Digital Repositories. Technical concepts and software platform options.

Emiliano Marmonti
IAEA
June 2015
Some definitions (from common sense to digital)

A Repository

A “site” or container to deposit and store objects (security & preservation)

A Digital Repository

A “system” to collect, store, preserve and give access to digital objects (images, documents, videos, etc).

An Institutional Repository

A digital repository that is institutionally defined: It means that contains **ONLY** material produced by the institution that host the service.
• Store, preserve and disseminate the production of the members of one institution (University, R+D center, etc).

• Some of them contains traditional academic material as graduate or postgraduate Theses, Journal Articles, and some others include learning objects, multimedia, historical digitalized documents, etc.

• Implement some or several ways to access to the content. The most common way is to organize the material in collections and allow to the final user to browse the content via different criterias.

Institutional Repositories

- Institutionally defined.
- Academic or scientific orientation.
- Cumulative and perpetual.
- Open.
- Interoperable.
Some IR examples
- Focused on a discipline.

- Some are created by Academic or Scientific organizations, others by gubernamental organizations.

- Can co-exist with institutional repositories from the same institution.

- Examples: Health Sciences, PubMed, Economics, RePEc, Information Sciences and Libraries, RLiCS

Could allow more than one discipline, for instance arxiv.org (physics, computer science, math, etc)
Disciplinary Repositories

RePEc (Research Papers in Economics) is a collaborative effort of hundreds of volunteers in 82 countries to enhance the dissemination of research in Economics and related sciences. The heart of the project is a decentralized bibliographic database of working papers, journal articles, books, books chapters and software components, all maintained by volunteers. The collected data is then used in various services as described below.

So far, over 1,600 archives from 81 countries have contributed about 1.4 million research pieces from 1,800 journals and 3,800 working paper series. Over 35,000 authors have registered and 70,000 email subscriptions are served every week. See below on how you can be part of this initiative.

RePEc services

The following are services that use (principle) and contribute RePEc data. They also report usage statistics that can be used towards the RePEc rankings.

- RePEc Author Service
- Munich Personal RePEc Archive
- IDEAS
- EconPapers
- RePEc Genealogy
- RePEc Biblio

Organic eprints is an international open access archive of electronic documents related to research in food and farming. The archive contains full-text papers together with bibliographic information, abstracts and other metadata. It also offers information on organisations, projects and the context of organic farming research.

Objectives of Organic Eprints are:

- Facilitate the communication about organic research,
- Improve the dissemination and impact of research findings,
- Document the research effort.

In 2003 the Research Institute of Organic Agriculture (FiBL) joined the project as its first international partner with editorial responsibilities for the German language version of Organic Eprints.

Additionally, Organic Eprints has national editors, currently in 26 countries, who assist in reviewing new eprints. See Contact Information.

Open access

Organic Eprints has been established to promote open access to research results. Open access means that the eprints are freely accessible to all who have access to the internet. The immediate benefit of an open access archive are several. Researchers and research institutions can document their production and make their research products available and searchable to a large audience.

Read more...
Primary Research data Repositories

- Store and preserve data that is produced in the context of a scientific publication.

- Shared under the concepts of replication and validation for the publication conclusions.

- Shared under the idea of re-use as starting point of another works. Open Science.

- Could be integrated as a collection of an institutional repository or as an independent repository.

Often (and should be) Integrated as a part of a Data Management plan.

NSF Data Management Plan Requirements
Some examples Primary Research
Data Repositories

DataDryad.org is a curated general-purpose repository that makes the data underlying scientific publications discoverable, freely reusable, and citable. Dryad has integrated data submission for a growing list of journals; submission of data from other publications is also welcome.

