

XLIII CLEI / 46 JAIIO

SLMDI / AGRANDA

Repositorio Semántico para la Universidad Nacional de Chilecito

Autores:

**Jose Texier, Marisa De Giusti, Gonzalo Villarreal,
Ariel Lira y Jusmeidy Zambrano**

✉ jtexier@undec.edu.ar; dantexier@gmail.com
🐦 @dantexier



Plan de exposición

- Contexto de la propuesta
- Objetivos
- Pilares de la propuesta
- Linked Open Data y DSpace
- Metodología Heath y Bizer
- Propuesta
- Trabajos referentes
- Metadatos de los Sistemas Nacionales
- Trabajos futuros

Contexto de la propuesta

- Ministerio de Ciencia, Tecnología e Innovación Productiva (MinCyT) y el Consejo Interinstitucional de Ciencia y Tecnología (CICyT).
 - Iniciativa conjunta del Programa de Grandes Instrumentos y Bases de Datos.
 - Metadatos estandarizados de SNRD, SNDB y SNDG.
 - Integración semántica.
- Ley 26899: Creación de Repositorios Digitales Institucionales de Acceso Abierto, Propios o Compartidos (2013).
- Ley 27275: Derecho de acceso a la información pública (2016).
- SEDICI, Servicio de Difusión de la Creación Intelectual, repositorio institucional de la UNLP
- Web Semántica y el Linked Open Data

Objetivos

- Desarrollar un repositorio semántico para la Universidad Nacional de Chilecito que unifique los recursos de los Sistemas Nacionales de datos biológicos, genómicos y repositorios digitales
- Brindar un conjunto de herramientas/aplicaciones web a la comunidad de la UNdeC
- Fomentar la cultura de compartición de recursos y vinculación de personas
- Consolidar el movimiento de Acceso Abierto en la comunidad
- Ofrecer otras posibilidades de temas de investigación a las disciplinas relacionadas de la propuesta

Pilares de la propuesta

- Los cuatro principios:
 - Usar URI como nombres de recursos (lugares, personas, eventos,...)
 - Usar HTTP URI (recursos) de manera que se puedan localizar.
 - Presentar información usando los estándares RDF o SPARQL, es decir, cuando se busque un URI.
 - Incluir enlaces a otros URI de manera que se puedan encontrar más recursos relacionados.
- RDF: representación del conocimiento basado en redes semánticas. No es un formato, es un modelo de datos.
 - Expresión de un metadato (tripleta): **sujeto** → el recurso sobre el que se declara algo, **predicado** → el aspecto que se describe, y **objeto** → el valor de lo que se describe sobre el sujeto
- DSpace: Soft. referente para la gestión de repositorios, licencia abierta
- OAI-PMH: protocolo para la extracción de metadatos desde repositorios digitales, independiente de la infraestructura usada. Integración a nivel sintáctico.

