

Bojan Marinković
(Matematički fakultet)

EAD DTD
Encoded Archival Description Document Type Definition
-- standard za opis arhivskih dokumenata --

Prikazujemo strukturu standarda za opis arhivskih dokumenata EAD DTD i to verzije 2002. Ovaj standard je napravljen od strane The Society of American Archivists Encoding Archival Description Working Group radi olakšavanja razvoja programa za pretraživanje arhivske građe.

Ključne reči: standard, internet, SGML, XML

Uvod

EAD DTD (Encoded Archival Description Document Type Definition) standard nastao je 1993. godine, kao projekat kalifornijskog univerziteta Berkli, a kao potreba za pomoć pri registraciji i pretraživanju dokumenata koji se nalaze u arhivama, bibliotekama i muzejima. Ciljevi tok projekta bili su:

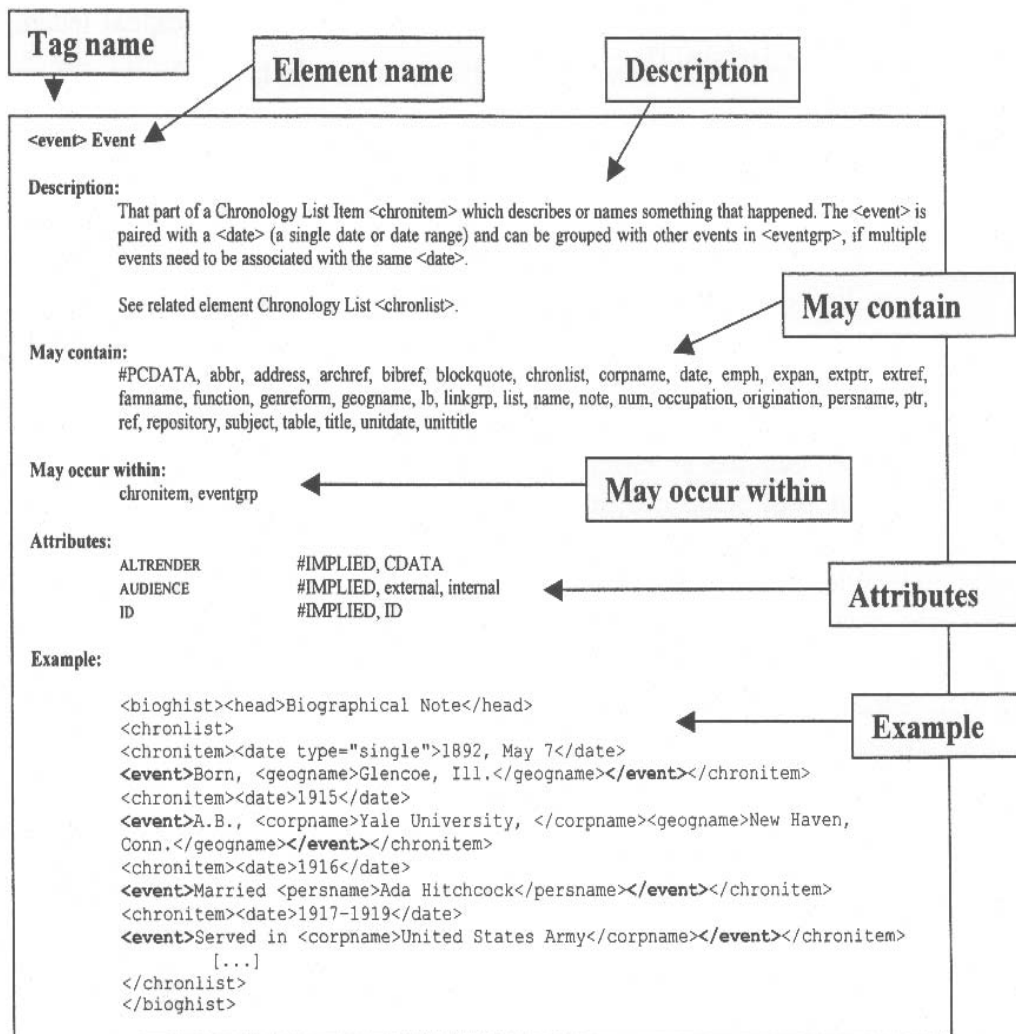
1. mogućnost prezentacija informacija, nađenih od strane pretraživača, sa svim svojim značenjima i međusobnim odnosima
2. mogućnost prikazivanja hijerarhijske povezanosti između različitih nivoa opisa
3. mogućnost prikazivanja informacije koje se nasleđuju prilikom prelaska sa različitih hijerarhijskih nivoa
4. mogućnost kretanja unutar hijerarhijske strukture
5. mogućnost specifičnog indeksiranja i oporavka od gubitka informacija