Browse for data
Recently Published Data


Definitions. Learning object

“a collection of content items, practice items, and assessment items that are combined based on a single learning objective”. Hodgins (1994)

Koper (2001)

Stolen from http://dspace.ou.nl/bitstream/1820/4921/1/OER_atelier_btabuenca.pdf
Open Educational Resources

OER are teaching, learning, and research resources that reside in the public domain or have been released under an intellectual property license that permits their free use and repurposing by others. Open educational resources include full courses, course materials, modules, textbooks, streaming videos, tests, software, and any other tools, materials, or techniques used to support access to knowledge.

Hewlett Foundation: http://www.hewlett.org/programs/education/open-educational-resources
The 5 R’s of OER

• Reuse
• Remix
• Revise
• Retain
• Redistribute

/Libre
Creative Commons Licenses

[CC BY] Attribution
Others may distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you for the original creation.

[CC BY SA] Attribution Share Alike
Others may remix, tweak, and build upon your work even for commercial reasons, as long as they credit you and license their new creations under the identical terms.

[CC BY ND] Attribution No Derivatives
Others may redistribute your work, commercially and non-commercially, as long as it is passed along unchanged, unedited, and attributed.

[CC BY NC] Attribution Non-Commercial
Others may remix, tweak, and build upon your work non-commercially, as long as they credit you.

[CC BY NC SA] Attribution Non-Commercial Share Alike
Others may remix, tweak, and build upon your work non-commercially, as long as they credit you and license their new creations under the identical terms. New work based on yours will carry the same license.

[CC BY NC ND] Attribution Non-Commercial No Derivatives
Others may share your work unedited and non-commercially as long as they credit you. (This is the most restrictive of the six licenses.)
<table>
<thead>
<tr>
<th>TYPE</th>
<th>CREATOR</th>
<th>OWNER</th>
<th>CC LICENSE DESCRIPTION</th>
<th>LICENSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>MERLOT info Help documents</td>
<td>Staff</td>
<td>MERLOT</td>
<td>Attribution, Non Commercial, Derivative, Share-Alike</td>
<td>by-nc-sa</td>
</tr>
<tr>
<td>MERLOT public website content &amp; images</td>
<td>Staff</td>
<td>MERLOT</td>
<td>Attribution, Non Commercial, No Derivative</td>
<td>by-nc-nd</td>
</tr>
<tr>
<td>MERLOT Logos</td>
<td>Staff</td>
<td>MERLOT</td>
<td>Attribution, Non Commercial, No Derivative</td>
<td>by-nc-nd</td>
</tr>
<tr>
<td>MERLOT portal content</td>
<td>Partner</td>
<td>MERLOT</td>
<td>Attribution, Non Commercial, Derivative, Share-Alike</td>
<td>by-nc-sa</td>
</tr>
<tr>
<td>Peer reviews</td>
<td>Partner</td>
<td>MERLOT</td>
<td>Attribution, Non Commercial, No-Derivative</td>
<td>by-nc-nd</td>
</tr>
<tr>
<td>Learning material metadata</td>
<td>Member</td>
<td>MERLOT</td>
<td>Attribution, Non Commercial, Derivative, Share-Alike</td>
<td>by-nc-sa</td>
</tr>
<tr>
<td>Assignments</td>
<td>Member</td>
<td>MERLOT</td>
<td>Attribution, Non Commercial, No-Derivative</td>
<td>by-nc-nd</td>
</tr>
<tr>
<td>Member Comments Snapshots</td>
<td>Member</td>
<td>MERLOT</td>
<td>Attribution, Non Commercial, No-Derivative</td>
<td>by-nc-nd</td>
</tr>
<tr>
<td>Bookmark Collections</td>
<td>Member</td>
<td>MERLOT</td>
<td>Attribution, Non Commercial, Derivative, Share-Alike</td>
<td>by-nc-sa</td>
</tr>
<tr>
<td>Course ePortfolios</td>
<td>Member</td>
<td>MERLOT</td>
<td>Attribution, Non Commercial, Derivative, Share-Alike</td>
<td>by-nc-sa</td>
</tr>
<tr>
<td>Discussion Board Creation/submission</td>
<td>Anyone</td>
<td>MERLOT</td>
<td>Attribution, Non Commercial, No Derivative</td>
<td>by-nc-nd</td>
</tr>
<tr>
<td>MERLOT International Conference</td>
<td>Anyone</td>
<td>MERLOT</td>
<td>Attribution, Non Commercial, Derivative, Share-Alike</td>
<td>by-nc-sa</td>
</tr>
</tbody>
</table>

