

Universidad Complutense de Madrid

Facultad de Ciencias de la Documentación



Desarrollo Informático de Aplicaciones Archivísticas: EAD en ARCHON

Manuel Blázquez Ochando
Juan-Antonio Martínez-Comeche

Madrid, 2008

16.- EAD ENCODED ARCHIVAL DESCRIPTION

¿Qué es EAD?

EAD es un lenguaje de codificación de descripciones archivísticas basado en el lenguaje de marcado extensible XML. EAD facilita la consulta y recuperación de las descripciones de las distintas agrupaciones documentales del archivo. Esto se logra gracias a la disposición detallada de cada uno de los niveles que conforman un verdadero instrumento de descripción.

¿Qué es XML?

Tal y como lo define el W3C (World Wide Web Consortium), XML es "... Lenguaje de Etiquetado Extensible (eXtensible Markup Language). Es un lenguaje con una importante función en el proceso de intercambio, estructuración y envío de datos en la web. Describe los datos de tal manera que es posible estructurarlos utilizando para ello etiquetas como lo hace HTML, pero que no están predefinidas, delimitando de esta manera los datos, a la vez que favoreciendo la interoperabilidad de los mismos..."

Tipos de documentos XML, Bien Formados y Validados:

Se debe distinguir que todo documento XML puede ser bien formado o validado.

- Son Well Formed o bien formados aquellos que sí cumplen con las características de construcción principales. Es decir, apertura y cierre de etiquetas correcto, empleo de caracteres aceptados, correcto anidamiento o raíz, correctacitación de los atributos de las etiquetas utilizadas, tal y como se establece en el W3C.
- Son Valid o validados aquellos que siendo bien formados, están contruidos no sólo conforme a una sintaxis correcta, si no a unas normas de estructuración y definición de los contenidos de sus etiquetas. Por regla general esto se lleva a cabo mediante archivos DTD o Schemas XSD

EAD, un XML Validado.

Los archivos EAD son documentos XML validados conforme a unas reglas de construcción o sintaxis de las etiquetas y contenidos de los que puede estar compuestos. Normalmente suelen ser de naturaleza archivística, por lo que se viene aplicando las Normas ISAD-G de descripción archivística. Por ello, un EAD validado estará formado por dos apartados esenciales:

- Archivo EAD con extensión XML: Dónde reside la codificación de los contenidos de la descripción archivística.
- Archivo DTD o Schema XSD: Dónde se codifican las normas que atañen a la sintaxis, codificación y tipologías de contenidos en cada etiqueta empleada.

Extracto de archivo codificado en EAD (Cabecera)

Fuente. (http://www.loc.gov/ead/tglib/appendix_ca.html)

<?xml version="1.0"?>

Desarrollo Informático de Aplicaciones Bibliotecarias y Documentales

Manuel Blázquez Ochando, Juan-Antonio Martínez-Comeche

<!DOCTYPE EAD PUBLIC "+//ISBN 1-931666-00-8//DTD ead.dtd (Encoded Archival Description (EAD) Version 2002)//EN" []>

<ead>

<eadheader audience="internal" countryencoding="iso3166-1" dateencoding="iso8601" langencoding="iso639-2b" repositoryencoding="iso15511">

<eadid countrycode="us" mainagencycode="cu-i" publicid="//us::cu-i//TEXT us::cu-i::p29.sgm//EN">Mildred Davenport Dance Programs and Dance School Materials, MS-P29

</eadid>

<filedesc>

<titlestmt>

<titleproper>Guide to the Mildred Davenport Dance Programs and Dance School Materials</titleproper>

<author>Processed by Adrian Turner; machine-readable finding aid created by Adrian Turner</author>

</titlestmt>

<publicationstmt>&hdr-cu-i-spcoll;

<date>© 2001</date>

<p>The Regents of the University of California. All rights reserved.</p>

</publicationstmt>

<notestmt>

<note>

<p>

<subject source="cdl">Arts and Humanities--Dance--Dance Performance</subject>

<subject source="cdl">Arts and Humanities--Dance--Dance History and Criticism</subject>

<subject source="cdl">Area, Interdisciplinary, and Ethnic Studies--African American Studies</subject>