Osnova za razvoj ovog standarda bio je SGML (Standard Generalized Markup Language). SGML je skup pravila za definisanje i izražavanje logičke strukture dokumenata i na taj način omogućuje programima kontrolu pretraživanja, oporavka i strukturne prezentacije tih dokumenata. Pravila se primenjuju u obliku tagova koje se mogu uneti u opis dokumenta i time omogućiti identifikovanje i pravljenje odnosa između strukturnih celina.

DTD određuje skup SGML tagova koji omogućuju obeležavanje delova dokumenata i gledanje na te celine kao na klase. Poslednja verzija standarda EAD DTD finalizovana je decembra 2002. godine i u nju je implementiran standard ISO 8879, što znači da je unet skup novih pravila za podršku XML (eXtensible Markup Language) specifikaciji. Nedavni testovi izvršeni u Francuskoj i Nemačkoj pokazali su da se EAD može dobro prilagoditi i upotrebljavati u drugim sredinama.

Opis elemenata standardne biblioteke

Standardna biblioteka sadrži 146 tagova za opis dokumenata.



Informacije o svakom elementu standardne biblioteke prikazane su kao u gornjem primeru, tako da svaka informacija sadrži:

- **Ime taga** (Tag name)
Ime taga je u formi mnemonika imena elementa koji se koristi u mašinski čitljivom dokumentu. To je prva reč na vrhu strane. Ime se navodi unutar uglastih zagrada, npr. `<event>`, osim u listama koje govore gde se ovaj tag može naći i koje druge elemente može sadržati. Uvek se javlja napisan *malim slovima*.
- **Ime elementa** (Element name)
Ime elementa je proširena verzija imena taga koja služi za bolji opis značenja elementa. To je najčešće reč ili fraza koja identifikuje svrhu elemenata. Nalazi se iza imena taga i napisano je *Velikim Početnim Slovima*, npr. `<event> Event`.

- Opis (Description)
Prvi paragraf opisa definiše element u terminima osnovnih ili arhivskih rečnika, odnosno rečnika podataka kakvi su: *ISAD*, *MARC21*, *the Categories for the Description of Works of Art* ili *the Text Encoding Initiative DTD*. Svaki sledeći paragraf opisuje upotrebu elementa, opis datog elementa od adekvatnih elemenata, ukazuje na slične elemente ili korisne attribute, daje ilustrativan primer. Koriste se osnovne fraze da ukažu prirodu elementa koji se opisuje.
- Može sadržati (May contain)
Ova sekcija daje spisak elemenata koji se mogu javiti prilikom definisanja nekog elementa. Spisak je sortiran po alfabetskom redu imena tagova. Ovaj deo može biti prazan, sadržati tekst (nazvan *#PCDATA*, *Parsed Character DATA*), druge elemente ili biti mešavina navedenog. Skraćenica *#PCDATA* znači da se tekst može naći direktno unutar elementa, ali taj tekst ne može sadržati tekst koji se od strane parsera može shvatiti kao neka naredba. Lista ugnježenih elemenata obično sadrži više elemenata nego što je generalno potrebno. Ovo je posledica činjenice da elementi koji dele neki sadržaj čine grupu, tako da se odabirom jednog automatski uključuju i svi elementi grupe.
- Može se pojaviti unutar (May occur within)
Ova sekcija sadrži spisak svih roditeljskih elementa u kojima se dati element može naći. Lista je sortirana po alfabetskom redu imena tagova. Na osnovu ovoga se može zeključiti koliko je neki element u upotrebi, kao i još po neku informaciju o uporebi datog elementa.
- Atributi (Attributes)
Ova sekcija sadrži listu svih atributa koji se mogu prikazati elementu. Atributi se predstavljaju *malim slovima*. Online EAD Tag Library, međutim, koristi konvenciju *Velikih Početnih Slova*.
- Primer (Example)
Najveći deo opisa elemenata sadrži tagovani primer da ukaže kako se zajedno koriste atributi i elementi. Deo primera je uzet iz realnih programa i problema, dok su neki namenski konstruisani za ovu priliku.

