In Bezug auf die spätere Nutzung ist vor allem zwischen Anforderungen an elektronische Akten und Datenbank-Informationen zu differenzieren. Während bei elektronischen Akten in der Regel die Nutzung auf eine möglichst genaue inhaltliche Recherche (einschließlich Volltextrecherche) und die Darstellung der Informationen in ihrem ursprünglichen Layout beschränkt bleibt, kommen bei Datenbanken weitere Nutzungsmöglichkeiten hinzu. Für archivierte Datenbanken lassen sich prinzipiell drei Nutzungsarten unterscheiden:

- die Recherche und Selektion einzelner Fälle aus einer Datenbank (Archivobjekte) aufgrund bestimmter Einzelmerkmale des Archivobjekts mit der vollständigen Darstellung aller Informationen
- die Recherche (einer Auswahl) von Archivobjekten nach bestimmten Kriterien, die mit einer Teilmenge der erfassten Informationen dargestellt werden sollen
- der Export (einer Teilmenge) archivierter Objekte zum Zweck der weiteren Bearbeitung mit speziellen Auswertungssystemen (Statistikprogrammen, grafischen Informationssystemen etc.)

In Anwendungssystemen, insbesondere in den heute verbreiteten relationalen Datenbanksystemen, werden Datenstrukturen vor allem unter dem Gesichtspunkt einer möglichst redundanzfreien Speicherung von Informationen definiert. Dies hat zur Folge, dass ein „Anwendungsobjekt“ aus Informationen besteht, die in der Datenbank auf verschiedene Tabellen (Relationen) verteilt sind. Erst im Anwendungsprogramm werden diese Informationen zu komplexen Anwendungsobjekten aufbereitet. Insofern liegen in diesen Fällen auch für die Aussonderung quasi nur Rohdaten vor, die zum Zweck der Aufbewahrung und Nutzung im Archiv als Archivierungsobjekte rekonstruiert und neu definiert werden müssen.

Für die Aufbereitung von Fachdaten aus komplexen Anwendungssystemen ist daher im Rahmen der Bewertung, Aussonderung und Aufbereitung zu prüfen, ob die im Anwendungssystem gespeicherten Informationen zu einem Fall bzw. Anwendungsobjekt vollständig mit der in der Bearbeitung vorgefundenen Datenstruktur archiviert werden sollen. Alternativ ist zu prüfen, ob für die dauerhafte Aufbewahrung eine Komplexitätsreduktion der Informations-

struktur vorgenommen werden kann, weil für die Archivierung z.B. auf Hilfsmerkmale verzichtet werden kann, die nur für die Bearbeitung wichtig waren.

Bei der Aussonderung von Datenbankinformationen ist daher die Frage zu klären, ob tatsächlich der gesamte Datenbestand oder nur ein „view“ auf den Bestand der Anwendungsdaten übernommen werden muss. Auch elektronische Akten können u.U. nur in Bezug auf einzelne Schriftstücke überlieferungsrelevant sein, wenn es z.B. nur um den Nachweis von Verträgen geht, der laufende Schriftwechsel zwischen den Vertragspartnern aber nicht aufbewahrt werden müsste. Die elektronische Übernahme erlaubt in diesen Fällen die Selektion bestimmter Informationen aus Überlieferungseinheiten, was bisher aus Praktikabilitätsgründen nicht möglich war.

Als Ergebnis der Aussonderung liegen Daten in einem Aussonderungsformat vor, dass aufgrund der verwendeten Zeichen- und Dateiformate möglichst viele Optionen zur dauerhaften Aufbewahrung offen lässt. Prinzipien für die Auswahl eines Aussonderungsformats sind die Verwendung möglichst verbreiteter Zeichen- und Dateiformate und das Bestreben, für ausgesonderte Unterlagen möglichst „flache“ Informationsstrukturen zu verwenden.<sup>7</sup>

Der Zwischenschritt der Aussonderung in ein spezielles Aussonderungsformat ist derzeit noch notwendig, weil ein generelles Archivierungsformat noch nicht feststeht. Die Konvertierung in ein Aussonderungsformat wird überflüssig, wenn die zu archivierenden Informationen unmittelbar aus dem Anwendungsformat in das Archivierungsformat übernommen werden können.

