



Consultora de Ciencias de la Información

Buenos Aires

Argentina

Serie

DOCUMENTOS DE TRABAJO

Área: Archivología

**EAD y EAC: Aplicación de metadatos en la descripción de colecciones
archivísticas**

Lic. Marcelo de la Puente

Enero 2018

N° 068

ISSN 1852 - 6411

Copyright Consultora de Ciencias de la Información

Editor: Patricia Allendez Sullivan. Asistente Editorial: Analía Bedrosian

Puente, Marcelo de la

EAD y EAC: aplicación de metadatos en la descripción de colecciones archivísticas. Buenos Aires: Consultora de Ciencias de la Información, 2018.

ISSN 1852 – 6411

1. EAD. 2. EAC. 3. Metadatos. 4. Colecciones archivísticas

I. Título

Resumen

Los cambios que han producido las nuevas tecnologías han afectado la forma de gestionar la información así como el comportamiento de los usuarios. Actualmente, se genera una gran cantidad de información en formato electrónico, lo que requiere la aplicación de técnicas documentales complejas. En este trabajo nos vamos a centrar en los siguientes temas: metadatos, metalenguajes, la descripción archivística, entre otras temas a tratar, todos ellos relacionados con el trabajo en los archivos.

Introducción

La irrupción de las tecnologías de la información y de la comunicación (TICs) en bibliotecas, centros de documentación y archivos han producido un cambio fundamental tanto en el trabajo de los responsables de gestionar toda esta información y los documentos que integran los diferentes fondos documentales, como en el comportamiento de los usuarios frente a la búsqueda de información relevante. La cantidad exponencial de información generada en múltiples formatos, principalmente electrónicos y vinculados al fenómeno de Internet, red de redes que todo lo abarca, está provocando fuertes cambios en todos los niveles y sectores económicos de la sociedad y condicionan la evolución de la sociedad al influir en la conducta de sus miembros. Todo ello ocurre en el fenómeno denominado “Sociedad de la Información” por diferentes teóricos como Miguel Castells. Fenómeno que se enmarca en el marco de una economía globalizada y de una sociedad altamente competitiva, en la que existe y se genera una enorme cantidad de información sin parar y sin control, leyes bibliométricas, como la Ley de Price (crecimiento exponencial de las publicaciones científicas), la ley de obsolescencia (citación solo de los trabajos más recientes), la ley de Bradford (concentración de la información en un núcleo de fuentes), entre otras, dan cuenta de este fenómeno, de todos estos cambios que llevan a la inmaterialidad de la información. Todos estos factores llevan a

un cambio del paradigma de la posesión por el del acceso a la información, en el esta se desvincula de un soporte material físico y se vuelve algo intangible y accesible a través de múltiples soportes electrónicos y redes digitales, a la vez que un elemento de valor estratégico clave para la competencia económica y el logro de objetivos

En el contexto del proceso que implicó la automatización de bibliotecas y archivos a finales del siglo XX y a comienzos del siglo XXI, los profesionales encargados de la gestión documental se enfrentan a un conjunto de innovaciones y cambios en el grupo de tareas que se deben realizar para almacenar, organizar y recuperar la información producida por las organizaciones y registrada en un entorno totalmente digital, tanto desde el punto de vista de producción de la información, como de su gestión, en el que el procesamiento técnico tradicional ya no es eficiente y se requieren nuevas técnicas y formatos electrónicos para poder procesar toda esa cantidad de información. A ello se le suma, en el caso de los archivos, la particularidad de ser documentos, ya sean analógicos o digitales, únicos y auténticos de valor jurídico y con relaciones complejas multinivel, series, expedientes, etc. con una complejidad que dificulta su procesamiento. Todos estos cambios llevan a una necesidad de normalizar la información que implica la descripción archivística en un entorno digital con la ayuda de las Tecnologías de la Información, lo que permita llegar al descubrimiento de recursos archivísticos en el entorno web en formatos estructurados para este tipo de documentos

Las empresas y las instituciones generan flujos documentales en su mayor parte en formatos electrónicos, lo mismo ocurre en los organismos públicos en lo que suele denominarse la e-administración o administración de procesos y servicios por medios electrónicos y estos flujos requieren de normas y formatos adecuados para su procesamiento eficiente, para su gestión documental. Debe tenerse en cuenta que en este tipo de documentos en formatos electrónicos y de naturaleza hipertextual es difícil definir en que consiste la unidad documental, a

lo que se le suman los problemas de la falta de normalización y el carácter volátil y efímero de muchos documentos digitales, lo que cual complica su tratamiento y procesamiento documental a lo largo de todo el ciclo de vida del documento, entendiéndose por tal, un proceso integrado de gestión de la información contenida en el documento desde su creación, procesamiento, organización, búsqueda, recuperación y posterior diseminación de la información contenida en él.

Todo este escenario requiere la aplicación de técnicas documentales complejas basadas en normas y en la aplicación de metadatos para especificar el contenido, estructura y contexto de los registros a lo largo del tiempo. Solo recientemente se han desarrollado estos nuevos estándares en el campo archivístico que permiten la descripción archivística de documentos digitales.

Esta nueva forma de gestión documental implica enfocar la atención desde el punto de vista de normas que permitan describir la información de estos registros documentales basadas en las TICs, normas que permitan describir y utilizar los recursos de una forma más eficiente, lo cual significa tener en cuenta la importancia de la descripción de acuerdo a los nuevos estándares y es aquí cuando surge la importancia del concepto de metadatos y su uso para el almacenamiento, recuperación y diseminación de un conjunto de objetos de información digital.

En primer lugar que puede entenderse como metadatos, un concepto confuso que da lugar a múltiples definiciones, entre conceptos relacionados como datos, información, descripción, etc. Etimológicamente metadatos significa datos sobre datos, en general un objeto que describe o dice algo sobre otro objeto de información. Aunque el uso de la palabra "metadato" se masificó en un contexto que se refiriere a la era de la información digital, la generación de metadatos data de siglos atrás. Los bibliotecarios han creado metadatos que han tomado la forma de catálogos de libros, catálogos de tarjetas y en la actualidad catálogos

en línea. Hoy en día la generalización del concepto ha cubierto cualquier tipo de información descriptiva (estandarizada) sobre recursos, incluyendo los que no son digitales. De manera formal podríamos decir que un metadato es un dato que se encarga de mantener un registro sobre el significado, contexto o propósito de un objeto informativo, de tal forma de poder descubrir, entender, extraer y administrar dicho objeto. En general estos registros son de menor envergadura que los objetos que describen, y son creados en un formato corto y conciso de tal forma que puedan intercambiarse.

No obstante debe aclararse que el termino metadatos en el sentido actual va más allá de una equivalencia con el registro bibliográfico tradicional, se relaciona con el nuevo entorno de la información en el que la naturaleza del objeto descrito es diferente, es una entidad virtual o inmaterial, más que física y requiere registrar información sobre el contenido, la estructura del documento y su contexto, así como de su gestión a lo largo del tiempo. Tomando una definición de Méndez Rodríguez se puede decir que los metadatos son “Información estructurada o semi-estructurada que facilita la creación, gestión y uso de registros a través del tiempo, dentro del dominio en que fue creado o a lo largo de él. Los metadatos para la gestión de documentos digitales pueden usarse para identificar, autenticar y contextualizar registros; y los procesos y sistemas que los crean y gestionan y los mantienen y utilizan”.

La Norma UNE-ISO 15489-1 entiende la gestión de metadatos como una parte inseparable de la gestión de documentos, los metadatos se definen “como datos que describen el contexto, contenido y estructura de los documentos, así como su gestión a lo largo del tiempo”.

La Norma UNE-ISO 23081-1, los metadatos son “información estructurada o semiestructurada que posibilita la creación, registro, clasificación, acceso, conservación y disposición de los documentos a lo largo del tiempo y dentro de un mismo dominio o entre dominios diferentes. Cada uno de estos dominios,

representa un área del discurso intelectual y de la actividad social o de la organización desarrollado por un grupo propio o limitado de individuos que comparten ciertos valores y conocimiento”. La misma norma explicita el objetivo de los metadatos: “Los metadatos para la gestión de documentos pueden usarse para identificar, autenticar y contextualizar tanto los documentos como los agentes, procesos y sistemas que los crean, gestionan, mantienen y utilizan, así como las políticas que los rigen”.