</p>

</note>

</notestmt>

</filedesc>

<profiledesc>

<creation>Machine-readable finding aid derived from MS Word. Date of source:

<date>2001.</date></creation>

<language>Description is in <language>English.</language></language>

</profiledesc>

</eadheader>

<frontmatter>

<titlepage>

<titleproper>Guide to the Mildred Davenport Dance Programs and Dance School Materials</titleproper>

<num>Collection number: MS-P29</num>

<publisher>Special Collections and Archives

<lb>The UCI Libraries

<lb><extptr actuate="onload" show="embed" entityref="ucseal">

<lb>University of California

<lb>Irvine, California</publisher>

<list type="deflist">

<defitem>

<label>Processed by: </label>

<item>Adrian Turner</item>

</defitem>

<defitem>

<label>Date Completed: </label>

<item>2001</item>

</defitem>

<defitem>

Desarrollo Informático de Aplicaciones Bibliotecarias y Documentales

Manuel Blázquez Ochando, Juan-Antonio Martínez-Comeche

```
<label>Encoded by: </label>
  <item>Adrian Turner</item>
</defitem>
</list>
  <p>&copy; 2001 The Regents of the University of California. All rights reserved.</p>
</titlepage>
</frontmatter>...
```

Extracto de XSD para EAD (Niveles o Agrupaciones Documentales)

Fuente: (<http://www.loc.gov/ead/ead.xsd>)

```
<xs:simpleType name="av.level">
  <xs:restriction base="xs:token">
    <xs:enumeration value="class"/>
    <xs:enumeration value="collection"/>
    <xs:enumeration value="file"/>
    <xs:enumeration value="fonds"/>
    <xs:enumeration value="item"/>
    <xs:enumeration value="otherlevel"/>
    <xs:enumeration value="recordgrp"/>
    <xs:enumeration value="series"/>
    <xs:enumeration value="subfonds"/>
    <xs:enumeration value="subgrp"/>
    <xs:enumeration value="subseries"/>
  </xs:restriction>
</xs:simpleType>
```

Extracto de DTD para EAD: (Autoridades, Categorías)

Fuente: (<http://www.loc.gov/ead/ead2002.zip>)

```
<!-- Corporate Name -->
<!ELEMENT corpname (#PCDATA | %m.phrase.bare; | subarea)* >
<!ATTLIST corpname
  %a.common;
  %a.access;
  %am.corpname.role;
  %am.corpname.encodinganalog; >

<!-- Family Name -->
<!ELEMENT famname (#PCDATA | %m.phrase.bare;)* >
<!ATTLIST famname
  %a.common;
  %a.access;
  %am.famname.role;
  %am.famname.encodinganalog; >

<!-- Geographic Name -->
<!ELEMENT geogname (#PCDATA | %m.phrase.bare;)* >
<!ATTLIST geogname
  %a.common;
  %a.access;
  %am.geogname.role;
  %am.geogname.encodinganalog; >

<!-- Genre or Form -->
<!ELEMENT genreform (#PCDATA | %m.phrase.bare;)* >
<!ATTLIST genreform
```

Desarrollo Informático de Aplicaciones Bibliotecarias y Documentales

Manuel Blázquez Ochando, Juan-Antonio Martínez-Comeche

```
%a.common;  
%am.genreform.type;  
%a.access;  
%am.genreform.encodinganalog; >
```

<!-- Name -->

```
<!ELEMENT name (#PCDATA | %m.phrase.bare;)* >  
<!ATTLIST name  
  %a.common;  
  %a.access;  
  %am.name.role;  
  %am.name.encodinganalog; >
```

<!-- Occupation -->

```
<!ELEMENT occupation (#PCDATA | %m.phrase.bare;)* >  
<!ATTLIST occupation  
  %a.common;  
  %a.access;  
  %am.occupation.encodinganalog; >
```

<!-- Function -->

```
<!ELEMENT function (#PCDATA | %m.phrase.bare;)* >  
<!ATTLIST function  
  %a.common;  
  %a.access;  
  %am.function.encodinganalog; >
```

<!-- Personal Name -->

```
<!ELEMENT persname (#PCDATA | %m.phrase.bare;)* >  
<!ATTLIST persname  
  %a.common;  
  %a.access;  
  %am.persname.role;  
  %am.persname.encodinganalog; >
```

