GT2 – Calidad de descripción e indización LILACS

Interoperabilidad semántica y sintáctica para garantizar calidad de los datos provenidos de otras bases de datos

23 de abril de 2014

biblioteca virtual en salud

BIREME • OPS • OMS

Objectivo de la sesión

- Aclarar que significa interoperabilidad, su importancia, a quién se destina y los procedimientos que los CC necesitan que sean candidatas a interoperabilidad de datos
- Dar ciencia a los CC que para interoperabilidad de datos de suceso, es necesario que el contenido de los campos **adopción efectiva de los estándares**.

Que es Interoperabilidad?

- “Interoperabilidad es la habilidad de los sistemas para trabajar juntos, en general **gracias a la adopción de estándares**.

- La interoperabilidad no es solamente la habilidad de intercambiar información sanitaria, sino que requiere la habilidad de **entender lo que se ha intercambiado**”

(Institute of Medicine, 2004 apud Indarte, 2012).

Reuso de metadatos

Quién puede hacer Interoperabilidad?

- Centros Cooperantes que adopten **sistemas distintos del LILDBI-Web** para gestión de su biblioteca y usa LILDBI-Web sólo para cooperación con LILACS realizando trabajo doble de descripción e indización.
- Ej:
  - Sistemas como Aleph, VTLS, Pergamum, u otro;
  - Repositorios institucionales basados en Dspace u otro.

Para que sirve hacer Interoperabilidad?

- **Optimización de recursos** por medio de reuso de los metadatos hechos por otra institución
- Para **evitar** trabajo doble!
- Publicación de los metadatos de manera mas eficiente para que estén listos para **recuperación en LILACS en menos tiempo**.

Puntos importantes para Interoperabilidad de datos

- URL para cosecha de datos via OAI-PMH provider
- Documentación acerca del sistema para chequeo de consistencia de los valores de los datos.
- Compatibilización de los campos de datos y contenido de los campos
- Pruebas de carga y contenido de los campos

Cuidados con la Calidad y Consistencia del Metadato

"…es evidente que lo que prevalece es la necesidad de garantizar la **calidad de la información** asistencial basándose en el principio:

de un dato, un significado".

(Indarte, 2012)

biblioteca virtual en salud
BIREME • OPS • OMS

Interoperabilidad Semática y Sintática

- Los estándares centrados en la comunicación proveen estructuras de datos que determinan un nivel **semántico** mínimo (**tipos, formatos, codificación, campos, tamaños y otros**) y proveen también alguna **sintaxis** para representar esas estructuras en un formato comunicable vía informática.

(Indarte, 2012)

Interoperabilidad Semática y Sintática

- El nivel **semántico** mínimo es necesario para que la información pueda ser enviada y recibida, pero **no garantiza la interpretación efectiva** de la información comunicada.

- Este tipo de interoperabilidad se denomina interoperabilidad sintáctica. Un concepto importante es que **sin interoperabilidad sintáctica no es posible implementar ningún otro tipo de interoperabilidad**.

(Indarte, 2012)

biblioteca virtual en salud
BIREME • OPS • OMS

Interoperabilidad Sintática

- **Sintaxis** se refiere a la **orden de los elementos** dispuestos para la representación de los recursos informacionales.

- La sintaxis del registro bibliográfico es la parte que volve su atención al estudio de la **forma, arranjo y disposición** en que cada elemento debe ser descripto.

(Zafalon e Santos, 2012)

“Herramientas” para la interoperabilidad

- Crosswalks: tabla o un gráfico que muestra las relaciones y equivalencias (y resalta las diferencias inevitables) entre dos o más formatos de metadatos.