Los metadatos definen las relaciones entre los objetos, como las tuplas en una base de datos o clases en orientación a objetos, generando estructuras, etc. Puede decirse que dependiendo de la función y de las necesidades los metadatos se pueden aplicar en tres niveles:

- Documentos individuales
- Colecciones distribuidas de documentos u objetos
- Sistemas completos de gestión de documentos

Existen clasificaciones en base a la función desempeñada

En cuanto a la tipología, Méndez Rodríguez (2003) enumera cinco tipos esenciales de metadatos aceptados por todas las comunidades o dominios de metadatos:

1. Metadatos descriptivos, aquellos que dependen del propio documento y sirven para representar o identificar los objetos de información digital en su fase de organización. En el nivel local permiten la búsqueda y recuperación y en un nivel permiten a los usuarios descubrir recursos. Ejemplos: Dublin Core, Darwin Core, etc. Vocabularios controlados digitales como el tesoro sobre Arte y Arquitectura
2. Metadatos administrativos o metadatos para la gestión de recursos, los utilizados para la gestión y el procesamiento de las colecciones digitales,

a corto y a largo plazo Ejemplos (MOA2 o Administrative Metadata Elements, National Library of Australia. Preservation Metadata for Digital Collections)

3. Metadatos técnicos, aquellos creados por, o generados para, un sistema automatizado.
4. Metadatos de uso, generalmente creados de forma automática, relativos al nivel de utilización y al tipo de usuarios de un determinado servicio de información.
5. Metadatos para la conservación, aquellos metadatos destinados a gestionar la preservación de las fuentes de información, como Preserving Access to Digital Information (PADI)

Sin embargo, quizás la tipología más completa es la que se funda en la complejidad estructural de los modelos de metadatos, como la establecida por Demsey y Heery, quienes basándose en la riqueza estructural y semántica y en su complejidad, agrupan los metadatos y los esquemas de metadatos en 3 tipos que denominan bandas o zonas.

Estas bandas o zonas permitirían agrupar los metadatos según un *continuum* de crecimiento de riqueza semántica y estructural que iría desde una menor riqueza y complicación (banda 1) hasta una mayor riqueza y complejidad en la descripción (banda 3), pasando por una banda 2 en la que tenemos formatos concisos, estructurados en campos y estándares de facto, más que internacionales como el Dublín Core, en la banda 3 los formatos son mucho más estructurados y complejos, como el formato MARC y se basan en estándares internacionales. Existirían, además, metadatos situados en determinadas zonas con características comunes. Además, esta tipología englobaría tanto aspectos referidos a los atributos de los metadatos, como a los distintos modelos de metainformación que comportan. Heery ha llegado a desdoblar la banda 3 en una cuarta banda, donde se desglosarían los modelos muy elaborados y

especializados tales como EAD, y entre los que hoy cabría incluir los esquemas de metadatos basados en XML/RDF.

Los metadatos desestructurados se corresponden con la indización propia de la Web, que carece de semántica y realiza las búsquedas por texto libre, los metadatos semiestructurados se corresponden con modelos que tienen una sintaxis atributo-valor y que cuentan con algún tipo de particularidad semántica que permite la búsqueda por campos -tal como los metadatos Dublin Core - y los metadatos estructurados se corresponden con los lenguajes de marcado tipo SGML o XML que ya poseen una gran carga sintáctica y semántica, lo que permite tanto la conservación como la recuperación de los recursos digitales en dominios muy específicos.

En cuanto a la creación de los metadatos hay que tener en cuenta dos aspectos importantes: quién los asigna y cómo los asigna.

En la inmensa maraña de la Web, la mayor parte de los documentos y recursos digitales son creados por autores personales individuales sin ninguna experiencia en lenguajes hipertextuales, metadatos, etc. puesto que existen sistemas de gestión de hipertexto y herramientas que permiten la edición de páginas web de forma muy sencilla. Sin embargo, los diseñadores y editores *web* son cada vez más conscientes de que si quieren tener verdadera presencia en la red, deben facilitar que los buscadores, los robots y agentes inteligentes sean capaces de indizar sus páginas para que los usuarios puedan encontrarlas y recuperarlas. Desde hace algunos años, los principales editores web permiten la inclusión de etiquetas META, ya que los buscadores rastrean en la cabecera de los documentos para extraer las etiquetas **<META NAME="KEYWORDS">** y **<META NAME="DESCRIPTION">** e indizar las páginas.

La implementación de la ISO 15489 la norma establece los siguientes tipos de metadatos:

1. Metadatos sobre el documento mismo.
2. Metadatos sobre procedimientos, políticas y regulaciones.
3. Metadatos sobre agentes.
4. Metadatos sobre actividades de gestión o procesos de negocio.
5. Metadatos sobre actividades de gestión de documentos

En general su uso se restringe a los metadatos de descubrimiento y descripción de recursos digitales, como en el caso del formato MARC o el de el Núcleo de Dublín, un esquema de metadatos utilizado con un consenso amplio para la descripción de recursos digitales, pero existen metadatos de diferentes clases, metadatos para el descubrimiento de recursos, para su descripción, para la gestión y preservación de los mismos, metadatos técnicos, que tratan garantizar la conservación, el funcionamiento etc., podemos decir que metadatos según definiciones de la norma ISO 11179.1 son datos que definen y describen otros datos o procesos, son representaciones de hechos, conceptos, instrucciones de manera formalizada adecuada para la comunicación, interpretación o procesamiento por humanos o por máquinas, en algún contexto Lo que los metadatos describen son objetos digitales que pueden ser un ítem o conjunto de ítems, documentos individuales o agregados orgánicos de documentos individuales como expedientes, series, etc., es decir , representaciones o datos sobre los objetos reales vinculados a través de diferentes relaciones multinivel

Básicamente, los metadatos sirven para:

- 1) Certificar la autenticidad y el grado en que el contenido está completo, permitiendo su accesibilidad y disponibilidad a lo largo del tiempo
- 2) Establecer y documentar el contexto del contenido, facilitando así la comprensión del documento

- 3) Identificar y explotar las relaciones estructurales que existen entre y dentro de objetos de información, es decir, proporcionar vínculos lógicos entre los documentos y su contexto de creación, posibilitando su mantenimiento de forma estructurada e inteligible
- 4) Proporcionar un amplio rango de puntos de acceso intelectuales para un crecientemente diverso rango de usuarios
- 5) Proporcionar parte de la información que un profesional de la información podría haber proporcionado en un entorno físico de referencia o investigación
- 6) Contribuir a garantizar la autenticidad, fiabilidad e integridad de los documentos.
- 7) Respalda la gestión del acceso, la privacidad y los derechos de propiedad intelectual.
- 8) Servir de base para una recuperación eficiente
- 9) Respalda las estrategias de interoperabilidad, permitiendo que se incorporen oficialmente al sistema documentos creados en diversos entornos administrativos y técnicos y que se mantengan durante tanto tiempo como sea necesario
- 10) Facilitar la migración eficiente y exitosa de documentos electrónicos de un entorno o plataforma informáticos a otro, o cualquier otra posible estrategia de conservación

En particular en el caso de las colecciones archivísticas que han sido tradicionalmente muy difíciles de normalizar por sus características especiales, si es importante llegar a normalizar principios generales que subyacen a las prácticas comúnmente utilizadas. En términos generales, las normas ofrecen el mecanismo para utilizar los recursos existentes de una manera más eficiente y un paso clave era el diseño de instrumentos de descripción estructurados de tal forma que permitiesen la difusión a nivel internacional de la información de archivo. En la descripción, como en otros procesos archivísticos, la unanimidad no es, ni siquiera aproximadamente, total. Múltiples factores (variedad y unicidad

de los materiales de archivo, tamaño y complejidad de la organización original, variación en los niveles de descripción, variación en la profundidad o complejidad de la descripción, diversidad terminológica, etc.) inciden en la descripción archivística y han hecho que en cada época y en cada país se desarrolle una terminología, una técnica y una metodología distinta, lo que hizo especialmente difícil la normalización internacional de la descripción archivística.

Metalingüajes

SGML (ISO, 1986) es un metalingüaje, es decir un medio de describir formalmente un lenguaje, en este caso, un lenguaje de codificación etiquetado.