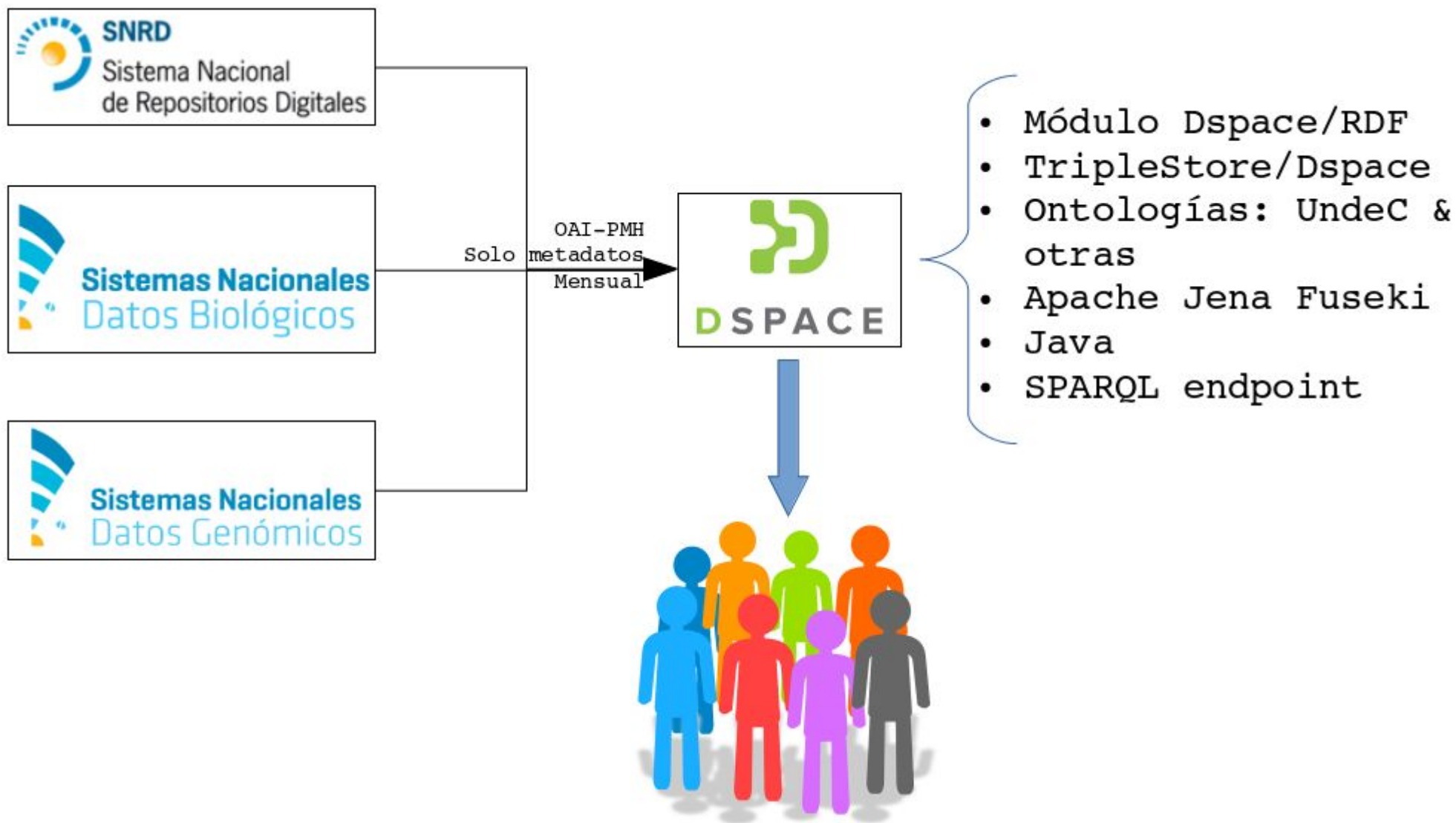
Linked Open Data y DSpace

- Linked Open Data
 - Sinónimo de datos semánticamente estructurados
 - Datos publicados (Open Data) y enlazados (Linked Data) mediante la estructura de la Web Semántica
- DSpace:
 - Módulo RDF: `[dspace]/config/modules/rdf.cfg`
 - SPARQL 1.1 Query Language y SPARQL 1.1 Graph Store HTTP Protocol
 - Reglas de conversión de los metadatos:
 - `[dspace]/config/modules/rdf/metadata-rdf-mapping.ttl`
 - Convertir los recursos a RDF (`rdifier`)
 - Vincular Fuseki-Server con TripleStore de DSpace
 - `[dspace]/config/modules/rdf/fuseki-assembler.ttl`
- Apache Jena Fuseki y JAVA
- Consultas:
 - Browser y SPARQL en Jena Fuseki

Metodología Heath y Bizer

- **Determinación de la población** a examinar: sistemas nacionales
- **Recolección y selección de datos**: OAI-PMH (verificación y "mensuales")
- **Validación** para asegurar el control de **calidad** de los datos: se analizará la estructura del conjunto de datos cosechados y reutilizables.
- **Conversión de los datos**: conversión a RDF (`rdflizer`).
- **Almacenamiento centralizado** de datos: los datos se almacenarán bajo el esquema RDF/DSpace.
- **Uso del software**: DSpace v. 6.1.
- Aplicación del **módulo RDF** y **visualización** de los datos.
- **Análisis de los resultados**: matriz FODA o técnicas de Benchmarking.

