

Möglichkeiten und Grenzen

des Einsatzes von XML für die Archivierung

Ursula Welsch, UWNM
(www.welschmedien.de)

Tagung „Digitales Archiv“
17. Juni 2004

Agenda

- XML-Grundlagen
- Einsatzmöglichkeiten im Archivwesen
- Grenzen des Einsatzes
- Ausblick

XML – Was ist das?

- Extensible Markup Language
W3C Recommendation von 1998
- Medienneutrale Dokumenten-
Beschreibungssprache
- Instrumentarium zur Formulierung von
eigenen Beschreibungssprachen

XML – Technische Grundlagen

Vielseitige Einsatzmöglichkeiten durch:

- Datenformat auf Betriebssystemebene
(keine proprietären Codierungen)
- Strukturbeschreibung im Zeichenformat
(Begrenzer im Datenstrom)
- Frei formulierbare Codierungen
(inhaltsorientierte Namensgebung)

XML – Beispiel

```
<archiv>
  <titel>Deutsche Bahn AG</titel>
  <akte>
    <titel>Teilstrecke München-Holzkirchen</titel>
    <aktenzeichen>1987.54.9</aktenzeichen>
    <provenienz>Bahnhof Darching</provenienz>
    <laufzeit>1.5.1987-30.9.1987</laufzeit>
    <inhalt>Kursbuch, zwei Aushangpläne,
      eine Rolle Fahrkartenvordrucke</inhalt>
  </akte>
</archiv>
```

XML – Stärken

Datenaustausch

- bei der Erarbeitung der Erschließungsdaten
- zum Zweck der Veröffentlichung der Erschließungsdaten, z.B. im Internet
- zum Zweck der Kooperation mit anderen Archiven

Mehrfachnutzung der erarbeiteten Daten

XML-Anwendungen im Archivwesen

- Erschließungsdaten
- elektronischer Dokumente
- Veröffentlichung von Archiv-Inhalten
- Langzeitarchivierung elektronischer Dokumente

Erschließungsdaten

State of the Art =

Erschließung mit Datenbank-Mitteln:

- Abbildung der Gliederungsstruktur oder des Aktenplans als Datenbank-Tabellen

Erschließung mit XML:

- Abbildung der Gliederungsstruktur oder des Aktenplans als Strukturbaum

Erschließungsdaten

XML versus Datenbank:

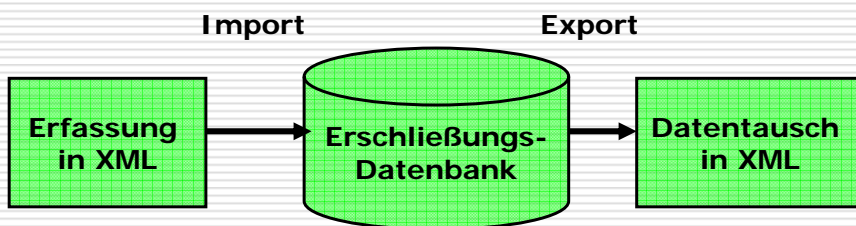
- Etwas flexiblere Strukturen
- Datenaustausch auf Dateiebene ohne Konvertierung möglich

XML als Ergänzung:

- Als Erfassungsformat
- Als Austauschformat

Erschließungsdaten

XML als Ergänzung



Erschließungsdaten

XML-Beispiel:

- Encoded Archival Description (EAD)
- Projekt der University of California, Berkeley unter Mitwirkung der Society of American Archivists
- www.loc.gov/ead/

XML – Beispiel

```
<ead>
  <eadheader>Deutsche Bahn AG</eadheader>
  <archdesc>
    <unititle>Teilstrecke München-Holzkirchen</unititle>
    <unitid>1987.54.9</unitid>
    <origination>Bahnhof Darching</origination>
    <unitdate>1.5.1987-30.9.1987</unitdate>
    <physdesc>Kursbuch, zwei Aushangpläne,
      eine Rolle Fahrkartenvordrucke</physdesc>
  </archdesc>
</ead>
```

Elektronische Dokumente

State of the Art =

Dokumenten-Managementsystem:

- Verwaltung verschiedener Dateiformate in einem Datenbanksystem
- XML ist eines davon