Es un sistema "descriptivo" que se sirve de códigos que simplemente ofrecen nombres para categorizar e identificar partes de un documento. Esto significa que SGML es una norma elaborada para expresar estructuras de contenido en lugar de apariencia de documentos. Es decir, usa códigos de marcado (etiquetas) que proporcionan nombres para categorizar las partes de un documento. Con metalingüajes de etiquetado descriptivo, como SGML, se diferencia claramente entre contenido y presentación, esto permite que el mismo documento pueda ser procesado fácilmente con muchos tipos diferentes de software, cada uno de los cuales puede aplicar diferentes instrucciones de procesamiento.

Todo ello es posible porque SGML introduce la noción de "tipo de documento", es decir una document type definition o definición de tipo de documento(DTD).

La DTD es la herramienta para "validar" un documento SGML. En la DTD, el tipo de un documento es definido formalmente por sus partes constituyentes y su estructura. Es el "libro de instrucciones", que permite definir libremente vocabularios de etiquetas que podrán ser procesados sobre cualquier tipo de plataforma e implica que diferentes documentos del mismo tipo pueden ser

procesados de una manera uniforme. Por otra parte, SGML proporciona un mecanismo de aplicación general para la sustitución de cadenas (string substitution), que es una forma simple de asegurar la independencia de sistemas concretos.

Lo que hace es proporcionar las herramientas que permiten diseñar las instrucciones para que una máquina "entienda" un documento por su definición estructural.

El documento concreto, que se denomina "modelo de documento" será un texto (en nuestro caso un instrumento de descripción) etiquetado. Las etiquetas utilizan delimitadores para ser distinguidas del texto en sí. Los delimitadores son "<" (inicio de etiqueta) y ">" (final de etiqueta). Normalmente una etiqueta inicio indicará el comienzo de un elemento y una etiqueta final (añadiendo una barra inclinada al inicio de etiqueta "</") determinará su terminación (decimos normalmente ya que el final de un elemento puede venir indicado por el inicio del siguiente, en cuyo caso se pueden aplicar reglas de minimización).

Las etiquetas indican la aparición de un elemento. Elemento (element) es la unidad estructural simple SGML (existen, también, atributos que normalmente son cualificadores del contenido de elementos-; y entidades partes nominadas de un documento marcado, independiente de cualquier consideración estructural-). Cada elemento deberá ser definido (declarado) atendiendo a su contenido estructural.

La declaración de elemento especificará el nombre (identificador genérico) de dicho elemento y su modelo de contenido (es decir qué elementos lo pueden constituir). La DTD presentará jerárquicamente las declaraciones de todos los elementos (junto a las declaraciones de entidades y de listas de atributos) que puede contener el tipo de documento en cuestión. En consecuencia, una representación gráfica imaginaria de la DTD sería similar a un árbol genealógico,

iniciándose en un ancestro (componente superior) que será la declaración del tipo de documento y descendiendo hasta los elementos más inferiores, cuyo modelo de contenido ya no serán otros elementos, sino cualquier cadena de caracteres.

La DTD describe la estructura de un documento, como un esquema de base de datos describe los tipos de información que trata y las relaciones entre campos.

Así, pues, una DTD proporciona una estructura para los elementos que constituyen un documento (por ejemplo, capítulos, encabezamientos, secciones, materias...); y también especifica las reglas para las relaciones entre elementos. Estas reglas ayudan a asegurar que los documentos tienen una estructura coherente y lógica. Las etiquetas marcan el comienzo y el fin de cada parte de la estructura. En el siguiente ejemplo, "**<Título>**" indica comienzo de un título, y "**</Título>**" indica el final

De acuerdo con el mismo autor, SGML se compone de tres piezas básicas: - Elementos, que se declaran mediante etiquetas. Toda la información contenida en un documento SGML está anidada en estas etiquetas, aunque, dependiendo de los elementos, éstos pueden contener la información directamente, o bien otros subelementos. Una DTD controla las funciones y uso de los elementos. - Atributos, que aparecen dentro de las etiquetas de inicio, y son modificadores de los elementos. También vienen controlados por la DTD. -Entidades, que son sistemas abstractos para describir cosas reales, por ejemplo, juegos de caracteres. El procedimiento de crear una referencia abstracta a un archivo externo es más fiable que utilizar el archivo externo mismo.

XML

La definición de XML es ambigua: XML es una subserie de SGML. Nació, por una parte, para solventar las irregularidades que los grandes navegadores habían introducido en HTML (Hypertext Markup Language), y que lo estaban convirtiendo en algo así como un lenguaje propietario según que navegador se utilizara; y por otra, para etiquetar “informativamente” los documentos web, es decir, para clasificar su contenido, no para facilitar exclusivamente la navegación, tal y como hace HTML.

XML no es una serie predefinida de etiquetas, del tipo de las definidas para HTML, que puedan utilizarse para marcar documentos, ni una plantilla normalizada para producir tipos determinados de documentos; y tampoco fue diseñado como un modo normalizado de codificar texto. Básicamente, esto significa que XML sólo define la gramática que deben satisfacer otros lenguajes de etiquetado para construir documentos XML válidos, y que, por tanto, mediante XML se garantiza la interoperabilidad y la compatibilidad de muchos potenciales lenguajes: desde un lenguaje para recetas de cocina hasta un lenguaje para instrumentos de descripción archivísticos.

Descripción archivística: norma ISAD (G)

La actividad de normalización descriptiva en archivos se afianza con la adaptación de capítulos de las AACR2 para la descripción de archivos en la década del 80 en EEUU. La evolución de los proyectos de normalización nacionales de Estados Unidos, Gran Bretaña y Canadá influyeron decisivamente, como en el proceso internacional desarrollado por el Consejo Internacional de Archivos (CIA) que culmina en 1994 con la aparición de la primera edición de la Norma Internacional General de Descripción Archivística ISAD(G).

La principal aportación de la norma es la estandarización de la estructura de datos, es decir, determina cuáles son los datos descriptivos que puede contener una descripción (un total de 26 elementos) y estructura jerárquicamente dichos datos en siete áreas: el área de mención de identidad o identificación, que contiene información mínima descriptiva; el área de contexto, empleada para recoger los datos provenientes del análisis contextual; el área de contenido y estructura, para los elementos relacionados con el análisis de contenido; las áreas de condiciones de acceso y utilización, de documentación asociada, de notas y de control de la descripción para información descriptiva proveniente de lo que se podría denominar "análisis externo".

La variabilidad con respecto a los niveles de descripción está solventada mediante la utilización de las denominadas reglas de descripción multinivel. La descripción multinivel consiste en llevar a cabo la descripción de un fondo y todas sus partes componentes (secciones, series, expedientes, documentos y todas las categorías intermedias), utilizando siempre los elementos apropiados para cada unidad de descripción y relacionando las descripciones resultantes de forma jerárquica.

A cada nivel de clasificación (agrupación) documental le corresponde un nivel de descripción. Las descripciones resultantes, relacionadas jerárquicamente, suponen la representación de la estructura de clasificación mediante la descripción.

La ISAD (G) establece un grupo de 5 niveles de descripción que van de lo general a lo específico: nivel de fondo (fonds), nivel de subfondo (sub-fonds), nivel de serie (series), nivel de expediente (file) y nivel de unidad documental (item). No obstante, la norma no obliga el nivel de descripción al que se debe llegar. Para una correcta aplicación de la descripción multinivel, la ISAD (G) propone cuatro reglas generales: regla de descripción de lo general a lo

particular, regla de información pertinente para el nivel de descripción, regla de vinculación de las descripciones y regla de no repetición de la información.

La ISAD ha supuesto un "paso de gigante" hacia la necesaria búsqueda de rentabilidad en la explotación de los recursos archivísticos, ya que permite compartir, a nivel internacional, una misma estructuración descriptiva, lo que posibilita el intercambio. Sin embargo, el contenido de los elementos descriptivos que componen dicha estructura no está normalizado y ello limita las posibilidades de su aplicación en un entorno digital para poder garantizar una recuperación de la información archivística más eficiente y precisa, lo que conlleva también problemas en la representación de interrelaciones en la información descriptiva, sobre todo en la propiedad de herencia de propiedades entre nivel jerárquicos superiores e inferiores y en las posibilidades de navegación a lo largo de esta estructura, todo lo cual dificulta la indización y posterior recuperación de elementos particulares.