Propuesta



Trabajos referentes

- El Repositorio de Investigadores de Ecuador:
 - <http://redi.cedia.org.ec/>
- Semantic Scholar:
 - <https://www.semanticscholar.org/>
- Semantic Data Repository EU:
 - <http://digital-agenda-data.eu/data/>
- Upapers:
 - <https://upapers.dcc.uchile.cl/index/>
- PaperCube:
 - <http://papercube.peterbergstrom.com/>
- Semantic Search for DSpace (en prueba):
 - <http://swig.hpclab.ceid.upatras.gr/dspace-ss-demo/>
- EconStor:
 - <https://www.econstor.eu/>

Metadatos de los SN (1/4)

- Metadatos comunes de un recurso en tres Sistemas Nacionales:
 - autor
 - fecha de creación
 - identificador
 - tipo de recurso
 - descripción
 - idioma
 - título
 - tipo de licencia



Sistemas Nacionales
Grandes Instrumentos, Facilidades y Bases de Datos



Ministerio de Ciencia,
Tecnología e Innovación Productiva
Presidencia de la Nación



● EL PROGRAMA

El Programa de Grandes Instrumentos y Bases de Datos es una iniciativa conjunta del Ministerio de Ciencia, Tecnología e Innovación Productiva y el Consejo Interinstitucional de Ciencia y Tecnología (CICYT).

Su propósito es alcanzar una utilización eficiente de los grandes equipamientos y una mejor organización y acceso a las bases de datos científicos existentes en las distintas instituciones de ciencia y tecnología de todo el país. Los Sistemas Nacionales forman parte del Programa y trabajan en conjunto para alcanzar este objetivo.