### 3.3 XML als Format für die dauerhafte Archivierung

Für die dauerhafte Archivierung elektronischer Unterlagen wird XML zur näheren Untersuchung vorgeschlagen. Diese Entscheidung liegt nahe, weil XML als Beschreibungssprache für Informationen entwickelt wurde, die möglichst unabhängig von speziellen Anwendungsprogrammen angezeigt und verarbeitet werden sollen. Dabei stand die Präsentation und weitere Bearbeitung von Informationen im World Wide Web im Vordergrund.

Im Unterschied zur bisher verbreiteten Seitenbeschreibungssprache HTML (HyperText Markup Language) sind die durch XML bereitgestellten Sprachkonstrukte nicht fest vorgegeben. XML besitzt nur einen minimalen Bestand an vordefinierten Sprachelementen<sup>8</sup>, die je nach Anwendungsgebiet um eigene Definitionen erweitert werden können. Allerdings kann XML auch ohne Erweiterungen zur Informationsmodellierung, Verarbeitung, Übermittlung und Speicherung benutzt werden.

---

<sup>7</sup> Vgl. Frank M. Bischoff: Elektronisches Grundbuch in Nordrhein-Westfalen. Möglichkeiten der Überlieferungssicherung aus archivischer Perspektive, in: Schäfer; Bickhoff (Hrsgg.), Archivierung (wie Anm. 5), S. 106; Udo Schäfer: Geographische Informationssysteme in der Landesverwaltung, in: ebd., S. 125 f.; Schärli 1999, S. 208; Michael Wettengel: Archivierung digitaler Datenbestände aus der DDR nach der Wiedervereinigung, in: Schäfer; Bickhoff (Hrsgg.), Archivierung (wie Anm. 5), S. 229; Fonnes, Methoden (wie Anm. 5), S. 219.

<sup>8</sup> Vgl. die Sprachbeschreibung von XML unter <http://www.w3.org/TR/1998/REC-xml-19980210>: „XML 1.0 Recommendation“ vom 10. Februar 1998 des W3C (World Wide Web Consortium). Weitere Informationen sind auf der Web-Seite des W3C unter <http://www.w3.org/Consortium/Legal/copyright-documents.html> zu finden.

Ein weiterer Vorteil von XML besteht darin, dass mit dieser Meta-Sprache Informationen nicht nur formal beschrieben, sondern auch gespeichert werden können. Dies ist einerseits die Voraussetzung dafür, dass XML-Dokumente mit gängigen Web-Browsern angezeigt werden können. Die von bestimmten Anwendungsprogrammen und Dateiformaten unabhängige Speicherung der Informationen erlaubt es aber auch Programme zu entwickeln, mit denen die in XML gespeicherten Informationen je nach Nutzungsanforderungen ausgewertet werden können. Außerdem kann auf diese Weise sichergestellt werden, dass elektronisches Archivgut auf Dauer für programmgestützte, automatische Migrationen und Konvertierungen zugänglich ist. Damit eröffnet sich ein technischer Realisierungsansatz, mit dem die Migrationsstrategie der dauerhaften Archivierung elektronischer Unterlagen auch unter den Bedingungen nicht dauerhafter Standards praktisch umgesetzt werden kann.

Das Hauptziel von XML ist die Modellierung von Informationsstrukturen. Dabei ist zu unterscheiden zwischen der abstrakten Beschreibung eines Schemas für Informationsobjekte und der konkreten, strukturierten Speicherung von Informationsobjekten in XML-Dokumenten. Ein konkretes Archivobjekt (z.B. eine konkrete elektronische Akte oder eine Datenbank) wird als *XML-Dokument* beschrieben und gespeichert.