CDWA	MARC	EAD	Dublin Core
Object/Work-Type	655 Genre/form	<controlaccess> <genreform>	Type
Titles or Names	24Xa Title and Title—Related Information	<unittitle>	Title
Creation–Date	260c Imprint—Date of Publication	<unitdate>	Date.Created
Creation-Creator-Identity	1XX Main Entry 7XX Added Entry	<origination> <persname> <origination> <corpname> <origination> <famname> <controlaccess> <persname> <controlaccess> <corpname>	Creator

(Woodley)

http://wiki.bireme.org/pt/index.php/Metadados_mínimos_para_Artigos_de_Periódicos_(S_as)

“Herramientas” para la interoperabilidad

ESQUEMAS DE CODIFICACIÓN		
ESQUEMAS DE CODIFICACIÓN DE CONTENIDO	Estándares de estructura de datos (*Data structure standards*)	• Esquema de metadatos: estructura descriptiva en directorios, campos, subcampos, código de subcampo, indicadores. Ej. MARC; LILACS
	Estándares de intercambio de datos (*Data exchange standards*)	• Protocolo ANSI/NISO Z39.50; Norma ISO 2709
ESQUEMAS DE CODIFICACIÓN DE VALORES	Estándares de contenido de datos (*Data content standards*)	• Estándares o códigos para descripción de información. Ex.: AACR, ISBDs, RDA
	Estándares de valores de datos (*Data value standards*)	• Tesauros, listas de autoridad, léxicos, esquemas de clasificación, vocabularios controlados

(Alves & Santos, 2013)

Inconsistencias en el metadato de: Autor

Tag <dc:creator>

- <dc:creator>
  Azad Khan, A K
  </dc:creator>
- <dc:creator>
  Hasan, M
  </dc:creator>



Tag <dc:creator>

- <dc:creator>
  Ribeiro,Jorge Pinto
  </dc:creator>
- <dc:creator>
  <ags:creatorPersonal>
  Ribeiro, Jorge **Pinto(Hospital** de Clínicas de Porto Alegre ,Universidade Federal do Rio Grande do Sul Faculdade de Medicina)
  </ags:creatorPersonal>



Inconsistencias en el metadato de: Autor

Tag <Author>

- <Author>
  <![CDATA[Lu Zhu]]>
- </Author>
  <Author>
  <![CDATA[Xing-hua Pan]]>
- </Author>

Tag <dc_creator>

- <dc_creator>
  Dávila González, Hermenegildo**,1846-1908.autor**
  </dc_creator>

Espacio en branco antes del nombre

Nombre estandarizado
Punto
“autor” al final

bvs biblioteca virtual en salud
BIREME • OPS • OMS

Inconsistencias en el metadato de: Autor

Tag Medline

- <LastName>
  de Carvalho
  </LastName>
  <ForeName>
  Leiliane GonÃ§alves
  </ForeName>
  <Initials>
  LG
  </Initials>

Autor con entrada incorrecta

Abreviatura del nombre incompleta y sin espacio

Abreviatura del nombre incompleta

- <dcterms:monographicAuthor>
  Barbosa, Tyara
  <dcterms:monographicAuthor>

En tags distintas

- <LastName>Teixeira</LastName>
  <ForeName>PatrÃcia de FÃ¡tima Dos **Santos**</ForeName>
  <Initials>**Pde F**</Initials>

- <LastName>Panico</LastName>
  <ForeName>Ana Luiza **BrandÃ£o Galotti**</ForeName>
  <Initials>AL</Initials>

biblioteca virtual en salud

BIREME • OPS • OMS

Inconsistencias en el metadato de:
Afiliación del Autor

- <Affiliation>

<Affiliation>
Faculdade de Medicina, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brazil.
</Affiliation>

Afiliación en orden invertida

<Affiliation>
LaboratÃ³rio de Endocrinologia Molecular, Centro de Pesquisas do Genoma J.F. Perez, Departamento de Morfologia, **scola** Paulista de Medicina, Universidade Federal de SÃ£o Paulo, SÃ£o Paulo, SP.
</Affiliation>

Ortografia incorreta

bvs biblioteca virtual en salud
BIREME • OPS • OMS

Inconsistencias en el metadato de:

Otras indicaciones de responsabilidad

Tag <dcterms:advisor>

- <dcterms:advisor xml:lang=""“>
  Amaral, Sueli Angélica do
  </dcterms:advisor>