Opis atributa elemenata standardne biblioteke

Atributi su pridruženi većini elemenata sadržanih u EAD. Oni reflektuju većinu osobina elemenata i mogu uzeti različite vrednosti, u zavisnosti od konteksta u kojem se nalaze. Dodela vrednosti atributa moguća na jedan od sledećih načina:

```
<[tag elementa] [atribut]=``[vrednost]''>
```

ili ako element poseduje više atributa:

```
"<[tag elementa] [atribut1]=``[vrednost1]'' [atribut2]=``[vrednost2]''>"
```

Atributi mogu biti jednog od sledeća tri tipa:

#IMPLIED - Atribut nije neophodan i ako mu nije dodeljena neka vrednost, program za obrađivanje može mu dodeliti neku vrednost.

#REQUIRED - Atribut mora biti specifikovan na početku taga elementa.

#FIXED - Atribut može imati samo jednu moguću vrednost definisanu definicijom tipa dokumenta (DTD)

Većina atributa je opciona, tj. tipa **#IMPLIED**. Dva atributa su naznačena kao **#REQUIRED**. To su atributi **LEVEL**, elemenata `<archdesc>` i `<archdescgrp>`, i atribut **COLS** elementa `<tgroup>`. Dva atributa su tipa **#FIXED**.

U zavisnosti od tipa vrednosti koje mogu atributi uzeti ili posebnih zahteva vezanih za te vrednosti, atributi se dele na:

CDATA: Vrednosti su iz skupa karaktera, bez specijalnih znakova <, >, &, ' i ".

ENTITY: Neobrađeni entitet deklarisan u lokalnoj deklaraciji dokumenta.

Program koji obrađuje dokument ovakve attribute uzima kao reference na druge elemente.

ID: Jedinstveni identifikator. Većina elemenata imaju svoj atribut **ID**, tako se taj jedinstveni kod može upotrebiti za referisanje tog specifičnog elementa.

Vrednosti atributa **ID** su upravo ovoga tipa. Programi za obradu potvrđuju da su vrednosti ovog tipa jedinstvene. Vrednosti ovog tipa atributa moraju početi slovom, dok ostali karakteri mogu biti slova, cifre, ili neki od sledećih karaktera: . (tačka), : (dvotačka), - (crtica), _ (podvlaka) i ne sme sadržati praznine.

IDREF: Identifikator reference, ID vrednost nekog predhodno unetog elementa.

IDREFS: Lista IDREF-a.

NMTOKEN: Imenovani token koji se može sastojati od slova, cifara, ili nekih od sledećih karaktera: . (tačka), : (dvotačka), - (crtica), _ (podvlaka) i ne sme sadržati praznine.

NMTOKENS : Lista NMTOKEN-a.

Vrednosti pojedinih atributa mogu biti samo iz zatvorenog ili poluzatvorenog skupa vrednosti.

Opis dokumenta EAD DTD standardom

Pre opisa dokumenta potrebno je proveriti da li pretrživač podržava XML i UNICODE standarde koji su ugrađeni u EAD DTD standard.

```
<?xml version="1.0" encoding="utf-8"?> <!DOCTYPE EAD PUBLIC
"+++ISBN 1-931666-00-8//DTD ead.dtd (Encoded Archival Description
(EAD) Version 2002)//EN">
```

Opis dokumenta počinje tagom <ead>, kao oznake da je reč o dokumentu opisanom EAD DTD standardom. Tagom <eadheader> specificuju se kodne strane koje se koriste za identifikaciju porekla dokumenta, njegovog opisa, podrška za druge sisteme za pretragu, kao i podaci vezani za sam dokument, koji su dostupni kroz druge podtagove ovoga taga. Tag <eadid> eksplicitno identifikuje porekla dokumenta (jezik, zemlju, prava korišćenja i sl.).

```
<ead>
  <eadheader countryencoding="iso3166-1" repositoryencoding="iso15511"
  langencoding="iso639-2b" relatedencoding="marc">
  <eadid countrycode="us" repositorycode="mnhi">terr06</eadid>
```

