



# La Representación de Recursos en los Repositorios Institucionales

Caso de estudio: SEDICI, Repositorio Institucional de la Universidad Nacional de La Plata

Jose Texier; Marisa De Giusti; Nestor Oviedo; Gonzalo Villarreal; Ariel Lira  
[jtexier@unet.edu.ve](mailto:jtexier@unet.edu.ve)



- En el 2003 nace SEDICI (*Celsius DL*)
- Plataforma de repositorios de SEDICI (hardware & software)
- Crecimiento de funcionalidades, más compleja!!!
- Análisis en el 2011
- En el 2012, llegó Dspace a SEDICI
- Mayor flexibilidad para cambios en el software!!!
- Inconvenientes en la migración de recursos de Celsius DL a Dspace
- El beneficios de implementar Dspace @ SEDICI

- Todo objeto, físico o digital
- Conjunto de datos específicos (metadatos)
- Por ejemplo: papers, reviews, tesis, libros, etc.
- Actualmente SEDICI tiene 27.000 items y 25 tipos de recursos
- En el contexto de esta investigación, se limitó a la siguiente tipología:
  - Artículo
  - Tesis (grado, especialización, maestría, doctorado)
  - Libro
  - Entidades abstractas:
    - Autor,
    - Institución
    - Revista

# Esquema de Metadatos



- Conjunto de campos de diferentes tipos, relacionados y organizados a través de reglas (semánticas y sintácticas)
- Objetivo de registrar y organizar los metadatos
- Gran cantidad de esquemas de metadatos
- En las Ciencias de la Información, los metadatos se emplean para referirse a registros de recursos de información disponibles
- Mayor uso en la Ciencia de la Información: DC, MARC, MODS, MADS, METS, PREMIS, entre otros



# Otros Esquemas

- Darwin Core.
- DDI (Data Documentation Initiative for Social and Behavioral Sciences Data).
- DIF(Directory Interchange Format for Scientific Data).
- EML (Ecological Metadata Language).
- FGDC/CSDGM (Content Standard for Digital Geospatial Metadata).
- NBII (National Biological Information Infrastructure).
- MIAME (Minimum Information About a Microarray Experiment).
- MINSEQE (Minimum Information about a high-throughput SeQuencing Experiment).
- TEI (Text Encoding Initiative).
- EAD (Encoded Archival Description).
- ETD (Electronic Theses and Dissertations).
- MPEG-7 (Multimedia Content Description Interface).

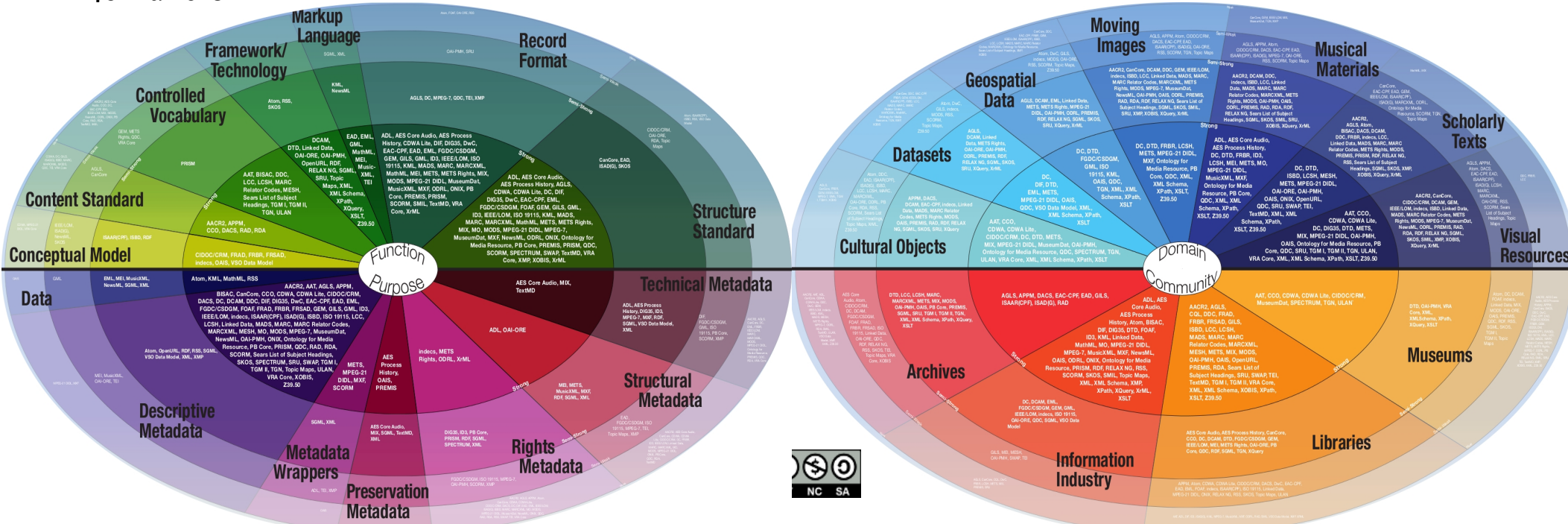
**Los esquemas de metadatos junto con los recursos se convierten en el elemento central del diseño de software para repositorios.**