XML-basiertes Content-Managementsystem:

- Abbildung der Datenstrukturen im Datenbanksystem

Elektronische Dokumente

XML versus DMS:

- Dokumentinhalt und zugehörige Erschließungsdaten in demselben Format
- Sicherung der langfristigen Nutzbarkeit durch Migration

Elektronischer Dokumente

XML-Beispiel:

- ❑ Text Encoding Initiative (TEI)
- ❑ Internationale Initiative von Geisteswissenschaftlern
- ❑ www.tei-c.org

XML – Beispiel

```
<tei.4><title>Deutsche Bahn AG</title>
  <teiHeader>
    <titleStmt>
      <title>Teilstrecke München-Holzkirchen</title>
    </titleStmt>
    <extent>Kursbuch, zwei Aushangpläne,
      eine Rolle Fahrkartenvordrucke</extent>
  </teiHeader>
  <text><title>Kursbuch</title>
  <p>...</p></text>
</tei.4>
```

Veröffentlichung von Archiv-Inhalten

Zweck:

- Öffnung für die wissenschaftliche Forschung

Mittel:

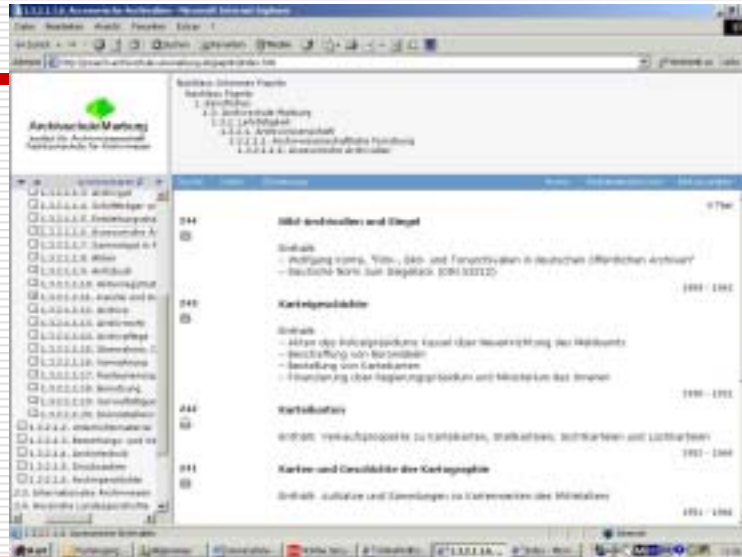
- z.B. Online-Bereitstellung von Findbüchern

Veröffentlichung von Archiv-Inhalten

XML-Beispiel:

- Midosa XML und Midosa Online
- Projekt der Archivschule Marburg
- www.midosas.de

Midosa – Online-Findbuch



Tagung "Digitales Archiv" (© UWNM)

19

Langzeitarchivierung elektronischer Dokumente

Annahme:

In absehbarer Zeit gibt es kein allgemein gültiges Speicherformat, das archivierte Daten auf unbestimmte Zeit zugänglich hält.

Anforderungen an Datenstrukturen:

- selbsterklärendes Format
- Beschreibung und Bildung komplexer Archivobjekte
- Zusammenführung verstreut liegender Informationen
- Nutzung der archivierten Daten unabhängig von Anwendungssystem
- leicht und verlustfrei konvertierbar
- etc.

Tagung "Digitales Archiv" (© UWNM)

20

Langzeitarchivierung elektronischer Dokumente

XML als Migrationsstrategie heisst:

Eine **einzig**e Codierungssystematik

- für die Dokumente einschließlich ihrer Erschließungsdaten (XML)
- für die Beschreibung der Dokumentlogik (XML Schema)
- für die Beschreibung von Grafiken (SVG)
- für die Navigationslogik im Dokument (XPath)
- für Retrieval über die Dokumente (XQuery)
- für die Transformation (XSLT)
- für die seitenorientierte Darstellung (XSL-FO)

Grenzen von XML

XML kann kein Ersatz sein,
für die intellektuelle Arbeit
bei der Bewertung und
Erschließung eines Bestands!

Ausblick

- XML breitet sich weiterhin rasant aus
- kompletter Lifecycle in XML
- zukünftig Regeln für eine automatisierte Archivierung