

# Recursos informativos en las redes

## Bases de datos bibliográficas y xml: el caso Medline

Por Núria Ferran

La *National Library of Medicine* de los EUA apuesta por xml (con la correspondiente dtd) para producir su base de datos bibliográfica *Medline*, la más utilizada en el mundo de la medicina.

Como es sabido, xml (*extensible markup language*) es un lenguaje de marcas que define dentro del mismo documento electrónico los contenidos y la estructura de

los datos. Mientras que html cuida del aspecto, xml pone el énfasis en el significado. Es la definición del tipo de documento (*document type definition*, dtd) la que indica cómo están estructurados los datos dentro del documento.

La decisión forma parte de un proyecto de modernización de la *NLM*, que se engloba, a su vez, en la planificación del gobierno fede-



Núria Ferran

### DTD de Medline

ver:

<http://www.nlm.nih.gov/databases/dtd/nlmmedline.dtd>

<!-- NLM Medline DTD

This is the current DTD which NLM has written for

External Use. If you are a data Licensee, use the information from the MedlineCitation Set.

Comments and suggestions are welcome.  
March 19, 2001

-->

<!-- ===== -->

<!-- NLM Medline DTD -->

<!-- Typical usage:

<!DOCTYPE MedlineCitationSet PUBLIC "-//NLM//DTD NLM//EN">

-->

<!-- ===== -->

<!-- Revision Notes Section

The following changes were made in the nlm-medline\_010319.dtd:

- The entity reference changed to: nlmmedlinecitation\_010322.dtd.
- Added PMID entity

- The entity reference changed back to nlmmedlinecitation\_010319.dtd.

- Added GrantID, Agency, Acronym entity references

- Added MedlineID entity

-->

<!-- ===== -->

<!ENTITY % ArticleTitle.Ref "ArticleTitle">

<!ENTITY % ISSN.Ref "ISSN?">

<!ENTITY % DateCreated.Ref "DateCreated">

<!ENTITY % PubDate.Ref "PubDate">

<!ENTITY % PMID.Ref "PMID">

<!ENTITY % MedlineID.Ref "MedlineID">

<!ENTITY % GrantID.Ref "GrantID?">

<!ENTITY % Agency.Ref "Agency">

<!ENTITY % Acronym.Ref "Acronym?">

<!-- ===== -->

<!-- Reference to Where the NLM MedlineCitation DTD is located -->

<!ENTITY % MedlineCitation PUBLIC "-//NLM//DTD MedlineCitation, 19th March 2001//EN"

"nlmmedlinecitation\_010319.dtd">

%MedlineCitation;

<!-- ===== -->

<!ELEMENT MedlineCitationSet (MedlineCitation\*, DeleteCitation?)>

<!ELEMENT DeleteCitation (MedlineID+ | PMID+)>

<!-- ===== -->



## OC 2.0 GESTIÓN DE ARCHIVOS

Producto abierto, Parametrizable, Cliente/Servidor, Bases de Datos O.D.B.C. (Access, Oracle, SQL Server...etc.)

Compatible con todo tipo de Sistemas Operativos y Redes.

Integrado con IADOC-G.E.D. e IADOC-REGISTRO E/S

- ✓ Gestión de Entidades o Empresas del Sistema, de Funciones, Organigrama, Centros de Trabajo, Archivos, Depósitos.
- ✓ Gestión de Fondos, Secciones, Series, Tipos Documentales.
- ✓ Gestión de Usuarios y Permisos.
- ✓ Cuadro de Clasificación, Conservación y Expurgo.
- ✓ Generador de Fichas Descriptivas (incluida ISAD-G) e Índices Auxiliares.
- ✓ Descripción Multinivel.



- ✓ Transferencias.
- ✓ Consultas (Argumentos relacionales y/o documentales).
- ✓ Gestión de Préstamos.
- ✓ Traslados y Eliminaciones.
- ✓ Gestión de Espacio en Depósitos.
- ✓ Informes y Estadísticas.
- ✓ Gestión de Documentos Electrónicos y Digitalizados.



Innovación en Archivos y Documentación, S.A.

C/ Arroyofresno, nº 19, 2ª Izda. (Puerta de hierro) / 28035 Madrid \*\* Tfno.: 91/373 76 25 - Fax.: 91/386 65 23  
<http://www.iad.es> - e-mail: infoiad@iad.es

ral estadounidense para utilizar la sintaxis xml para intercambiar información por internet<sup>1</sup>.