<!-- Subject -->

```
<!ELEMENT subject (#PCDATA | %m.phrase.bare;)* >  
<!ATTLIST subject  
  %a.common;  
  %a.access;  
  %am.subject.encodinganalog; >
```

Estructura Esencial de un Documento EAD.

```
<ead>  
  <eadheader>  
    <filedesc></filedesc>  
    <revisiondesc></revisiondesc>  
    <requiredft></requiredft>  
  </eadheader>  
  
  <frontmatter>  
    <titlepage></titlepage>  
    <div></div>  
  </frontmatter>
```

Desarrollo Informático de Aplicaciones Bibliotecarias y Documentales

Manuel Blázquez Ochando, Juan-Antonio Martínez-Comeche

```
<findaid>
  <archdesc>
    <did></did>
    <admininfo></admininfo>
    <bioghist></bioghist>
    <controlaccess></controlaccess>
    <note><note>
    <odd></odd>
    <scopecontent></scopecontent>
    <tspec></tspec>
    <thead><thead>
    <c></c>
    <c01></c01>
    <dsc></dsc>
  </archdesc>

  <add>
    <head></head>
    <bibliography></bibliography>
    <div></div>
    <fileplan></fileplan>
    <index></index>
    <relatedmaterial></relatedmaterial>
    <separatedmaterial></separatedmaterial>
  </add>
</findaid>
</ead>
```

<eadheader>

<filedesc> descripción del instrumento **</filedesc>**

<revisiondesc> descripción revisada **</revisiondesc>**

<requiredft> estado de borrador **</requiredft>**

</eadheader>

<eadheader> es la etiqueta que presenta la información del instrumento de descripción y esta compuesta por varios subelementos como **filedesc**, **revisiondesc** y **requiredft** que permiten introducir una descripción de las características de dicho instrumento de descripción, posibles revisiones a los planteamientos del instrumento y notas de aclaración en caso de constituir aún un borrador

<frontmatter>

<titlepage> título del instrumento para la descripción **</titlepage>**

<div> resúmenes, introducciones, prefacios **</div>**

</frontmatter>

<frontmatter> es la etiqueta que posibilita incluir información adicional respecto del instrumento de descripción y por lo tanto el eadheader. Permite añadir datos esenciales como título del instrumento, introducciones, resúmenes, etc.

<findaid>La descripción del material archivístico en su conjunto**</findaid>**

<findaid> es una etiqueta contenedora de toda la descripción archivística, contextual, administrativa y complementaria que se puede esperar. Las etiquetas que conforman el grupo findaid tienen una correlación directa con las normas ISAD-G

<archdesc>Descripción archivística propiamente dicha**</archdesc>**

<add>Información complementaria o asociada**</add>**

<archdesc> es la etiqueta que comprende la descripción archivística propiamente dicha por lo que contiene las principales áreas de descripción de las normas ISAD-G. Suele estar conformada por los siguientes atributos

- **“level”** que permite indicar la agrupación que se está describiendo, por ejemplo Fondo, Subfondo, Serie, Subserie, Expediente, es decir (Collection, Fonds, Recordgrp, series, subgrp, subseries, file, item, otherlevel...).
- **“audience”** que permite especificar el tipo de destinatario de la descripción archivística. Puede adoptar valores *Internal* o *External*.
- **“Langmaterial”** para especificar el idioma en el que está escrito. Por ejemplo spa, eng, fra...
- **“legalstatus”** en relación al estatuto legal que determina el tipo de acceso, público o restringido. (Public, Restricted)
- **“relatedencoding”** atributo que especifica la codificación complementaria a la que se está empleando que por defecto se basa en las normas ISAD-G,
- **“Type”** permite diferenciar qué tipo de instrumento de descripción se trata, o un inventario o un registro. (register, inventory)

<add> es la etiqueta con información complementaria asociada a la descripción codificada en la etiqueta <archdesc>