Debido a todos estos problemas es que surgió la necesidad de crear una representación digital de estos instrumentos de descripción archivística, de forma tal que se pudiera representar toda esta información descriptiva altamente estructurada en un formato codificado mediante un lenguaje de etiquetado o marcado normalizado.

Control de autoridades: ISAAR (CPF)

Es la norma internacional sobre registros de autoridad para instituciones, personas y familias. El objetivo fundamental de esta Norma es ofrecer reglas generales para la normalización de las descripciones archivísticas en relación a los productores de los documentos de archivo

Una normalización de la información sobre los productores de los documentos permite:

- El acceso a los archivos y sus documentos, a través de una descripción de su contexto de producción relacionada con las descripciones de los propios documentos que, en muchos casos, son diversos y se encuentran dispersos físicamente
- La producción de registros de autoridad archivísticos que puedan garantizar puntos de acceso consistentes
- Identificar de una manera precisa los productores de los documentos, incorporando también la descripción de las relaciones entre las diferentes entidades, especialmente aquella relación relativa al cambio administrativo en las instituciones o el cambio de circunstancias en individuos y familias
- Crear la posibilidad de intercambiar y compartir registros de autoridad de interés de más de una institución.

El concepto de "registro de autoridad" proviene del área de bibliotecas pero, la propia norma, expone que los registros de autoridad de archivos deben responder a un número mayor de exigencias que los registros de autoridad de bibliotecas, sobre todo debido a la importancia que en los sistemas de descripción archivísticos tiene la información sobre el productor y el contexto de producción de los documentos.

En la década de 1990 se produjeron importantes avances en la normalización de las pautas que deben guiar la descripción archivística. Estos esfuerzos chivists, que combina APPM con ISAD(G), establece los elementos que deben usarse en la redacción de descripciones e incluye ejemplos de su utilización en descripciones codificadas mediante EAD y MARC.

EAD

El resultado del proyecto, iniciado en 1993 por Daniel Pitti (1997) en la Biblioteca de la Universidad de Berkeley, fue el diseño de una DTD que definía una clase de documentos *finding aids* (instrumentos de descripción) que, en términos generales, constaban de una página de título opcional, la descripción del material archivístico y unos apéndices también opcionales. La página de título incluiría variados elementos como la identificación del material o el tipo de instrumento de descripción. La descripción, propiamente dicha, constaría de una breve descripción del material descrito, incluyendo elementos etiquetables como título, fechas, alcance y contenido, El diseño de la versión beta de EAD puso rápidamente de manifiesto algo que, de forma natural, ocurrió en muchas otras aplicaciones SGML. Las ventajas de uso de SGML son enormes, pero debido precisamente a esto, las posibilidades de implementación con la tecnología actualmente operativa son nulas. Era indispensable aprovechar las capacidades de la Web, para ello podría haberse utilizado HTML (que nació como una DTD SGML). No obstante, HTML es un lenguaje de anotación genérico y además híbrido que especifica muy poco sobre la estructura del documento, combina formato y contenido al mismo tiempo, porque a un nivel bibliográfico se dificulta la aplicación de metadatos Por todo ello, las últimas versiones de EAD, emplean vocabulario XML, una adaptación de SGML para posibilitar su uso en la Web. Es su primera versión del año 1998 ya se contemplaba emplear, entonces, XML de la misma manera que en la última versión del año 2002. La DTD como la documentación asociada, manuales, guías es accesible a través de la web de la Library of Congress con el respaldo de la Society of American Archivists (<https://www.loc.gov/ead/>)

Los componentes de la DTD se dividen en distintos módulos, con otros cuatro archivos asociados que funcionan con una entidad, ead.dtd, ead.base.ent (entes SGML para EAD), eadnotat.ent (archivos de imágenes), eadchars.ent (caracteres), y eadsgml.dcl (elementos, atributos en SGML). La Society of American Archivists mantiene distinta documentación que funciona como guías de ayuda, preguntas comunes, aplicaciones, etc.

EAD es “una norma de codificación en Standard Generalized Markup Language (SGML) diseñada específicamente para marcar información contenida en instrumentos de descripción archivísticos. Con ello se quiere decir: -Es un lenguaje que se adecua a la gramática SGML (y, por añadidura, es compatible con la gramática XML). En su segunda versión, no obstante, la DTD ha sido escrita de acuerdo con XML. -No es, en sentido estricto, un lenguaje de descripción de materiales archivísticos, sino un lenguaje de codificación, y, en esta medida, destinado a ser interpretado por máquinas para devolver una visualización legible por humanos. -Es un lenguaje para codificar instrumentos de descripción, no los materiales archivísticos en sí mismos.

La finalidad del proyecto era investigar la factibilidad de una norma de codificación no-propietaria, para instrumentos de descripción legible por máquina y procedente de distintos depósitos culturales. Los directores del proyecto eran conscientes del papel creciente de las redes a la hora de acceder a los fondos, y estaban dispuestos a proporcionar información más allá de la ofrecida por los tradicionales registros MARC. Los criterios para el desarrollo de esta norma fueron:

- Capacidad para representar información descriptiva heredada por un nivel jerárquico de otro; -capacidad para moverse dentro de una estructura informativa jerárquica y; -soporte a la indización y recuperación de elementos específicos. SGML fue elegida como herramienta de trabajo debido a la relevancia de algunas de sus características
- Capacidad para representar entidades archivísticas y sus diferentes atributos en un formato de codificación de alcance internacional que permita el intercambio de información entre instituciones
- El lenguaje está formado por una serie de reglas que definen y expresan la estructura lógica de los documentos, es un formato de estructura de datos, no una norma sobre el contenido de los mismos, es decir, se

- focaliza en el contenido estructural de la descripción archivística, no en su presentación
- Capacita por tanto a los productos de software para controlar la búsqueda, recuperación y visualización estructurada de esos documentos.
 - Las reglas se aplican en la forma de etiquetas de marcado que pueden anidarse en un documento electrónico para identificar y establecer relaciones entre partes estructurales.
 - Puesto que un marcado coherente de documentos estructurados de manera similar es clave para el éxito del procesamiento electrónico, SGML estimula la coherencia introduciendo el concepto de definición de tipo de documento (DTD). Una DTD prescribe la serie ordenada de etiquetas de marcado SGML disponibles para codificar las partes de los documentos de una clase similar.
 - Los instrumentos de descripción archivísticos, que comparten partes y estructuras similares, forman una clase de documentos para las que se puede desarrollar, y de hecho se desarrolló, una DTD.

La versión de marzo de 1995 de la DTD Berkeley Finding Aid Project (BFAP o FINDAID DTD) definió una clase de documentos que, en general, constaban de una portada opcional, la descripción de una unidad de material archivístico, y una contraportada opcional. En 2002 se lanza la segunda edición y todos sus elementos son compatibles con la norma ISAD(G), se añaden nuevos elementos respecto de la primera versión y se produce una reestructuración en un orden más lógico

Estructura

De forma general, la EAD es una estructura de datos normalizada que reproduce en formato digital los instrumentos de descripción archivística.

La DTD ha sido diseñada para reflejar la jerarquía natural que presenta la organización de los fondos, subfondos, series, expedientes e items en conjunción con la jerarquía intelectual que imponen los archiveros con sus prácticas descriptivas. En esencia, EAD determina los tipos de elementos utilizables y el contenido posible de cada elemento particular. Desde la primera versión se definen un conjunto de 145 elementos, cada uno con un nombre y un identificador unívoco, para cada elemento se debe aclarar si su uso es obligatorio u opcional, si es repetible, es decir, si puede presentar ocurrencias, si incluye componentes con un orden de presentación específico, etc.

El equipo acordó que, en su nivel básico, el documento de un instrumento de descripción consta de dos segmentos:

- Un segmento que proporciona información sobre el instrumento de descripción mismo (en su versión definitiva, este segmento será el elemento, **eadheader** que permite básicamente identificar la instancia de documento entre el conjunto de documentos web, tal como lo haría, por ejemplo, Dublin Core, o la cabecera codificada de cualquier otro documento),
- Otro segmento que proporciona información sobre el cuerpo del material archivístico (en la versión definitiva, este segmento vendrá definido por el elemento de nivel superior, **archdesc** y sus subelementos anidados).