NUEVO

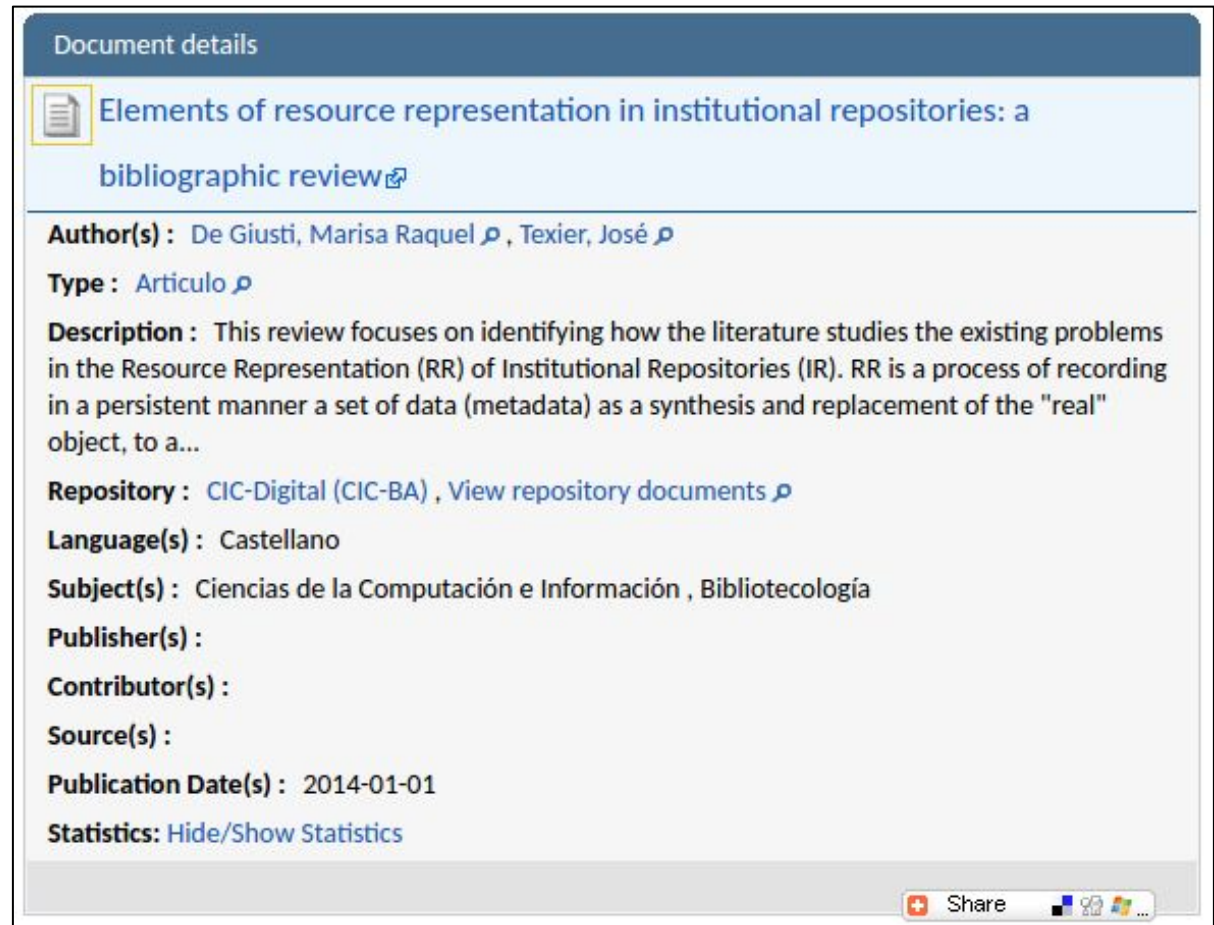
Los Sistemas actúan en forma coordinada en diversas áreas de interés para la comunidad científica, con el fin de poner a disposición de los investigadores tanto las bases de datos como los equipos registrados en el patrimonio de un organismo público. Asimismo, brindan apoyo económico a las instituciones adheridas para garantizar su actualización constante.

GRANDES INSTRUMENTOS	BASES DE DATOS	
Sistemas Nacionales Datos del Mar	Sistemas Nacionales Datos Biológicos	Sistemas Nacionales Datos Climáticos
Sistemas Nacionales Repositorios Digitales	Sistemas Nacionales Datos Genómicos	



SISTEMA DE
GESTIÓN DE TURNOS

Metadatos de los SN (2/4)

- Un recurso del SN de Repositorios Digitales
 - páginas
 - fulltext
 - peer review
 - revista
 - volumen
 - issue
 - archivo digital del recurso



Document details

 Elements of resource representation in institutional repositories: a bibliographic review [↗](#)

Author(s) : De Giusti, Marisa Raquel [↗](#) , Texier, José [↗](#)

Type : Artículo [↗](#)

Description : This review focuses on identifying how the literature studies the existing problems in the Resource Representation (RR) of Institutional Repositories (IR). RR is a process of recording in a persistent manner a set of data (metadata) as a synthesis and replacement of the "real" object, to a...

Repository : CIC-Digital (CIC-BA) , [View repository documents ↗](#)

Language(s) : Castellano

Subject(s) : Ciencias de la Computación e Información , Bibliotecología


Publisher(s) :

Contributor(s) :

Source(s) :

Publication Date(s) : 2014-01-01

Statistics: [Hide/Show Statistics](#)

[+](#) Share 

<https://digital.cic.gba.gob.ar/handle/11746/3255>

Metadatos de los SN (3/4)

- Un recurso del SN de Datos Biológicos:
 - colección
 - número de catálogo
 - tipo de registro
 - datasets asociados
 - taxonomía
 - geolocalización
 - registro de calidad de datos