<archdesc level="fonds">

<did>área de identificación**</did>**

<admininfo>área de información administrativa**</admininfo>**

<bioghist>Historia de la Entidad o Nota Biográfica**</bioghist>**

<controlaccess>Vocabulario controlado**</controlaccess>**

<note>área de notas**</note>**

<dao>asociación con documentación digital**</dao>**

<scopecontent>área de alcance y contenido**</scopecontent>**

<odd>otros datos descriptivos**</odd>**

<c>área de componente**</c>**

<c01>Información de componente 01**</c01>**

<dsc>Contenedor con el Tipo de descripción de los elementos subordinados**</dsc>**

</archdesc>

A su vez la etiqueta <archdesc> consta de múltiples subelementos que permiten codificar las distintas áreas de descripción que caben esperar en las normas ISAD-G. A saber, el área de identificación representado por la etiqueta <did>, área de información administrativa <admininfo>, área de nota biográfica o historia de la entidad correspondiente al sujeto productor <bioghist>, área de control de autoridades y vocabulario normalizado <controlaccess>, área de notas <note>, <odd>

<did>

<repository> Archivo **</repository>**
<origination> Origen o Procedencia (Sujeto Productor) **</origination>**
<unittitle> título de la agrupación **</unittitle>**
<unitdate> fechas extremas **</unitdate>**
<physdesc> Descripción física (Volumen) **</physdesc>**
<abstract> Resumen de la agrupación
 <title> Posible título destacable **</title>**
</abstract>

</did>

La etiqueta <did> contiene las etiquetas correspondientes a los campos de identificación en ISAD-G. En este caso se presentan los más importantes; repository para incluir la denominación del archivo o institución que dispone de la agrupación documental. Etiqueta origination para determinar el origen, procedencia o sujeto productor de la agrupación. El título de la unidad o agrupación queda consignado en la etiqueta unittitle. Los datos de volumen de la descripción física estarán dispuestos en la etiqueta physdesc. Finalmente abstract contendrá un breve resumen que aporte información general sobre la agrupación a modo de contexto, de la cual puede extraerse un título destacable o atribuido mediante la etiqueta title.

<admininfo>

<acqinfo> Información sobre la adquisición de la documentación ya sea por medio de transferencia, donación, compra... **</acqinfo>**
<custodhist> Información sobre la historia de la custodia de la documentación **</custodhist>**
<accessrestrict> Restricciones de acceso **</accessrestrict>**
<userrestrict> Restricciones de uso o consulta **</userrestrict>**
<altformavail> Otros formatos físicos alternativos y su descripción física **</altformavail>**
<prefercite> Citas relevantes predefinidas **</prefercite>**
<appraisal> Información sobre la valoración de la agrupación documental, como por ejemplo, valores primarios y secundarios, procesos de expurgo, conservación, etc. **</appraisal>**
<accruals> Ingresos por medio de transferencias que se han efectuado en relación a la agrupación documental, según calendario de transferencia **</accruals>**
<processinfo> Información en relación al procesado de la agrupación documental **</processinfo>**

</admininfo>

La etiqueta <admininfo>, presenta información esencial en relación a la gestión administrativa de la agrupación documental en el archivo. Atiende a los efectos de la transferencia, procesos de valoración, selección, eliminación o expurgo, restricciones de acceso y uso, determinados por la legislación aplicable y vigente.

<bioghist>

<head>Denominación del campo o Nota Biográfica o Historia Institucional**</head>**

<title>Título de la nota biográfica o historia de la Entidad**</title>**

<persname>autoridad personal (Sujeto Productor)**</persname>**

<corpname>autoridad corporativa, entidad (Sujeto Productor)**</corpname>**

<event>autoridad, evento reseñable**</event>**

<date>autoridad, fecha reseñable asociada a un evento**</date>**

<geogname>autoridad, lugar geográfico reseñable asociado a un evento y una fecha concreta**</geogname>**

</bioghist>

La etiqueta <bioghist> permite definir la procedencia y la historia del sujeto productor de la documentación que se describe. Además se facilita la incorporación de autoridades de muy diferente ámbito como eventos, fechas, lugares geográficos que facilitan la organización y estudio cronológico-historiográfico.