El EAD contiene, entonces, varios tipos de elementos: aquellos que codifican puntos específicos en la descripción de partes componentes del instrumento de descripción o el material que describe (elementos descriptivos - título de la unidad, fecha de la unidad, productor, etc.-); los utilizados para el acceso (nombre de entidad `corpname-`, nombre de persona, etc.); de enlace y aquellos que podrían codificar cualquier característica del documento (elementos genéricos). Se establece un número de 145 elementos, estableciendo para cada

uno de ellos un nombre completo y un identificador genérico. La especificación del contenido de cada elemento (modelo de contenido en la DTD) significa aclarar si es de uso obligatorio u opcional, si es repetible o no, si incluye componentes que deben presentarse en una secuencia concreta, etc.

- Información jerárquicamente organizada que describe una unidad de registros, junto con sus partes o divisiones componentes (es decir, el instrumento de descripción);
- Información adjunta que quizá no describe directamente registros, pero que facilita su uso a los investigadores (es decir, posibles ficheros asociados al instrumento de descripción).

La jerarquía de la información descriptiva, reflejando los principios de la organización archivística, comienza con un sumario del todo y procede a la delineación de las partes como una serie de visiones contextuales. Las descripciones de las partes heredan la información de las descripciones del todo.

Las descripciones de las partes heredan la información de las descripciones del todo. Este acuerdo hizo posible que se re-evaluaran los elementos de codificación del modelo BFAP. Los elementos que sobrevivieron a la evaluación formaron dos categorías:

- Elementos descriptivos, o aquellos que serían etiquetados en puntos específicos y predecibles de la descripción de unidades o partes componentes;
- Elementos genéricos, o aquellos que serían etiquetados en cualquier lugar dentro del documento.

Como dijimos, la DTD también declara los componentes que se van a utilizar en el marcado de instancias de documentos. Estos componentes son básicamente los elementos y atributos, aunque la DTD define igualmente entidades externas e internas, cuyo significado y función veremos algo más adelante. Los elementos se definen como los bloques principales de codificación que especifica la DTD; se representan mediante etiquetas y proporcionan el marco necesario para la codificación de una instancia de documento. Elementos son, por ejemplo, Resumen, Fecha, Autor, Acceso de nombre personal; cada uno de ellos cumple una función, especificada en el Repertorio de etiquetas¹³⁵, y cada uno de ellos tiene una representación única mediante etiquetas

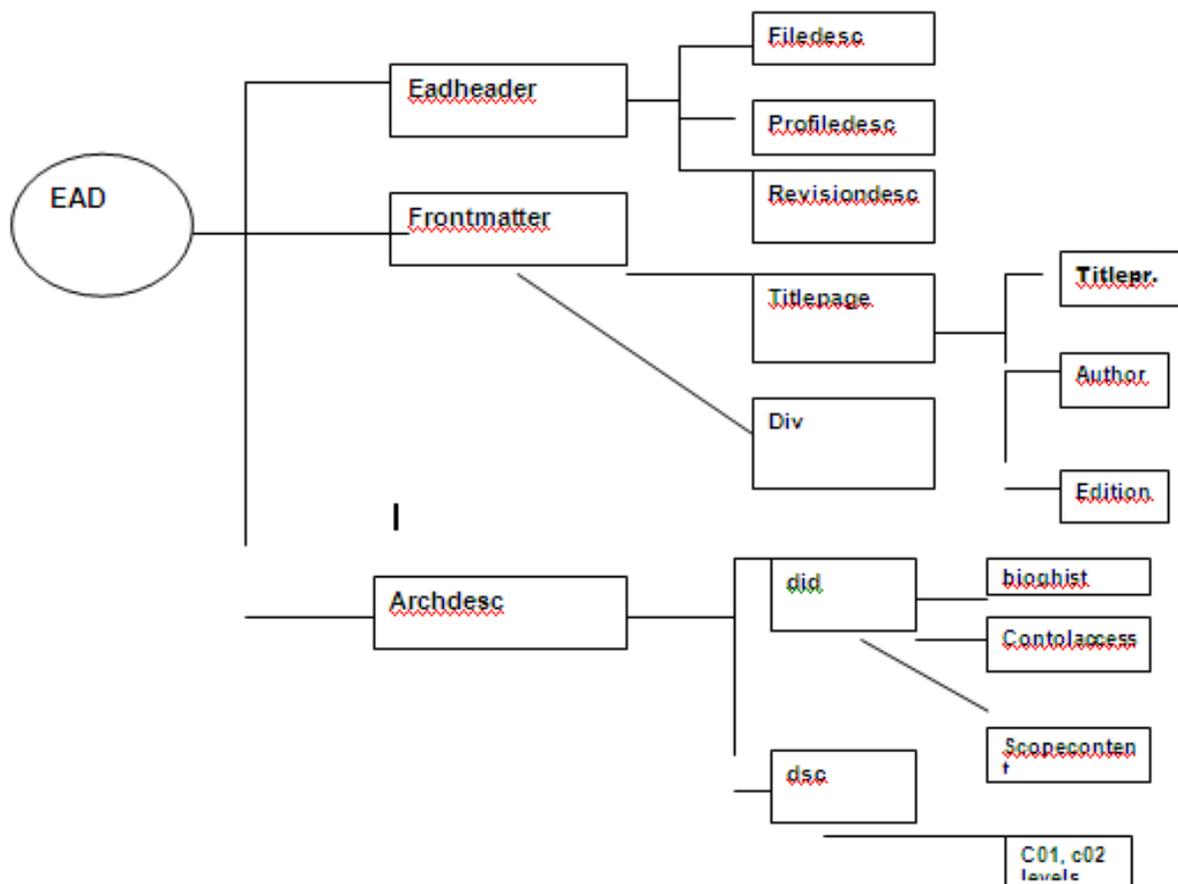
En cuanto a su sintaxis, la declaración de un elemento en la DTD está formada por:

- La cadena ELEMENT
- El nombre del elemento, element_name
- Un modelo del contenido: content_model
 - ¡ELEMENT, cadena constante para designar elementos
 - Ead nombre del elemento
 - (Eadheader, frontmatter, archdesc) es su modelo de contenido

Un elemento EAD tiene las siguientes características:

- El tipo de información que puede existir dentro de un elemento
- El orden que debe establecerse en dicha información
- La frecuencia con la que pueden aparecer subelementos dentro del elemento

Los principales componentes informativos que constituyen el modelo se pueden representar mediante una estructura arbórea que refleja claramente las relaciones jerárquicas entre los elementos.



EAD usa el término instrumento de descripción que en la primera versión se denominaba **finding aids**, para referirse a cualquier herramienta jerárquica que haya sido codificada usando EAD y que permitirá a un creador de registros o a un usuario acceder a los materiales que están siendo descritos.

A un nivel muy básico, un documento "instrumento de descripción" codificado utilizando EAD, consta de tres segmentos: uno que proporciona información sobre el instrumento de descripción en sí mismo (su título, compilador, fecha de compilación), **<eadheader>**; un segundo componente que incluye las cuestiones preliminares necesarias para la publicación formal del instrumento de

descripción, **<frontmatter>**; y un tercero que proporciona la descripción del material archivístico en sí misma, además de la información contextual y administrativa asociada, **<archdesc >**.

El **<eadheader>** abarca cuatro subelementos (**<filedesc>**, **<profiledesc>**, **<revisiiondesc>** y **<requiredft>**) para capturar o contener la mayoría de la información que normalmente es registrada sobre la creación, publicación y la distribución de un documento "instrumento de descripción". Otra información adicional sobre el instrumento de descripción puede ser recogida en el elemento **<frontmatter>** que, incluida en dos elementos como **<titlepage>** y **<div>**, refleja detalles de tipo introductorio (prefacio, introducción, etc.) necesarios, como ya hemos indicado, para la publicación formal del instrumento de descripción.