Aus Sicht der Archivierung besteht ein XML-Dokument im Wesentlichen aus Elementen und Entitäten, mit denen Informationen und deren Beziehungen untereinander beschrieben werden können. Ein *Element* hat einen Namen und wird im XML-Dokument durch eine Start- und Endmarkierung (Start- bzw. End-Tag) von übrigen Informationseinheiten abgegrenzt (vgl. **Beispiel 1**). Durch den Elementnamen und Kommentare zum Element kann die Bedeutung einzelner Elemente im XML-Dokument näher erläutert werden.

In XML werden einfache (atomare) und zusammengesetzte (komplexe) Elemente unterschieden. Komplexe Elemente bestehen aus einfachen und/oder zusammengesetzten Elementen. Auf diese Weise können sequentielle und hierarchische Beziehungen zwischen Elementen definiert und XML-Dokumente als Baumstrukturen dargestellt und verarbeitet werden.

Beispiel 1: Auszug aus einem XML-Dokument

```
<E-Akte Name="IMKA-Sitzungen">
  <A_Deckblatt>
    <Abgeb_Behoerde> BMI </Abgeb_Behoerde>
    <Akt_Stelle> O 1(KBSt)</Akt_Stelle>
    <AZ> 195 052-1/15 </AZ>
    <Akte_Inhalt> Einladungen und Protokolle </Akte_Inhalt>
  </A_Deckblatt>
```

Zu XML-Elementen können Attribute definiert werden (vgl. im Beispiel 1 das Attribut „Name“ zum Element „E-Akte“), so dass auch Elemente des gleichen Typs anhand von Merkmalsausprägungen unterscheidbar bleiben. Über Attributdeklarationen kann die Bearbeitung von XML-Dokumenten auf Teilmengen eingeschränkt werden. Ein Spezialfall ist die Deklaration von Identifier-Attributen, mit denen ein-elementige Mengen gebildet und einzelne XML-Elemente oder XML-Dokumente identifiziert und bearbeitet werden können.



Ein weiteres, für die Langzeitarchivierung wichtiges Sprachelement zur Strukturierung von XML-Dokumenten sind *Entitäten*. Mit Entitäten können Verweise auf ein definiertes Element innerhalb und außerhalb eines XML-Dokuments spezifiziert werden. Interne Entitäten definieren einen Bereich, auf den innerhalb eines XML-Dokuments verwiesen werden kann. An der Verweisstelle wird der Inhalt der referenzierten Entität eingefügt. Externe Entitäten enthalten Elemente, die in einer separaten Datei gespeichert sind und über einen Verweis in das XML-Dokument eingebunden werden. Der entscheidende Vorzug für die Langzeitarchivierung besteht darin, dass mit externen Entitäten die Speicherung einzelner Komponenten eines XML-Archivobjekts auf verschiedene Dateien verteilt werden kann. Dies hat wiederum den Vorteil, dass in ein XML-Archivobjekt auch Dateien mit unterschiedlichen Dateiformaten eingebunden werden können. Auf diese Weise ist es möglich, beispielsweise Meta- und Primärinformationen eines Archivobjekts in einer Datenstruktur zu kapseln, selbst wenn die zugehörigen Informationen in verschiedenen Dateien und Speicherformaten vorliegen (z.B. Charakter- und Graphikformate).

XML bietet darüber hinaus die Möglichkeit, die Informationsstruktur von Objekten in einem Schema, einer sogenannten *Document Type Definition (DTD)*, zu beschreiben. In einer DTD wird festgelegt, welche Elemente, Attribute und Entitäten zur Beschreibung eines XML-Dokuments verwendet werden können bzw. müssen. Sie kann entweder einem XML-Dokument unmittelbar hinzugefügt werden oder als separate Datei abgespeichert werden. Ein XML-Dokument, das den Definitionen einer DTD entspricht, wird gültig genannt. Dokumente, für die kein spezielle DTD definiert wurde, aber die den Syntaxregeln der XML-Spezifikation entsprechen, heißen wohlgeformt. Existiert zu einem XML-Dokument eine DTD, so kann der XML-Parser die Übereinstimmung mit der spezifizierten Informationsstruktur prüfen.