Tagom <filedesc> definišu se naslov, autor, vlasnik, izdanje i slični podaci vezani za dokument koji se opisuje.

```
<filedesc>
  <titlestmt>
  <titleproper>MINNESOTA TERRITORIAL ARCHIVES. Territorial Governor:
  </titleproper>
  <subtitle> An Inventory of Territorial Governor Willis A. Gorman
  </subtitle>
  <author>Finding aid prepared by Lydia Lucas.</author>
  </titlestmt>
</filedesc>
```

Tagom <profiledesc> daju se bliže informacije o samom procesu opisivanja dokumenta, osobama ili firmama koje opisale dokument, ako i datum kada je opis završen.

```
<profiledesc>
  <creation>Finding aid encoded by Lyda Morehouse,
  <date>July 17, 2002.</date>
</creation>
</profiledesc>
</eadheader>
```

Jedan od važnijih tagova u opisu dokumenta je <archdesc>. Pomoću njega i njegovih podtagova daje se informacije o dokumentu koje se odnose na kontekst i sadržaj dokumenta, kao i druge informacije o dokumentu (administrativne i slične prirode). Ove informacije organizovane su kroz hijerarhijske strukture, koje omogućavaju detaljniji uvid u pojedine delove dokumenta.

```
<archdesc level="series">
  <did id="a1">
    <head>OVERVIEW OF THE RECORDS</head>
    <repository label="Repository:">Minnesota Historical Society
    </repository>
    <origination label="Creator:">
    Agency:Minnesota. Governor (1853-1857 : Gorman).
    </origination>
    <unittitle label="Title:">
    Series Records of territorial governor Willis A. Gorman,
    </unittitle>
    <unitdate label="Date:">1852-1857.</unitdate>
    <abstract label="Abstract:">Subject files, correspondence,
    appointments, pardon records, reports from territorial officers,
    requests for return of fugitives, and miscellany.</abstract>
    <physdesc label="Quantity:">0.75 cu. ft. (23 folders in 2 partial
    boxes).
    </physdesc>
    <physloc label="Location:">
    See Detailed Description section for box locations.
    </physloc>
  </did>
```

Tagom <bioghist> daju se bibliografske ili istorijske informacije o dokumentu. Ovde se mogu naći informacije o životu autora, osobe o kojoj se govori u dokumentu, istorijske okolnosti nastanka dokumenta i sl. Ove informacije mogu se naći kao običan tekst (podeljen u paragrafe <p>) ili kao niz datuma povezanih sa nekim događajima (tag <chronlist>). Pojedine delove teksta moguće naglasiti prilikom prikazivanja tagom <emph> definisanjem vrednosti atributa **RENDER**.

```
<bioghist>
  <head id="a2">BIOGRAPHICAL SKETCH OF GOVERNOR WILLISA.GORMAN
  </head>
  <p>
```

Willis Arnold Gorman was born in Kentucky on January 12, 1816. His early career was as a lawyer and state legislator (Democrat) in Indiana. During the Mexican War he raised and commanded an Indiana rifle battalion and an infantry regiment, and subsequently was elected to two terms in the U.S. Congress. President Franklin Pierce appointed him Governor of Minnesota in May 13, 1853, and he served in that office until 1857. He was subsequently a delegate to the state constitutional convention(1857) and was elected to the first state legislature (1858).

```
</p>
```

<p>

At the outset of the Civil War he commanded the First Minnesota Infantry Volunteers (April-October, 1861) until his promotion to brigadier general. Following his discharge from military service in 1864, he practiced law in St. Paul in partnership with Cushman K. Davis. In 1869 he was appointed St. Paul city attorney, and held that office until his death on May 20, 1876.

</p>

<p>

Sources:

Minnesota Biographies, pp. 267-268;
Minnesota in the Civil and Indian Wars,
 Vol. I, p. 14; William Watts Folwell,
A History of Minnesota
 (St. Paul, rev. 1956), Vol. I, pp. 377-383; Theodore C. Blegen,
Minnesota, A History of the State
 (1963), pp. 171, 240-242, 293.</p>

</bioghist>

Tagom <scopecontent> daje se kratak sadržaj svih relevantnih informacija vezanih za dokument koji se opisuje. Ovaj tag, na taj način, pomaže pretraživaču pregled dokumenta i može skratiti vreme pretrage.