En el elemento **<archdesc >** pueden presentarse dos tipos de información: de tipo descriptivo y complementaria, de tipo bibliográfico y administrativo. El elemento "descripción archivística" o **<archdesc>** presenta información jerárquicamente organizada que describe una unidad de registros junto con sus partes componentes o divisiones. Muchos elementos EAD pueden contener un elemento **<head>**, para incluir información de "entrada", **<archdesc>**, puede presentar, además, el elemento **<runner>**, que permite incluir una cabecera para el conjunto de los datos descriptivos. Estos datos descriptivos se reúnen en un elemento contenedor (**<did>**) llamado identificación descriptiva.

EAD permite incluir, también, información suplementaria opcional que no describe directamente los registros pero facilita su uso por parte de los investigadores ,por ejemplo, una bibliografía (en la versión 1.0 los elementos que estructuraban esta información eran agrupados en el elemento "complemento a los datos descriptivos" o **<add>**); e información de descriptiva pero de tipo "administrativo", como detalles sobre el ingreso, datos de valoración o condiciones de acceso (información representada por elementos agrupados en **<admininfo>** de la versión 1.0).

Los descriptores (de nombre, de familia, geográficos, de materia, etc.) pueden ser representados en EAD con los elementos <name>, <famname>, <geogname>, <subject>, etc. agrupados en el genérico <controlaccess> (Controlled Access Headings).

Para enlazar un instrumento de descripción a copias digitalizadas de los materiales descritos o a cualquier otros recurso, EAD proporciona el elemento <dao> (Digital Archival Object), que lógicamente facilita la navegabilidad. Los mecanismos de enlace EAD son compatibles con XML Linking and Addressing Language, pero no se ajustan a la especificación. La localización del dao se especifica mediante <daoloc>, y ambos pueden estar vacíos o contener información que describa el dao contenida en el elemento <daodesc>. Para otro tipo de enlaces, como el que conectara con otro instrumento de descripción se emplea el elemento <extref> (extended reference), que estaría sujeto a condicionantes semejantes a los de dao. El componente dao es subelemento de did, como los más importantes elementos de tipo descriptivo como identificación de la unidad (<unitid>), título de la unidad (<unititle>), fecha (<unitdate>), descripción física (<physdesc>), etc.

Los subelementos de *did*, naturalmente pueden presentar atributos que cualifican su contenido como *label*, *type* o *encodinganalog*. Este último atributo permite la compatibilidad explícita con el ISAD(G) y otros formatos, puede relacionar cada elemento descriptivo con el correspondiente elemento ISAD o campo MARC, lo que facilita la generación de registros ISAD o MARC desde archivos EAD y potencia la accesibilidad, además de mejorar las posibilidades de indización.

De cualquier forma, en la edición EAD 2002, se explicita claramente la compatibilidad entre la codificación EAD y el modelo normativo ISAD (de hecho, en las directrices de aplicación de la versión 1.0 ya se recogía un listado de correspondencias entre elementos EAD y elementos ISAD). En un instrumento

de descripción codificado esta compatibilidad se puede poner de manifiesto expresamente mediante el atributo *relatedencoding*, en el elemento raíz *ead*, especificando en cada elemento descriptivo el elemento ISAD correspondiente mediante el citado atributo *encodinganalog*:. Mediante este elemento puede enlazarse descripción a elementos de páginas web, a registros en formato MARC u otros formatos en OPACs, a registros ISAD en bases de datos, etc.

Por otra parte, la EAD (Encoded Archival Description) ha abierto nuevas posibilidades a la representación de la vinculación entre los diferentes niveles de descripción archivística. Presenta un elemento, "descripción de elementos subordinados" *<dsc>* (en "archdesc") que permite incluir información sobre las unidades de descripción inferiores dependientes, jerárquicamente, de la unidad descrita.

Este elemento se utiliza de forma combinada con el elemento "componente" *<c>* que proporciona información particular sobre las unidades de descripción subordinadas y a su vez puede subdividirse en otros componentes tantas veces como así lo requiera la jerarquía de las unidades de descripción (*<c01>*, *<c02>*, *<c03>*...).

La descripción multinivel puede realizarse con el elemento *<dsc type="in-depht">*, con *analyticover*, para el nivel de subseries documentales, *combined*, para series y subseries documentales, *in-depth*, para información detallada y *other types*, para otros modelos.

Con el objeto de detallar el tipo de nivel, se recoge un atributo de nivel (como hemos visto anteriormente "level" en "archdesc"), que puede tomar nueve valores distintos: "collection", "fonds", "recordgrp", "series", "subgrp", "subseries", "file", "item" y "otherlevel".

Para indicar el tipo de estructura jerárquica subordinada, el elemento <dsc> puede tener un atributo "type" que puede tomar los siguientes valores: "analyticover" (estructura jerárquica que desciende hasta nivel de subserie), "combined" (desciende hasta la indicación del contenido de las series o subseries), "in-depth" (estructura jerárquica con información detallada) y "othertype" (otros modelos).

Otros elementos obsoletos que se eliminan o para los que se desaconseja su uso en la edición del 2002, son <add>.(datos adjuntos descriptivos), <admininfo > (información administrativa), <organization > (organización) y <langmaterial>, lenguaje del material

Aplicaciones

Al realizarse la descripción archivística de recursos digitales en este formato codificado, se pueden recopilar descripciones en listas en una página web, efectuar enlaces a registros MARC en catálogos en línea o sobre registros ISAD(G) en bases de datos distribuidas, entre otras posibilidades dada la flexibilidad del XML, ya que consta de un conjunto de módulos que permiten estructurar, almacenar e intercambiar información entre diferentes sistemas y formatos, al separar el contenido de la presentación de los datos en un formato estructurado. Al estar la última versión basada en XML, eso significa que para la presentación es necesario proporcionarles un estilo a través de XSL, el lenguaje para expresar las hojas de estilo, aplicable tanto a HTML como a XML, basado a su vez en XSLT (Extensible Stylesheet Language o Hoja de estilo extensible), XSL es, en general, la especificación de hojas de estilo para documentos XML. Más potente que las CSS y DSSSL, toma, sin embargo, elementos de ambos lenguajes, y su gestión resulta más compleja. Además, y a diferencia de CSS, no se aplica directamente sobre el árbol de código, sino que crea su propio árbol de salida. Por último, XSL tiene, al igual que DSSSL, capacidades de conversión. Un lenguaje de transformación usado para reacomodar, agregar y

eliminar etiquetas y atributos. Una plantilla XSLT permite variar la organización de la presentación de los elementos del documentos XML fuente, ordenar distintos elementos y subelementos según distintos parámetros, determinar que elementos estarán presentes en el formato de salida, lo cual es importante es una descripción multinivel de elementos, incluso crear elementos que no estaban presentes en el documento fuente XML, convertir un documento XML en otro formato, como el HTML, entre otras aplicaciones.

Otra posibilidad es proporcionar el estilo mediante plantillas CSS u hojas de estilo en cascada. Las CSS nacieron para permitir a los programadores modificar el estilo por defecto HTML. Básicamente permiten enunciar, en un archivo de texto, características tales como los márgenes, la sangría, el tipo de letra, la familia, el grosor o el color. Tienen la ventaja de que se pueden asociar a más de un documento, de tal manera que el estilo de éstos se gestiona en bloque desde un fichero independiente.

Mediante reglas un selector identifica o selecciona elementos de una página web, luego siguen un bloque de declaraciones entre llaves que constan de propiedades y valores para cada entidad descrita. Las plantillas CSS pueden ser externas o estar contenidas en un archivo HTML

EAC-CPF (Encoded Archival Context for Corporate Bodies, Persons, and Families).

Esquema XML basado en ISAAR(CPF) para la codificación y el intercambio de registros de autoridad.