Tag <dc:contributor>

- <dc:contributor>
  Ramos, Dalton Luiz de Paula
  </dc:contributor>





Inconsistencias en el metadato de: Título

Tag <dc:title>

- <dc:title>
  Peptic ulcer in Bangladesh.
  </dc:title>



Tag <dc:title>

- <dc:title>
  Em busca de novos paradigmas para o manejo da cardiopatia isquêmica
  </dc:title>
- <dc:title>
  Efeito dos ácidos graxos ômega-3 sobre o relaxamento-dependente do endotélio em coelhos hipercolesterolêmicos
  </dc:title>

Inconsistencias en el metadato de:
Título de la Revista

Tag <MedlineTA>

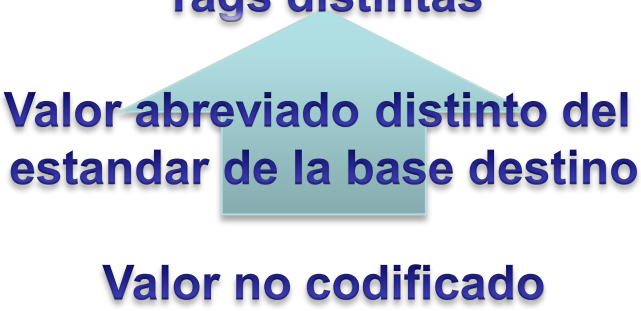
- <MedlineTA>**Arq Bras Endocrinol Metabol**</MedlineTA>

Tag <dc:description>

- <dc:description> Bangladesh Medical Research Council Bulletin. </dc:description>

Tag <dc:publisher>

- <dc:publisher> Arquivos Brasileiros de Cardiologia </dc:publisher>



bvs biblioteca virtual en salud
BIREME • OPS • OMS

Inconsistencias en el metadato de:
Publisher y paginación

<dc_publisher>

- <dc_publisher>
  Monterrey, N.L. : Edición de "la India"**,**
  </dc_publisher>

Tag XML propio

- <FirstPage>
  <![CDATA[7138]]>
  </FirstPage>
  <LastPage>
  <![CDATA[7143]]>
  </LastPage>



Inconsistencias en el metadato de:
Fecha de Publicación

Tag <dc:date>

- <dc:date>
  2013-08-07T14:09:34Z
  </dc:date>
  <dc:date>2013-08-
  07</dc:date>
  <dc:date>2009</dc:date>
  <dc:date>20090000</dc:date>

Valores distintos para la misma tag en un mismo registro

<dcterms:issued>

- <dcterms:issued>
  2013-05-14
  <dc:date>

  <dc_date>
  1882.
  </dc_date>

Punto al final del valor

bvs biblioteca virtual en salud
BIREME • OPS • OMS

Inconsistencias en el metadato de: Type

Tag <dc:type>

- <dc:type>
  Journal Article
  </dc:type>

- <dc:type>
  Research Support, Non-U.S. Gov't
  </dc:type>

Usos distintos para la misma tag

biblioteca virtual en salud
BIREME • OPS • OMS

Inconsistencias en el metadato de: Enlace para registro o texto completo

Tag <dc:identifier>

- <dc:identifier> http://imsear.hellis.org/handle/123456789/12 </dc:identifier>



- <dc:identifier> http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0066-782X1997000700001 </dc:identifier>

<dc:identifier scheme="dcterms:URI">

- <dc:identifier **scheme="dcterms:URI**"> <![CDATA[http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0066-782X1997000700001&lng=en&nrm=iso&tlng=en]]> </dc:identifier>

Inconsistencias en el metadato de:
Otros usos para <dc:identifier>

Tag <dc:identifier>

- <dc:identifier>
  http://imsear.hellis.org/handle/123456789/12
  </dc:identifier>

Enlace para registro

- <identifier>
  oai:scielo:S0066-782X1997000700001
  </identifier>

PID

- <dc:identifier>
  Azad Khan AK, Hasan M, Roy PK, Aziz S, Mahbubul Huq K, Ali SM, Haq SA. Peptic ulcer in Bangladesh. Bangladesh Medical Research Council Bulletin. 1987 Jun; 13(1): 29-42
  </dc:identifier>