Registro: EBIRD:OBS382183916

Lessonia rufa

Pruebas de calidad de datos

fallida: 0
 alertas: 2
 Aprobada: 24
 Faltante: 11
 desactivado: 47

[Ver informe de calidad de datos completo](#)

Data resource	EOD - eBird Observation Dataset
Institution	Aves Argentinas - Asociación Ornitológica del Plata Reemplazado código de institución "CLO"
Collection	Colección de Aves de eBird curadas en Argentina
Catalogue number	OBS382183916
Occurrence ID	URN:catalog:CLO:EBIRD:OBS382183916
Basis of record	HumanObservation Base reemplazada "HUMAN_OBSERVATION"
Fecha de registro	2015-12-10 Fecha reemplazada "2015-12-10T01:00Z"
Individual count	12
Collection code	EBIRD
Recorded by	obsr93553
Taxonomía	
Scientific name	<i>Lessonia rufa</i> Supplied scientific name " <i>Lessonia rufa</i> [Gmelin, 1789]"
Taxon rank	Species
Kingdom	Animalia

<http://datos.sndb.mincyt.gob.ar/generic-hub/occurrences/c012349a-027a-472e-bb2d-73790561c1d1>

Metadatos de los SN (4/4)

- Un recurso del SN de Datos Genéticos:
 - procedencia
 - secuenciación ADN
 - tipo de organismo
 - tipo de comunidad
 - datasets asociados

Genoma Yerba mate: Ilex Paraguarensis Genoma Yerba mate: Ilex Paraguarensis

Genomas de Bacterias Extremófilas

- ☰ Hemos ya secuenciado varios genomas de bacterias de diferentes orígenes.
Tipo de organismo: Bacteria [Varios, Exiguobacterium, Bizionia]
- 📁 **Proyecto:** Extremófilos
 El proyecto consiste en secuenciar bacterias extremófilas y analizar sus funciones.
- 👤 Marcelo Martí
- 👥 Walter McCormack, Farias
- ✅ Adherido al SNDG desde 07-08-2015
- 🏠 Instituto de Química Biológica de la Facultad de Ciencias Exactas y Naturales [101]
- 🌐 <http://www.iqubicen.fcen.uba.ar/home/>

Recursos

Pseudomonas fluorescens SF4c

Pseudomonas fluorescens SF4c

Aeromonas salmonicida pectinolytica 34mel

Datos	
Download Genome Sequence	
Download GFF annotation	
Type:	Soap de novo
Total Size	1.445.128.269 bp
Shortest Contig	500 bp
Scaffolds Count	65.682
Scaffolds > 1mb	count 34 (0.1%)
Scaffolds > 1kb count	53.487 (81.4%)

http://arkham.exp.dc.uba.ar/xomeq/genome/ILEX_PARA?a=1

Trabajos futuros

- Analizar la aplicación de otras metodologías para el desarrollo de ontologías (p.e. Methontology, Uschold y King, TERMINAE, Ontology Development 101, etc.)
- Ofrecer servicios web de consultas
- Ampliar las fuentes, por ejemplo: SN de Datos Climáticos & SN Datos del Mar.
- Interrelacionarse con otras bases de conocimiento.
- Lograr la integración virtual (*mediante tecnologías de LOD*), ya que en la propuesta presentada se pierde la idea de la Web (información distribuida).
- Desarrollar un módulo estadístico y la posibilidad de un cubo de datos RDF.

Preguntas, comentarios y/o dudas



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

Jose Texier

dantexier@gmail.com

@dantexier

XLIII CLEI / 46 JAIIO

SLMDI / AGRANDA

Repositorio Semántico para la Universidad Nacional de Chilecito

Autores:

**Jose Texier, Marisa De Giusti, Gonzalo Villarreal,
Ariel Lira y Jusmeidy Zambrano**

✉ jtexier@undec.edu.ar; dantexier@gmail.com
🐦 @dantexier