La historia de EAC comenzó en 1998, con el esfuerzo de Richard Szary, Wendy Duff y Daniel Pitti para concebir una norma destinada a la codificación e intercambio de información de autoridades sobre el contexto de los documentos de archivo. Esta norma constituiría un estándar de comunicación para el

intercambio de registros de autoridad basados en la Norma internacional sobre los registros de autoridad de archivos relativos a instituciones, personas y familias (ISAAR(CPF)). Sería una norma paralela a EAD (Descripción archivística codificada), el estándar de codificación para instrumentos de descripción archivísticos. Al igual que EAD posibilita la expresión práctica de la Norma internacional general de descripción archivística (ISAD(G)), el nuevo estándar permitiría la expresión de ISAAR(CPF). Esta norma paralela preservaría y reforzaría la dualidad esencial que caracteriza la descripción archivística cuando se presenta en los instrumentos de descripción archivísticos. En EAC se utiliza el término “contexto” en lugar de “autoridades”. Esto es así porque no sólo se persigue la normalización de las distintas formas disponibles para un nombre (como sucede en el control de autoridades tradicional en el ámbito bibliográfico), sino que también se pretende describir las circunstancias bajo las cuales se generaron los distintos fondos, EAC propone la creación de registros contextuales separados, que posteriormente se vincularán a las descripciones, la utilización de una sintaxis común, xml, y la posibilidad de crear enlaces hipertexto para relacionar los documentos xml, permite estructurar un sistema de información donde descripciones de documentos y registros de autoridad/contexto quedan vinculados y permiten la navegación a través del sistema.

Un estándar distinto allanaría el camino para eliminar algunos problemas prácticos existentes en el uso de EAD, que se había desarrollado como una solución global para codificar instrumentos de descripción independientes (el modelo de presentación dominante), que incluyesen todo tipo de datos descriptivos sobre los documentos de archivo. Dado que en muchos fondos o depósitos se pueden encontrar documentos de archivo creados por (o relativos a) una misma entidad, existe mucho esfuerzo redundante en el registro de información sobre dicha entidad. Además, estos esfuerzos duplicados pueden provocar una gran inconsistencia, problemática tanto para los usuarios (en lo que respecta a la búsqueda e interpretación de los documentos de archivo)

como para los archiveros (en lo referente a la creación de referencias correctas y completas para dichas entidades). En conclusión, el EAC tiene por objetivo fundamental el solucionar el problema del control de autoridades, la forma normalizada y sus diferentes variantes, de los nombres de personas, familias y entes corporativos asociados a documentos de archivo, ya que el contexto de creación y uso de los documentos de archivo es complejo y multidimensional, pudiendo estar implicadas personas, familias, instituciones, sociedades, funciones, actividades, procesos, lugares, acontecimientos y otras entidades. Entre estas entidades, son esenciales los agentes responsables de la creación o uso de los documentos de archivo (por lo general instituciones o personas).

Su principal objetivo es normalizar la codificación de las descripciones sobre los agentes, para permitir el intercambio, descubrimiento y presentación de esta información en un entorno electrónico. EAC-CPF posibilita vincular la información sobre un agente con la de otros agentes para mostrar o descubrir las relaciones entre entidades creadoras de documentos de archivo, y también vincular dicha información con descripciones de documentos y otras entidades contextuales. EAC-CPF es una estructura de comunicación para la información contextual archivística de instituciones, familias y personas. Permite el intercambio de registros de autoridad conformes con ISAAR(CPF), ya que ISAAR(CPF) determina el tipo de información que puede incluirse en un registro de autoridad archivístico y sirve de orientación sobre la forma de integrar estos registros en un sistema de descripción archivística. Entre los diferentes datos que puede contener están:

1. La información contextual archivística consiste en información que describe las circunstancias en que los documentos de archivo han sido creados, mantenidos y usados. Este contexto incluye (pero no está limitado a) la identificación y características de las instituciones, personas y familias (agentes) que han sido creadores, usuarios o materias de los documentos de archivo, así como las relaciones entre ellos.

2. La información contextual sobre los agentes no son datos que describen otros recursos de información, sino datos que describen entidades que forman parte del entorno en que esos recursos de información (por ejemplo, documentos de archivo) han existido.
3. El registro de información contextual sobre los agentes directamente en los sistemas de información archivísticos permite una descripción y comprensión más completa de los documentos de archivo, facilitando además su recuperación por procedencia a través del tiempo y de diferentes dominios.
4. . La información contextual sobre los agentes también puede tener valor como un recurso de información independiente, separada de su utilización de apoyo a la descripción, recuperación e interpretación de los documentos de archivo.
5. Este modelo también está orientado a apoyar el intercambio y compartición de información contextual sobre agentes, especialmente en aquellos depósitos que tienen fondos con información contextual común o interés en compartir este tipo de datos.
Estructura y contenido. Cada instancia de información contextual sobre agentes describe una sola entidad.
6. El modelo proporciona un marco para registrar toma la gama y profundidad de la información contextual sobre agentes, y sugiere un juego mínimo de elementos para describir una entidad, pero deriva las recomendaciones para el uso apropiado de otros elementos a las directrices de aplicación desarrolladas para implementaciones específicas.
7. El modelo define un universo de elementos utilizados para describir los agentes y la estructura de interrelaciones entre dichos elementos. Estos elementos y estructura permiten el descubrimiento, navegación y presentación de la información contextual sobre los agentes y la vinculación de esta información con descripciones de documentos de

archivo o con otras entidades contextuales, especialmente aquellas codificadas con EAD, MARC y otros estándares similares.

8. El modelo permite la vinculación de descripciones de entidades contextuales con representaciones digitales u otros sustitutos de dichas entidades.
9. El modelo está expresado como un lenguaje XML para favorecer la independencia de la plataforma y la portabilidad de la información. El modelo también puede implementarse utilizando otros métodos.

Elementos y conceptos

Los agentes son complejos. Por ejemplo, un agente puede tener múltiples identidades, o bien una sola identidad puede estar asociada con varios agentes diferentes. Por ello, es necesario que la estructura de este estándar tenga en cuenta las distintas maneras de representación de los agentes. Para acomodar esta diversidad se han definido varios conceptos:

- IDENTIDAD INDIVIDUAL: una persona (o institución o familia) con una sola identidad representada en una instancia EAC-CPF. (El supuesto más frecuente).

- IDENTIDAD MÚLTIPLE-MUCHAS EN UNA: dos o más identidades (incluyendo identidades oficiales), cada una representada por descripciones diferentes dentro de una instancia EAC-CPF. Mediante programación se puede convertir en “Identidad Múltiple Una en Muchas”. (Supuesto menos frecuente aunque no raro).

- IDENTIDAD MÚLTIPLE-UNA EN MUCHAS: dos o más identidades (incluyendo identidades oficiales), cada una representada en dos más instancias EAC-CPF interrelacionadas. Mediante programación se puede

convertir en “Identidad Múltiple Muchas en Una”. (Supuesto menos frecuente aunque no raro).

- CONJUNTO ALTERNATIVO: instancia EAC-CPF derivada, que incorpora (y está basada en) dos o más instancias EAC-CPF alternativas de la misma entidad. Utilizable por un consorcio o un servicio que proporcione un acceso conjunto a los registros de autoridad mantenidos en dos o más sistemas por dos o más agencias. Las instancias EAC-CPF alternativas pueden estar en la misma o en distintas lenguas.

- IDENTIDAD DE COLABORACIÓN: una sola identidad compartida por dos o más personas (por ejemplo, un seudónimo colectivo utilizado en la creación de una obra de colaboración). Use “Identidad Múltiple-Una en Muchas”

Tiene un elemento raíz, que contiene dos elementos obligatorios (*eachheader*) —: con datos para el control de la descripción, su autor, fechas de creación, modificación, etc.

(*condesc*): incluye información sobre la entidad a la que se refiere el registro de autoridad: descripción, nombres autorizados y alternativos, relación con otros registros EAC y con otros recursos

La raíz va acompañada de un atributo *type* que recoge los valores *corporatebody*, *person* o *family* para indicar si el registro hace referencia a una institución, persona o familia.

Por otro lado el elemento (*condesc*) constituye la base del registro de autoridad y contiene los siguientes elementos hijos:

1. (*identity*): recoge los nombres usados por la entidad, tanto los autorizados como los que no lo están.
2. (*desc*): descripción de la entidad, su historia/biografía, contexto, funciones, etc.
3. (*eacrelcs*): referencias a otros registros de autoridad EAC para personas, familias u organizaciones relacionadas.
4. (*resourcerels*)—: referencias a recursos archivísticos, bibliográficos o museográficos relacionados
5. (*funactrels*)—: descripciones de funciones o actividades relacionadas con la persona, familia o institución en un formato normalizado (procedente de la lista de autoridades).}

Estos elementos contienen a su vez a otros mediante los cuales se detalla la información que contienen. Existe una equivalencia entre los elementos definidos en la Isaar(CPF) y las etiquetas EAC.