Cita del registro

- <dc:identifier scheme="ags:**DOI**"/>

DOI

bvs biblioteca virtual en salud
BIREME • OPS • OMS

Inconsistencias en el metadato de: ID del registro en la base de origen

Tag <dc:identifier>

- <WPRIMID>
  <![CDATA[787780]]>
- </WPRIMID>

Enlace para registro

- <identifier>
  oai:scielo:S0066-782X1997000700001
  </identifier>

PID

bvs biblioteca virtual en salud
BIREME • OPS • OMS

Inconsistencias en el metadato de:
Fuente del registro

Tag <JournalIssue>

- <JournalIssue CitedMedium="Print">
  <Volume>48</Volume>
  <Issue>1</Issue>
  <PubDate>
  <Year>2004</Year>
  <Month>Feb</Month>
  </PubDate>
  </JournalIssue>

Tag <dc:identifier>

- <dc:identifier>
  Azad Khan AK, Hasan M, Roy PK, Aziz S, Mahbubul Huq K, Ali SM, Haq SA. Peptic ulcer in Bangladesh. Bangladesh Medical Research Council Bulletin. 1987 Jun; 13(1): 29-42
  </dc:identifier>

Cita con todos elementos juntos

Inconsistencias en el metadato de:
Fuente del registro

Tag <ags:citation>

<ags:citationTitle>Arquivos Brasileiros de Cardiologia</ags:citationTitle>

<ags:citationIdentifier scheme="ags:ISSN">0066-782X</ags:citationIdentifier>

<ags:citationNumber>vol.69 num.1</ags:citationNumber>

<ags:citationChronology>1997/07</ags:citationChronology>

</ags:citation>

Tag <dc:source>

- <dc:source> http://www.banglajol.info/index.php/BMRCB/index </dc:source>

Valores de volumen y número no separados

Inconsistencias en el metadato de: Idioma

Tag <dc:language>

- <dc:language>
  eng
  </dc:language>

- <dc:language>
  pt
  </dc:language>

Estándar de codificación distinto

Tag <Language>

- <Language>
  en**g**
  </Language>

Tag <dcterms:languageIso>

- <dcterms:languageIso
  xml:lang="pt_BR">
  Português
  </dcterms:languageIso>

Tags distintas
Valor no codificado

biblioteca virtual en salud
BIREME • OPS • OMS

Inconsistencias en el metadato de: ISSN

Tag XML propio

- <ISSN>
  <![CDATA[1673-8225]]>
  </ISSN>

Tag DC Simple <setSpec>

- <setSpec>
  0066-782X
  </setSpec>



Inconsistencias en el metadato de: DOI

Tag <ArticleId>

- <ArticleIdList>
  <ArticleId>
  <![CDATA[[doi]10.3969/j.issn.2095-4344.2013.40.018]]>
- </ArticleId>
  </ArticleIdList>

Tag propia

Tag <dc:identifier>

- <dc:identifier scheme="ags:**DOI**"/>

bvs biblioteca virtual en salud
BIREME • OPS • OMS

Inconsistencias en el metadato de: Resumen

Tag <dcterms:abstract>

- <dcterms:abstractxml:lang="**pt_BR**">
  O objetivo desta pesquisa de caráter exploratório descritivo (...) científica. (AU)
  <dcterms:abstract>

<dc:description>

- <dc:description>
  OBJETIVO: Avaliar se o enalaprilato, droga inibidora da enzima de conversão (...) colágenas.
  </dc:description>

bvs biblioteca virtual en salud
BIREME • OPS • OMS

Inconsistencias en el metadato de:
Tipo de Publicación

Tag MEDLINE

- <PublicationTypeList>
  <PublicationType>
  English Abstract
  </PublicationType>
  <PublicationType>
  Journal Article
  </PublicationType>

- <PublicationType>
  Research Support, Non-U.S. Gov't
  </PublicationType>
  <PublicationType>
  Review
  </PublicationType>
  </PublicationTypeList>

Tipos de Publicación no aceptas para LILACS

bvs biblioteca virtual en salud
BIREME • OPS • OMS

Inconsistencias en el metadato de: Descriptor

Tag MEDLINE

- <**Descriptor**Name MajorTopicYN="**Y**"> Hypokalemic Periodic Paralysis </DescriptorName>