<scopecontent>

<head id="a3">SCOPE AND CONTENTS OF THE RECORDS</head>

<p>

Subject files and correspondence of territorial governor Gorman. A large proportion relate to territorial officers' appointments and commissions. Letters received and sent address the following topics: concerns of federal civilian agencies, military and Indian affairs in general, the 1855 treaty with the Winnebago Indians, land sales, the Minnesota and Northwestern Railroad Company, and other territorial affairs. Also present are a legislative diary (Jan. 27-March 6, 1857), pardon records, petitions regarding local government organization, reports from territorial officers, and requests from other states for return of fugitives.

</p>

<p>

Gorman's governorship spanned a period of both tangible and speculative growth in the state's infrastructure and economy, with numerous railroads, educational institutions, and corporations being chartered, new counties established, new towns incorporated, and several Indian treaties ratified. The surviving records contain only sketchy information on these developments, and do not appear to contain any significant documentation on an abortive 1857 scheme to remove the state capital from St. Paul to St. Peter.

</p>

</scopecontent>

Tagom <arrangement> daje se informacija kako je dokument podeljen na manje celine, kao i unutrašnju organizaciju opisa dokumenta.

<arrangement>

<head id="a4">ARRANGEMENT OF THE RECORDS</head>

<p>

Arranged alphabetically by topic; thereunder chronologically within each folder.

</p>

</arrangement>

Tagom <controlaccess> definišu se glavne pristupne tačke opisanog materijala. One se mogu odnositi na sadržaj dokumenta, osobe, organizacije, geografske i slične pojmove koje se nalaze u materijalu.

<controlaccess>

<head id="a7">INDEX TERMS</head>

<p>

<emph render="italic">This collection is indexed under the following headings in the catalog of the Minnesota Historical Society.

Researchers

desiring materials about related topics, persons or places should search

the catalog using these headings.</emph>

</p>

<controlaccess>

<head>Topics:</head>

<subject>Civil-military relations--Minnesota.</subject>

<subject>Extraditions--Minnesota.</subject>

<subject>Federal government--Minnesota.</subject>

<subject>Indians of North America--Government relations--1789-1869.</subject>

<subject>Land titles--Minnesota--Registration and transfer.</subject>

<subject>Pardons--Minnesota.</subject>

<subject>Winnebago Indians--Treaties.</subject>

</controlaccess>

<controlaccess>

<head>Places:</head>

<geogname>Minnesota--Officials and employees--Selection and appointment.

</geogname>

<geogname>Minnesota--Politics and government--1849-1858.</geogname>

</controlaccess>

<controlaccess>

<head>Persons:</head>

<persname>Gorman, Willis Arnold, 1816-1876.</persname>

</controlaccess>

<controlaccess>

<head>Organizations:</head>

<corpname>Minnesota Territory.</corpname>

</controlaccess>

<controlaccess>

<head>Types of Documents:</head>

<genreform>Diaries--Minnesota.</genreform>

<genreform>Petitions--Minnesota.</genreform>

<genreform>Proclamations--Minnesota.</genreform>

<genreform>Territorial records--Minnesota.</genreform>

</controlaccess>

</controlaccess>

Tagom <prefercite> definiše način na koji se referiše opisani dokument prilikom daljeg korišćenja.

<prefercite>

<head>Preferred Citation:</head>

```
<p>
<emph render="italic">[Indicate the cited volume and page or item
and folder title here].</emph> Minnesota Territorial Archives.
Records of territorial governor Willis Gorman. Minnesota Historical
Society. State Archives.
</p>
```

```
<p>
<emph render="italic">See the Chicago Manual of Style for additional
examples.</emph>
</p>
```

```
</prefercite>
```

```
<acqinfo>
```

```
<head>Accession Information:</head>
```

```
<p>Accession number(s): none; 2001-46</p>
```

```
</acqinfo>
```