Beispiel 2: Ausschnitt aus einer DTD zur Definition des komplexen Elements E-Akte

```
<!DOCTYPE E-Akte> <!-- Elektronische Akte - DTD Version 1 -->
<!ELEMENT E-Akte (A_Deckblatt, A_Meta, Vorgang+)>

<!ELEMENT A_Deckblatt (Abgeb_Behoerde, Akt_Stelle, AZ, Akte_Inhalt)>

    <!ELEMENT AbgeB_Behoerde (#PCDATA)>
    <!ELEMENT Akt_Stelle (#PCDATA)>
    <!ELEMENT Aktenzeichen (#PCDATA)>
    <!ELEMENT Akte_Inhalt (#PCDATA)>

<!ELEMENT A_Meta (#PCDATA)>
...
```

XML-Dokumente werden in einem Standard-Zeichencode gespeichert. Damit sind XML-Dokumente mit der zugehörigen DTD prinzipiell mit einfachen Texteditoren darstellbar. Das mit XML vorgegebene (logische) Dateiformat ist sowohl unabhängig von dem verwendeten Betriebssystem als auch von speziellen Anwendungsprogrammen. Es kann mit allen gängigen Browsern angezeigt werden.

Für die Langzeitarchivierung ist von entscheidender Bedeutung, dass mit einer DTD die Struktur von Archivobjekten vollständig spezifiziert werden kann. Damit sind die Voraus-

setzungen geschaffen, in XML gespeicherte Informationen ohne Informationsverluste in andere Formate zu konvertieren. Selbst die spätere Konvertierung in noch nicht bekannte Formate ist auf diese Weise prinzipiell gewährleistet. Genauso ist es möglich, für die beschriebenen Datenstrukturen auch nachträglich Auswertungsprogramme zu entwickeln, mit denen die Nutzung von XML-definierten Archivobjekten über die bloße Anzeige hinaus unterstützt werden kann.

Schließlich können die mit einem XML-Schema erfassten Archivobjekte in konventionellen Datenbanken gespeichert und verwaltet werden. Aufgrund der geringen Anforderungen an das Speicherformat bietet sich XML auch als universelles Datenaustauschformat an.

### 3.4 Darstellungsformate

In XML beschriebene Archivobjekte können - wie bereits erwähnt - mit XML-fähigen Browsern angezeigt werden. Damit ist prinzipiell ein plattform-unabhängiger Zugriff auf Archivobjekte gewährleistet. Zur benutzerfreundlichen Aufbereitung der Archivinformationen ist es ratsam, spezielle Style-sheets zu definieren, mit denen Archivobjekte angemessener präsentiert werden können. Daneben lassen sich XML-Dokumente in HTML-Objekte integrieren oder einzelne Komponenten eines XML-Dokuments als externe Entities auch im Ursprungsformat von speziellen Anwendungsprogrammen anzeigen lassen. Die Darstellungsmöglichkeiten sind bei Verwendung von XML sehr vielfältig.

## 4 Überblick zur softwaretechnischen Realisierung der dauerhaften Archivierung mit XML

### 4.1 Elektronische Akten als Beispielfall

Die softwaretechnische Realisierung eines Archivierungsverfahrens mit XML soll im Folgenden am Beispiel elektronischer Akten erläutert werden. Das Beispiel orientiert sich an dem im DOMEA-Aussonderungskonzept beschriebenen Verfahren.<sup>9</sup>