<controlaccess>

<persname>Nombre de Persona**</persname>**

<corpname>Nombre de Instituciones**</corpname>**

<famname>Nombre de Familia**</famname>**

<geogname>Nombre Geográfico**</geogname>**

<name>Nombres**</name>**

<genreform>Género / Forma**</genreform>**

<function>Funciones**</function>**

<occupation>Ocupaciones**</occupation>**

<subject>Materias**</subject>**

<title>título**</title>**

</controlaccess>

<controlaccess> es un elemento que sirve para proporcionar al instrumento de descripción todas las autoridades normalizadas que pueden encontrarse en todo el documento EAD. De hecho en esta bolsa de contenidos, se validan a su vez todos los elementos marcados en cualquier apartado de la descripción, refrendando la disposición de archivo validado por la DTD correspondiente o XSD.

<c>Componente no numerado**</c>**

<c01>Componente numerado**</c01>**

Los componentes son etiquetas que engloban contenidos intelectuales fácilmente

reconocibles como elementos subordinados de la agrupación documental que se describe. Es decir, que podrían conformar verdaderos inventarios de contenidos a uno o varios niveles inmediatamente inferiores respecto al que se está describiendo. Por ello, embebido a cada etiqueta puede aplicarse todo el bloque de etiquetas <archdesc><did>... para describir a su vez cada uno de los componentes de manera anidada, quedando registro o constancia de cada descripción en bloque en un único instrumento de descripción.

<dsc type="combined">Contenedor con el Tipo de descripción de los elementos subordinados**</dsc>**

La etiqueta <dsc> permite definir el tipo de relación existente en la jerarquía de componentes subordinados además de establecer el tipo de descripción de dichos elementos, como por ejemplo, descripción combinada, analítica, en profundidad o de otro tipo (combined, analytic overview, in-depth, othertype). Dichos aspectos se consignan por medio del atributo "type". Es importante señalar que <dsc> es una etiqueta contenedora de componentes que por regla general suelen ser numerados y que a su vez pueden contener descripciones completas mediante las etiquetas <archdesc>, <did>, etc.

<add>

<bibliography>Documentación bibliográfica**</bibliography>**

<fileplan>Planos de Archivo**</fileplan>**

<index>Índices**</index>**

<otherfindaid>Otros instrumentos de descripción**</otherfindaid>**

<relatedmaterial>Documentación relacionada**</relatedmaterial>**

<separatedmaterial>Documentación separada**</separatedmaterial>**

<add>

Como se ha citado anteriormente la etiqueta <add> permite agregar información complementaria o asociada y a su vez se compone de documentación bibliográfica, planos de archivo, índices, otros instrumentos de descripción, documentación relacionada y separada.

EAD en Archon.

Archon posibilita la creación automática de documentos EAD, a partir de las descripciones elaboradas desde el módulo de colecciones. Esto significa que se pueden comprobar los elementos de codificación anteriormente mencionados en cada uno de los archivos generados, disponibles para el administrador desde el portal de archivos de archon correspondiente.

La documentación de cada agrupación documental, queda de esta forma codificada, y resulta susceptible de ser utilizada para muy diversos fines, a saber:

- Redifusión de contenidos archivísticos
- Desarrollo de catálogos de colecciones y fondos archivísticos
- Preparación de archivos para su exportación / importación en labores de migración de datos.

Desarrollo Informático de Aplicaciones Bibliotecarias y Documentales

Manuel Blázquez Ochando, Juan-Antonio Martínez-Comeche

- Transformación y presentación de la información al investigador.

Para demostrar la potencialidad y posibilidades del formato EAD, se puede comprobar el funcionamiento de algunos sistemas generadores de informes, específicamente desarrollados para facilitar a los archivos el servicio a los investigadores y proveer al usuario en general de una información, fácilmente legible y estructurada, que de otra forma podría resultar muy difícil de conseguir. Precisamente en esta línea la Biblioteca del Estado de Oregon desarrolló un generador de informes a partir de documentos codificados en EAD <http://eadpdfgen.library.oregonstate.edu/>

El resultado de aplicar dicho generador a un documento EAD, es un documento PDF listo para ser utilizado por el usuario o investigador. El uso intensivo de dicho recurso sobre la totalidad de documentos EAD de un archivo puede llegar a facilitar sobremanera el desarrollo de verdaderos catálogos archivísticos, facilitando con ello uno de los objetivos esenciales de la archivística.