# Mapa de Metadatos por Jenn Riley



Eje Función

Eje Dominio



Eje Propósito

Eje Comunidad

<http://www.dlib.indiana.edu/~jenlrile/metadatamap/seeingstandards.pdf>



- Almacenar de forma persistente los metadatos como síntesis y reemplazo del objeto "real" para poder identificarlo, recuperarlo y distribuirlo
- Hay muchos recursos y son muy variados
- Cantidad de esquema de metadatos, vocabularios controlados, tesauros, etc.
- Alternativas de base de datos para preservar los recursos
- Recuperar el recursos en forma eficiente
- Garantizar la preservación en el tiempo
- Interoperabilidad con otros repositorios
- **RESULTADO:** una representación de recursos compleja en el contexto de los repositorios institucionales

# Algunas Características en una Representación

- Los recursos deben estar catalogados en más de un esquema de metadatos a fin de evitar pérdida de información
- Diversidad en las estructuras (planas y jerárquicas) y restricciones (simples y complejos) de cada esquema de metadatos
- Muchos campos presentes en los esquemas de metadatos necesitan ser estandarizados (vocabularios controlados o tesauros)
- Entidades abstractas que representan información descriptiva propia, tienen distintas representaciones según el esquema de metadatos y la semántica deseada.
- Almacenamiento para cualquier tipo de paradigma de base de datos
- Aspectos que influyen: complejidad del software, interoperabilidad, y, escalabilidad y performance
- Diversidad en plataformas de software, adicional a los desarrollos propios

# Problemas de la Representación de Recursos dentro de los RI



- Formatos de metadatos:
  - Se pueden clasificar por su estructura (planos y jerárquicos) y por su especificidad (simples y complejos)
  - Su relación con la base de datos que se use
- Vocabularios controlados:
  - Tesoros, sistemas de clasificación del conocimiento, idiomas, referencias geográficas, tipos de recursos, materias, frecuencia de entrega
- Entidades abstractas:
  - Entendidas como un conjunto de elementos con información descriptiva propia son utilizadas en los procesos de catalogación, p.e. autores, instituciones, revistas y sus números, eventos y sus instancias.
- Representación física de los datos:
  - Tiempos de respuestas y consumo de recursos, mantener sus cualidades aumentando sus recursos, usuarios, etc.
  - Base de datos en XML, relacionales, orientadas a objetos, RDF, o soluciones mixtas.



# Caso de estudio: Celsius DL (SEDICI I)

- Formato de metadatos
  - Formato propio
  - Normas de catalogación (de forma obligatoria, recomendada u opcional)
- Vocabularios controlados
  - Conjunto restringido de términos controlados (UNESCO, Eurovoc, DEC's, entre otros)
  - En primer término una “macrocatalogación” (materias), luego una catalogación temática (descriptores) y finalmente, si las hay, mediante las palabras-clave proporcionadas en el texto por su autor.
- Entidades abstractas
  - 17 tipos de documentos, 4 de los cuales son entidades abstractas: serie documental o publicación periódica, entrega documental o número de publicación periódica y congresos. Además se encuentran las entidades abstractas autores e instituciones.
- Representación física de los datos
  - Estaba basada en un conjunto de tablas relacionales con MySQL

# Caso de estudio: Dspace-SEDICI (SEDICI II)



- Versión: DSpace 1.8
- Formato de metadatos
  - ☒ DSpace no soporta esquemas de metadatos jerárquicos, por ello, se tiene que limitarse a esquemas existentes o propios de estructuras no anidadas
  - ☒ Un esquema propio y combinado con dos esquemas conocidos como DC y una emulación de MODS (se usa el nombre del campo como elemento y el nombre del campo hijo como calificador)
- Vocabularios controlados & Entidades abstractas
  - ☒ Los vocabularios y entidades abstractas usadas se siguen manteniendo salvo una particularidad, para establecer un vínculo con los autores, en SEDICI-DSpace se creó una tabla de referencia que mantiene una relación con los autores que se usaban en Celsius DL
- Representación física de los datos.
  - ☒ Persistencia de los datos es asegurada bajo un modelo de base de datos relacional conocido como PostgreSQL



- Relevamiento bibliográfico de la representación de recursos en los repositorios institucionales
- Analizar y comparar los modelos de los repositorios y establecer los pro y los contra que tienen cada uno de ellos
- Visualización del repositorio a través del enfoque Model-Driven
- Evolución del concepto de los repositorios y de las bibliotecas digitales, el gran auge que los RI y el afianzamiento de la filosofía del acceso abierto en la comunidad de investigadores y académicos, son realidades incentivan el estudio y desarrollo de aplicaciones del dominio LIS.
- Generar un modelo general que permita una representación de recursos de manera indistinta a la plataforma que se use

# La Representación de Recursos en los Repositorios Institucionales

*Caso de estudio: SEDICI, Repositorio Institucional de la Universidad Nacional de La Plata*



감사합니다 Natick  
 Grazie Danke Ευχαριστίες Dalu  
 Thank You Köszönöm  
 Спасибо Dank Tack  
 谢谢 Merci Seé  
 ありがとう  
 Obrigado

<http://ayudawordpress.com/wp-content/uploads/2012/07/gracias-en-varios-idiomas-500x290.jpg>

Jose Texier; Marisa De Giusti; Nestor Oviedo; Gonzalo Villarreal; Ariel Lira  
 jtexier@unet.edu.ve