Elektronische Akten sind demnach komplexe Archivobjekte, bestehend aus Akten, Vorgängen und Dokumenten. *Akten* als Aussonderungs- und Archivierungseinheit bestehen aus mehreren Elementen: dem Deckblatt, Metainformationen und einer Liste von Vorgängen, die zur Akte angelegt worden sind. Das „Aktendeckblatt“ enthält als ASCII-Datensatz die Ordnungsinformationen der Archivierungsumgebung, einschließlich der Informationen über den Aussonderungsprozess (Zeitpunkt der Aussonderung etc.). Hinzu kommen anwendungsbezogene Metainformationen, die das Ordnungssystem der Akten im Entstehungs- und Anwendungszusammenhang abbilden. Schließlich gehört zum Aktenobjekt eine Liste der in der Akte enthaltenen Vorgänge.

*Vorgänge* besitzen als Elemente die überlieferten Metainformationen des Vorgangs, eine Liste der zum Vorgang gehörenden Dokumente sowie die Kontextinformationen, die im Verwendungszusammenhang entstanden sind und Auskunft über den im Vorgang abgebildeten Entscheidungsprozess geben. *Dokumente* als Elemente eines Vorgangs bestehen aus Meta-

<sup>9</sup> Vgl. KBSt (Hrsg.), Konzept (wie Anm. 2), Teil 1.

und Primärinformationen. Primärinformationen sind nach dem DOMEA-Aussonderungskonzept im TIFF-Format gespeicherte Schriftstücke, die dementsprechend layoutresistent dargestellt werden können. Metainformationen fassen in einem Datensatz, der im ASCII-Format übergeben wird, die Registraturdaten zum Dokument zusammen.

Elektronische Akten sind demnach sowohl im Anwendungs- wie im Archivierungszusammenhang abstrakte Datentypen, die zur dauerhaften Aufbewahrung und Nutzung in ein XML-Format (DTD) konvertiert werden müssen. **Abbildung 2** gibt eine Übersicht über die XML-Architektur zur Speicherung elektronischer Akten und stellt die verschiedenen Schritte der Konvertierung dar.

#### 4.2 Schritte der softwaretechnischen Realisierung

Um einen Aktenbestand zu archivieren, sind die im Folgenden beschriebenen Schritte notwendig. Dabei wird in der vorliegenden Beschreibung der Schwerpunkt auf die Darstellung der Vorgehensweise gelegt. Auf spezielle Anforderungen der Archivierung elektronischer Akten wird nicht gesondert eingegangen.

Im ersten Schritt der Übernahme eines Aktenbestands aus elektronischen Vorgangsbearbeitungssystemen ist eine XML-Document-Type-Definition (DTD) zu definieren, die den im Aussonderungskonzept beschriebenen Aufbau elektronischer Akten abbildet. In dieser DTD wird die Datenstruktur der Archivobjekte beschrieben.