EAC-CPF ha sido creado para acomodar esta variedad de identidades, e incluye varias maneras de expresar las complejidades asociadas con las preferencias de un depósito o agregador de datos particular. En este Repertorio de etiquetas el término “agentes” se refiere a los depósitos o servicios que crean o mantienen registros EAC-CPF, mientras que el término “entidades” se refiere a las instituciones, personas o familias descritas en dichos registros.

Cada documento EAC-CPF contiene elementos que incluyen datos utilizados en el control de la descripción de la entidad y proporciona el contexto para esta descripción. El elemento contiene información sobre las estructuras del nombre, elementos descriptivos y relaciones.. Estos d elementos contenedores incluyen elementos específicos para respaldar los objetivos funcionales del elemento

padre o contenedor y elementos jerárquicamente relacionados. Este elemento se utiliza para incluir uno o más nombres usados por o para la entidad, y también para registrar dos o más nombres paralelos en diferentes lenguas o escrituras, permitiendo la operación de reunión de nombres que representan la misma entidad en varios idiomas.

La descripción de la entidad, como componente de la descripción archivística, debe ponerse en relación con los otros componentes descriptivos. Las descripciones de entidades deben estar dinámicamente relacionadas con las descripciones de documentos de archivo (a las que aquellas proporcionan contexto), y con las descripciones de funciones y actividades (realizadas por dichas entidades y testimoniadas en esos documentos). Las entidades se relacionan con otras entidades, con funciones y actividades, y con documentos de archivo. Asimismo, las funciones y actividades se relacionan con otras funciones y actividades, con creadores de documentos y con documentos de archivo. Y los documentos de archivo se relacionan con otros documentos, con entidades, y con funciones y actividades. Cada descripción de entidad, de documentos de archivo o de función/actividad puede por tanto funcionar como un nodo en una red de relaciones.

El esquema EAC-CPF posibilita el establecimiento de enlaces a recursos externos utilizando un subconjunto limitado del estándar XLink, definido en <http://www.w3.org/TR/xlink/>.

Los atributos XLink pueden usarse para crear y describir enlaces entre recursos. En particular pueden emplearse para referenciar un conjunto más rico de relaciones que aquellas que es posible registrar con los atributos @cpfRelationType (Tipo de relación con la institución, persona o familia), @functionRelationType (Tipo de relación con la función) y @resourceRelationType (Tipo de relación con el recurso).

Conclusiones

La EAD es una DTD definida en XML que permite la codificación de la información descriptiva de colecciones de archivos y su difusión y recuperación de información a través de redes digitales, permitiendo reflejar las descripciones multinivel características de este tipo de documentos, de forma tal que sea compatible con normas de aplicación internacional como la ISAD(G).

Al estar basada en XML se puede utilizar distintas alternativas para el enlace de partes o de todos los documentos, integrar imágenes de los documentos empleando Xlink, un modo estándar de crear hipervínculos multidireccionales a un archivo XML referenciados desde una variedad de elementos, así como se pueden usar aplicaciones de programas de interfase orientadas a objetos digitales para el diseño de diferentes opciones de formatos de salida, mediante el DOM, un conjunto estándar de llamadas a funciones para manipular archivos XML y HTML desde un lenguaje de programación.

También se debe tener en cuenta el papel del lenguaje RDF (Resource Descripción Framework) en el desarrollo de ontologías y aplicaciones para la implementación de la web semántica. RDF es un formato de texto XML que soporta aplicaciones de descripción de recursos y metadatos, e integra las aplicaciones de la denominada web semántica es decir, documentos con un etiquetado semántico mucho más preciso y profundo, que permiten búsquedas muy detalladas basadas en el etiquetado semántico, por ejemplo, la identificación de personas en un conjunto de fotos en un sitio web y su relación con otros sitios en los que también aparezcan, ese deben tener en cuenta la posibilidades de integración con documentos definidos en XML , ya que este lenguaje se basa en la posibilidad de crear un etiquetado más personalizado a medida, separando el contenido de la estructura de los datos .El uso conjunto de los formatos EAD y EAC permite además de una descripción codificada de documentos archivísticos en formato electrónico, el poder establecer enlaces

entre documentos y entidades, establecer un control de autoridades sobre las entidades responsables de estos documentos y establecer enlaces con otros recursos, lo que posibilita la filosofía de datos abiertos compartidos u Open Linked Data, el enlace entre distintos recursos con etiquetado semántico, lo cual es fundamental para el establecimiento de la web semántica. Este objetivo puede lograrse tanto a través del lenguaje RDF como con el establecimiento de Ontologías, formalizaciones conceptuales de un dominio que describe en lenguaje formal todos los conceptos y sus relaciones recíprocas de un dominio temático dado, lo que permite compartir información sobre recursos y lograr de esta manera, una recuperación de la información más eficiente.

Se han realizado en ese sentido proyectos para transformar el formato EAC-CPF en una ontología para convertir los registros de autoridad contenidos en este formato en datos abiertos enlazados, lo que permitiría compartir información sobre recursos e integrar la información dentro de un lenguaje formal y normalizado, para lo cual se requiere:

- Identificar los recursos mediante URLs desreferenciables
- Proveer información sobre los recursos y sus relaciones en un lenguaje normalizado que permita reconocerlos inmediatamente
- Incluir en la descripción de los recursos la mayor cantidad de enlaces relevantes hacia otros recursos de información

Para lograr estos objetivos se transformaron los elementos y atributos del formato y sus relaciones en conceptos y propiedades de la ontología pero expresados en RDF, definiendo como clases separadas los tres tipos de entidades posibles, personas, familias y entes corporativos, con el propósito de lograr una riqueza semántica mayor, se buscó además, facilitar el enlace con otros recursos a través de URLs alternativos o con enlaces hacia otros sitios

web y con archivos de autoridad disponibles en la web, como el Archivo de Autoridad Virtual Internacional (VIAF).

Este tipo de investigaciones muestra que es posible expresar la información incluida en los formatos de descripción archivística en el lenguaje necesario para el etiquetado semántico y para el establecimiento de la web semántica y lograr conectar la información normalizada sobre recursos archivísticos con la información sobre entidades responsables a un nivel internacional, más amplio lo que permite la conexión con recursos y entidades similares, lograr la recuperación de la información por tipo de recursos o por tipo de entidades semánticamente relacionadas y lograr conectar la información de documentos archivísticos con la otras instituciones, como bibliotecas, museos etc., o con otros documentos relacionados con las entidades vinculadas a las colecciones archivísticas.

Bibliografía

Alvite Diez, L. (2014). Metadatos en el contexto archivístico: El reto de la gestión y conservación de documentos electrónicos. En: Jornadas Archivando la Nueva Gestión de Archivos, Leon, 6 y 7 de noviembre 2014

Delgado Gómez, A. (2004). *Normalización de la descripción archivística: Introducción a Encoded Archival Description (EAD)*, 2ª.ed. Cartagena: 3000 Informática.

Lamarca Lapuente, M. J. (2017). Metadatos, Hipertexto: El nuevo concepto de documento en la cultura de la Imagen. Disponible en: <http://www.hipertexto.info/documentos/metadatos.htm>

Mazini, S.: Rucci, F. (2011). *EAC-CPF Ontology and Linked Archival Data*, Proceedings of the 1st International Workshop on Semantic Digital Archives (SDA 2011), Disponible en: <http://ceur-ws.org/Vol-801/paper6.pdf>

Mendez Rodríguez, E. M. (2013). “La descripción de documentos electrónicos a través de metadatos: una visión para la archivística desde la nueva e-Administración”. En: Revista d´Arxlus, 47-82

Nogales Flores, T. (2002). XML aplicado a los instrumentos de descripción archivística: EAD (Encoded Archival Description) versión 2002 Disponible en: www.bib.uc3m.es/~nogales

Peis, E.; Rodríguez Ruiz, A. (2004). *EAD (Encoded Archival Description): Desarrollo, Estructura, uso y aplicaciones*, Hipertext.net, nº2, (2004), disponible en; <https://www.upf.edu/hipertext/numero-2/ead.html>

Vázquez Paulus, C. Metadatos: Introducción e historia. Diponible en: <https://users.dcc.uchile.cl/~cvasquez/introehistoria.pdf>