Merlot aggregator use of CC different licenses schemas for each material type

http://www.merlot.org/merlot/index.htm
Mostly related to the way to describe the objects (metadata). EducationalLevel, Context, possibilities to describe Chapters, interactivityType.

Most of the items of a OERR are complex objects, it means that are composed by several documents, tools, software, etc. It represents a challenge for the metadata schema, for digital preservation (one could think about exercises based on Java Applets for instance).

The repository / links / digital objects

The recommended practice is that every item in the repository will be composed by metadata and the digital asset. Some repositories maintain a reference or a link to a digital object as a way to disseminate the material, or in the meantime that the digital asset will be stored there. This situation is different of security levels of metadata and digital assets.
Some examples
Metadata & Metadata Schemas.
Consider:

- The **type of the material** that the repository will manage (general purpose, theses, articles, multimedia, OERs, etc).
- The **technology support** for the material (pdf documents, videos, software, etc).
- The **main objective** of the repository: Acess, Dissemination, Preservation.
- Depth of **description** (Librarian assistance? Taxonomies? )
- **Similar experiences**. To analyze if there are similar experiences and the decisions that have arrived.
- Search & analyze the **existent metadata schemas**. Prioritize those that are part of consortia, international organizations, and are actively maintained.
- **Connect! Interoperate!** Which networks I would connect / integrate. Which are required metadata schema?
Created at the end of 90's in Ohio, USA.

Keep it simple. Created under the concept that the author can “metadata” his/her own data.

To ease the identification and retrieval of web documents.

Maintained & updated by the DCMI (Dublin Core Metadata Initiative).

The elements are divided in three groups: Related with the content, related with intellectual property and freedoms of use and deployment, use & manipulation (of the described digital object).
**Dublin Core Metadata Element Set (DCMES)**
(version 1.1)

1. Title
2. Creator
3. Subject
4. Description
5. Publisher
6. Contributor
7. Date
8. Type
9. Format
10. Identifier
11. Source
12. Language
13. Relation
14. Coverage
15. Rights Management

The implementation of a data-model. Aimed to support: description, discovery and re-use of the material described.


- 9 categories (that includes elements)
  General         Educational
  Life Cycle      Rights
  Meta-Metadata    Relation
  Technical       Annotation
  Classification