  <**Qualifier**Name MajorTopicYN="**N**"> diagnosis </QualifierName>

  <**Qualifier**Name MajorTopicYN="**Y**"> etiology </QualifierName>

Otros valores que no solo de descriptor

<dcterms:decsPrimary>

- <dcterms:decsPrimary xml:lang="pt_BR"> Bibliotecas Virtuais </dcterms:decsPrimary>
- <dcterms:decsSecondary xml:lang="pt_BR"> Saúde Mundial </dcterms:decsSecondary>
- <dc_subject>Salud pública-**-México--Monterrey**</dc_subject>

BIREME • OPS • OMS

Recomendaciones generales

- Uso de **crosswalks** (mapeo de campos de datos) ya divulgados por los formatos bibliográficos

- Adopción de los formatos con la semántica y sintaxis conforme propuesto en el stándar

- Adopción **efectiva** de los stándares para
  - Codificación de idioma, pais, y otros
  - Control de autoridad de títulos abreviados, autoria
  - Tesauros y vocabularios controlados

- Divulgación de la documentación acerca de los campos para conversión de datos

- Control de calidad de la entrada de datos en las bases de datos de origen

Status de registros en LILACS

La contribución a LILACS va a ser dividida en status:

- **En revisión**:

Registros no indizados recogidos por cosecha de datos por OAI-PMH con Dublin Core Simple – cerca de 15 campos).

- **En proceso**

Registros sometidos por el editor o por cosecha de datos (casi LILACS-Express de hoy día):

  - Registros que adoptan metodologia SciELO - sin indización
  - Registros de cosecha de datos OAI-PMH (indizados – DCQ)
  - Registros de cosecha de datos ISIS-OAI-Provider
  - Registros LILACS-Express Editor

En ese status se propone algunas calificaciones:

  - **En proceso – SciELO**
  - **En proceso – DCQ**
  - **En proceso – LLXP Editor**

En ese status se **excluye** los registros que no cumplen con los criterios LILACS.

- **LILACS:** Registros que cumplen con los criterios LILACS y están completos, con descripción e indización del documento.

biblioteca virtual en salud
BIREME • OPS • OMS

Referéncias

- ALVES, Rachel Cristina Vesu, SANTOS, Plácida L. V. A. da C. Santos. Metadados: organização e acesso à informação no dominio bibliográfico. In: IX EIC - Encontro Internacional de Catalogadores e II Enacat - Encontro Nacional de Catalogadores. **Anais**... Rio de Janeiro, 2013. Disponible en: http://www.enacat.ufscar.br/index.php/eic-enacat/eic-enacat/paper/viewFile/52/22 Acceso en 23/04/2014.
- INDARTE, Selene. Capitulo XV: Interoperabilidad. In: Sociedad Española de Informática de la Salud (SEIS); Comisión Económica para América Latina y el Caribe (CEPAL). **Manual de salud electrónica para directivos servicios y sistemas de salud**. Pamplona: SEIS, 2012. p. 317-329. Disponible en: http://www.seis.es/documentos/informes/secciones/adjunto1/15_Interoperabilidad.pdf Acceso en 23/04/2014
- http://www.nlm.nih.gov/bsd/mms/medlineelements.html#stat Acceso en 23/04/2014
- WOODLEY, Mary S. Crosswalks, Metadata Harvesting, Federated Searching, Metasearching: Using Metadata to Connect Users and Information In: **Introduction to Metadata.** Online edition, versión 3.0. Disponible en: https://www.getty.edu/research/publications/electronic_publications/intrometadata/path.html. Acceso en 23/04/2014
- ZAFALON, Zaira Regina, SANTOS, Placida L.V.A. da Costa. Sintaxe e semântica de registros bibliográficos: princípios para a conversão de registros analógicos para o Formato MARC21 Bibliográfico: Scan for MARC. T**endências da Pesquisa Brasileira em Ciência da Informação,** vol. 5, n. 1 (2012). Disponible en: http://inseer.ibict.br/ancib/index.php/tpbci/article/view/73/134. Acceso en 23/04/2014

biblioteca virtual en salud
BIREME • OPS • OMS

Muchas Gracias!!

Dúdas??

Sueli M Yano Suga
sugasuel@bireme.ops-oms.org

BIREME • OPS • OMS