Informacije o podacima korišćenim u pripremama za opis materijala da ju se tagom `<processinfo>`. Ove informacije mogu se razdavajati po paragrafima, ili davati na više različitih mesta u opisu dokumenta, ukoliko je to potrebno.

```
<processinfo>
```

```
<head>Processing Information:</head>
```

```
<p>PALS ID No.: 0800011304</p>
```

```
<p>RLIN ID No.: MNHV90-A1172</p>
```

```
</processinfo>
```

Tagom `<dsc>` daje hijerarhijska organizacija opisanog materijala. Tip organizacije materijala daje se atributom **TYPE**.

```
<dsc type="combined" id="fruin" audience="external">
```

```
<head id="a9">DETAILED DESCRIPTION OF THE RECORDS</head>
```

```
<p id="MHSLoc">
```

```
<emph render="italic">Note to Researchers: To request materials,
please
```

```
note both the location and box numbers shown below.</emph>
```

```
</p>
```

Tagom `<c>` daje se sadržaj samog materijala. Ovaj tag uvek je ugnežđen unutar taga `<dsc>`. Ukoliko je potrebno materijal podeliti na različite hirerajhijske nivoe, to se čini tagovima `<c01>`-`<c12>`, i to tako da su oni sa većim brojem unutar onih sa manjim brojem. Atributom **LEVEL** daje se značenje takvoj hijerarhijskoj podeli (serija, član, fajl, podgrupa, potserija ...). Identifikaciju celina ovako podeljenog materijala dajemo tagom `<did>`.

```
<c01>
```

```
<did>
```

```
<physloc>115.I.19.8F</physloc> <container>[1]</container>
```

```
<unittitle>Appointments: Commission of Deeds, </unittitle>
```

```
<unitdate>1853-1857. </unitdate>
```

```
<physdesc>1 folder.</physdesc>
```

```
</did>
```

```
<scopecontent>
```

```
<p>
```

```
For each individual, may include any of the following: application,
petition, recommendations, letter of resignation.
```

```
</p>
```

```
</scopecontent>
```



```

</c01>
<c01>
  <did>
    <unittitle>Reports: Other Territorial Officers, </unittitle>
    <unitdate>1853-1855. </unitdate>
    <physdesc>7 items.</physdesc>
  </did>
  <c02>
    <did>
      <unittitle>- Agent to select university lands (December 12, 1855).
      </unittitle>
    </did>
  </c02>
  <c02>
    <did>
      <unittitle>- Commissioner for the Crystal Palace Exhibition
      (1853, 1854).
      </unittitle>
    </did>
  </c02>
</c01>
</dsc>
</archdesc>
</ead>

```

Zaključak

EAD DTD standard je pogodan za opis tipičnih arhiva i arhivskih dokumenata. Standardna biblioteka ne predviđa tagove kojima bi se opisali neki multimedijalni sadržaji, ali ovaj nedostatak se može „zaobići“ većim korišćenjem tagova za opis sadržaja dokumenta i definisanjem linkova na stvarne lokacije tih multimedijalnih sadržaja.

Standardna biblioteka tagova trenutno je dostupna na web-u [1] u html formatu. Takođe, do njene verzije pogodne za štampu moguće je doći preko linka sa istog sajta. Kako je poslednja verzija relativno skoro standardizovana, još nije dostupna verzija pogodna za direktno korišćenje u programima.

Kao što je u uvodu rečeno, EAD DTD je izašao iz okvira SAD, tako da je broj onih koji učestvuju u razvoju ovog standarda veliki. Pored Berklija, tu su i drugi poznati univerziteti (Jeil, Harvard, Glazgov, Sidnej), kao i veliki muzeji, arhivi i biblioteke i u svakoj od ovih institucija postoji osoba koju je moguće kontaktirati u vezi sa ovim standardom.

Reference

1. EAD DTD
<http://www.loc.gov/ead>
2. ISAD(G) General International Standard Archival Description
<http://www.ica.org/eng/mb/com/cds/descriptivestandards>
3. MARC21 Format for Bibliographic Data
<http://lcweb.loc.gov/marc/bibliographic/ecbdhome.html>

EAD DTD
Encoded Archival Description Document Type Definition

We present the structure of EAD DTD encoding standard for archival finding aid, version 2002. This standard was created by The Society of American Archivists Encoding Archival Description Working Group.

Key words: standard, internet, SGML, XML