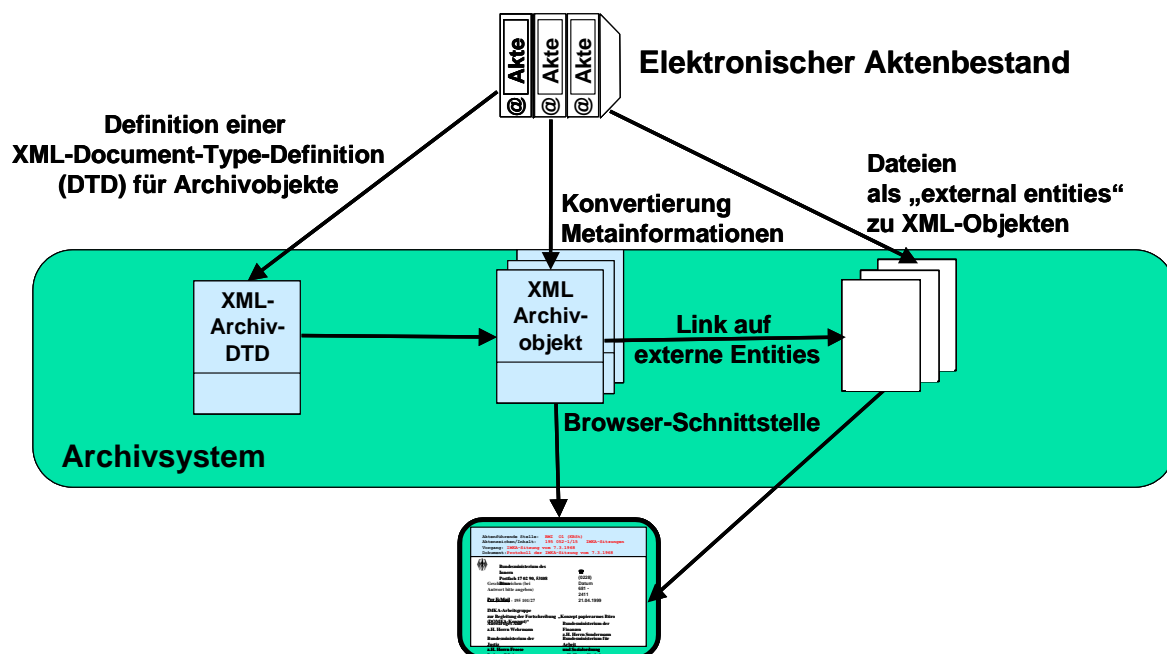


Abbildung 2: Realisierungsansatz für eine Architektur zur Speicherung von Archivobjekten mit XML

Die Definition einer DTD ist Voraussetzung dafür, dass die aus dem Anwendungssystem ausgesonderten elektronischen Akten mit entsprechenden Schnittstellenprogrammen automatisch in XML-Archivobjekte konvertiert werden können (2. Schritt). Die in elektronischen

Akten enthaltenen und im ASCII-Format ausgesonderten Metainformationen werden im ASCII-Format in die entsprechenden Elemente des XML-Dokuments übernommen. Komponenten, die im Grafikformat gespeichert sind (z.B. die Datei der Kontextinformationen oder die Primärinformationen der Dokumente), können entweder als externe Entities deklariert oder mit einer Deklaration der entsprechenden Anzeigeformate in einer Datei mit dem Archivobjekt gespeichert werden.

Nach XML konvertierte Archivobjekte sind nach dem Aussonderungsprozess in eine entsprechende Archiv-Datenbank zu übernehmen. Für die Anzeige der XML-Objekte ist ein XML-Style-Sheet zu entwickeln bzw. ein vorhandenes Style-Sheet zu deklarieren.

## **5 Ausblick**

Mit XML steht eine Beschreibungssprache zur Verfügung, mit der sowohl die logische Struktur (Document Typ Definition) als auch die Speicherstruktur (externe Entities) von Archivobjekten beschrieben werden kann. Die Beschreibung der Informationsstruktur von Archivobjekten ist eine notwendige Bedingung für die automatische Konvertierung von Archivobjekten. Die Möglichkeit zur automatischen Konvertierung ist wiederum eine entscheidende Voraussetzung, um die Aufbewahrung elektronischer Unterlagen dauerhaft zu sichern. Mit XML steht eine Methode zur Verfügung, die diese Anforderungen erfüllt.

XML gewährleistet aber nicht nur die dauerhafte Verfügbarkeit elektronischer Archivobjekte. Es werden auch neue Optionen für das Aussonderungsverfahren eröffnet, insbesondere im Hinblick auf die Bewertung und Auswahl von überlieferungswürdigen Informationen. Mit der elektronischen Archivierung ändert sich folglich nicht nur das Medium der Archivobjekte, auch die bisher angewendeten Prinzipien und praktizierten Verfahren der Bewertung, Aussonderung, Erschließung und Nutzung sind zu überdenken. Die Entwicklung von Verfahren zur dauerhaften Archivierung elektronischer Unterlagen wird daher die Arbeitsweise der Archivare erheblich verändern.