**«La nueva versión de *Medline* en xml se beneficia de la principal característica de este estándar: la capacidad de intercambiar cualquier información separando los contenidos de la presentación. Con este paso en firme que ha realizado la *NLM* se vislumbra algo que en poco tiempo puede convertirse en habitual»**

*NLM* fue precursora en la adopción de xml dentro del gobierno norteamericano, según su director de sistemas de información, **Simon Liu**, estableciendo unos es-

tándares para los editores que enviaban información bibliográfica para la base de datos *Medline*. Con este primer paso se integraban fácilmente datos de las fuentes más dispares. Desde enero de 2001 se realiza un paso más, entregándose a los distribuidores sólo en xml junto con una dtd para definir su estructura.

<http://www.nlm.nih.gov/bsd/licenses.html>

Tal como afirma **Roy Tennant**, las bibliografías no son más que conjuntos de elementos textuales estructurados, y precisamente xml está pensado para esto<sup>2</sup>.

Con la transición al nuevo formato la *NLM* aprovechó para realizar algunos cambios en la organización de sus datos, incorporando nuevos elementos, como por ejemplo el campo autor corporativo. Además *Medline* se ha expandido en 2001 en alcance y cobertura<sup>3</sup>: contiene más de 11 millones de re-

gistros bibliográficos extraídos de 40.000 publicaciones periódicas relacionadas con las ciencias de la salud procedentes de 70 países.

La *NLM* ha anunciado que este apoyo a xml no se detiene en *Medline*, sino que ya tiene planificados muchos más proyectos. Se han definido más dtds para otras bases de datos como *Toxnet* (reinención de *Toxline*) y para el vocabulario controlado *MeSH* (*Medical subject headings*).

Pueden teledescargarse muestras del nuevo formato en la dirección:

<http://www.nlm.nih.gov/mesh/filelist.html>

Otro de los proyectos susceptibles de utilizar xml es la información monográfica que ahora se proporciona en Usmarc<sup>4</sup>. Xml demuestra ser un lenguaje muy válido para acomodarse a los datos bibliográficos complejos e integrar-

los en recursos web, de manera que los formatos tradicionales bibliotecarios como Marc tienen que valorarse frente a las ventajas de este lenguaje de marcas.

### **Preservación del contenido electrónico**

Xml soporta *Unicode*, que permite a la *NLM* utilizar caracteres especiales, diacríticos, etc., un aspecto muy importante dado el alcance global que tiene *Medline*. Otro aspecto ventajoso del conjunto de caracteres *Unicode* es su neutralidad en cuanto a plataformas, y por tanto ofrece las posibilidades más seguras de preservación del contenido electrónico a lo largo del tiempo. Como no está ligado a ningún hardware, software o protocolo de red, los datos en formato xml seguirían siendo vigentes aunque cambiaran las plataformas<sup>4</sup>.

El proceso de apertura se complementa además con el de migración de la base de datos *Medline* desde una implementación propietaria en un mainframe *IBM* hacia una base de datos relacional comercial en Unix<sup>5</sup>.

### **Notas**

1. *Xml.gov. Government makes the most of xml.* Data Conversion Laboratory. Consultado en: 26/05/01.

<http://www.dclab.com/eGovernmentxml.asp>

2. **Tennant, Roy**, Xml: "The digital library hammer", *Library journal*, 2001, v. 126, n. 5, p. 30.

3. *Overview of NLM data, format, policy and practice changes for licensees of NLM leased databases in 2001.* U. S. National Library of Medicine. Última actualización: 31-01-01.

[http://www.nlm.nih.gov/bsd/2001\\_overview.html](http://www.nlm.nih.gov/bsd/2001_overview.html)

4. **Miller, Dick**, "Xml: Libraries' strategic opportunity". En: *School library journal*, 2000, v. 46, n. 8, p. 18.

5. *NLM to use xml and dtd for Medline data.* U. S. National Library of Medicine. Última actualización: 13-10-00.

<http://www.nlm.nih.gov/news/medlinedata.html>

*Reseña enviada a EPI por Núria Ferran.*

[nferranf@campus.uoc.es](mailto:nferranf@campus.uoc.es)