More info in:
Electrophoresis to separate DNA strands

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>dc:audience.educationlevel</td>
<td>Undergraduate (Upper Division)</td>
</tr>
<tr>
<td>dc:contact.email</td>
<td><a href="mailto:ambusch@utmb.edu">ambusch@utmb.edu</a></td>
</tr>
<tr>
<td>dc:contributor.affiliation-institution</td>
<td>utmb.edu</td>
</tr>
<tr>
<td>dc:contributor.affiliation-division</td>
<td>Clinical Laboratory Sciences</td>
</tr>
<tr>
<td>dc:contributor.author</td>
<td>Vicki Freeman</td>
</tr>
<tr>
<td>dc:contributor.submitter</td>
<td>Anne Busch</td>
</tr>
<tr>
<td>dc:date.created</td>
<td>2015-03-06T20:50:51Z</td>
</tr>
<tr>
<td>dc:date.accessioned</td>
<td>2015-03-19T21:28:00Z</td>
</tr>
<tr>
<td>dc:date.available</td>
<td>2015-03-19T21:28:00Z</td>
</tr>
<tr>
<td>dc:date.issued</td>
<td>2015-03-06T20:50:24Z</td>
</tr>
<tr>
<td>dc:date.digital</td>
<td>2015-02-25T08:00:00Z</td>
</tr>
<tr>
<td>dc:date.modified</td>
<td>2015-02-25T16:21:33Z</td>
</tr>
<tr>
<td>dc:description.abstract</td>
<td>This animation presents what is happening to the various sizes.</td>
</tr>
<tr>
<td>dc:description.comment</td>
<td>Partially funded by FIPSE</td>
</tr>
<tr>
<td>dc:description.peerreview</td>
<td>no</td>
</tr>
<tr>
<td>dc:description.peerreview-date</td>
<td>2015-03-16T02:22:16Z</td>
</tr>
<tr>
<td>dc:format.extent</td>
<td>0</td>
</tr>
<tr>
<td>dc:identifier.uri</td>
<td><a href="http://hdl.handle.net/2249.3/1542">http://hdl.handle.net/2249.3/1542</a></td>
</tr>
<tr>
<td>dc:instructionalmethod</td>
<td>Demonstrations</td>
</tr>
<tr>
<td>dc:language.iso</td>
<td>English</td>
</tr>
<tr>
<td>dc:language</td>
<td>English</td>
</tr>
<tr>
<td>dc:publisher</td>
<td>University of Texas Medical Branch</td>
</tr>
<tr>
<td>dc:relation.haspart</td>
<td>no</td>
</tr>
<tr>
<td>dc:rights.accessrights</td>
<td>Attribution-NonCommercial-ShareAlike CC BY-NC-SA</td>
</tr>
<tr>
<td>dc:rights.license</td>
<td>Creative Commons</td>
</tr>
<tr>
<td>dc:rights.rightsholder</td>
<td>University of Texas Medical Branch</td>
</tr>
<tr>
<td>dc:subject</td>
<td>Molecular Biology</td>
</tr>
<tr>
<td>dc:subject.gender</td>
<td>No Selection</td>
</tr>
<tr>
<td>dc:subject.keyword</td>
<td>Electrophoresis, separation of DNA fragments, molecular techniques</td>
</tr>
<tr>
<td>dc:title</td>
<td>Electrophoresis to separate DNA strands</td>
</tr>
<tr>
<td>dc:subject.course</td>
<td>Molecular Biology</td>
</tr>
<tr>
<td>dc:type</td>
<td>Illustration</td>
</tr>
<tr>
<td>lom:classification.taxon</td>
<td>Clinical Laboratory Sciences</td>
</tr>
<tr>
<td>lom:educational.difficulty</td>
<td>Easy</td>
</tr>
<tr>
<td>lom:general.aggregationlevel</td>
<td>Learning Object (L3)</td>
</tr>
<tr>
<td>lom:rating</td>
<td>5</td>
</tr>
<tr>
<td>lom:subject.course</td>
<td>Molecular Biology</td>
</tr>
<tr>
<td>lom:subject.language</td>
<td>English</td>
</tr>
<tr>
<td>lom:subject.rightsholder</td>
<td>University of Texas Medical Branch</td>
</tr>
<tr>
<td>lom:subject.rightsholder</td>
<td>University of Texas Medical Branch</td>
</tr>
<tr>
<td>lom:subject.rightsholder</td>
<td>University of Texas Medical Branch</td>
</tr>
</tbody>
</table>

**Ficheros en el ítem**

<table>
<thead>
<tr>
<th>Nombre del fichero</th>
<th>Tamaño</th>
<th>Formato</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMPLETE-OBJECT-a66G/RLT7g-14.zip</td>
<td>4.692Kb</td>
<td></td>
</tr>
<tr>
<td>Instructions - LO27.txt</td>
<td>2.176Kb</td>
<td>Texto</td>
</tr>
<tr>
<td>MolAn2-27v2.swf</td>
<td>3.339Kb</td>
<td></td>
</tr>
</tbody>
</table>

Este documento presenta lo que está sucediendo...
Another approach (embedded tags) LRMI

Based on the approach of schema.org. Structured data. Microdata. Supported by Google, bing, CC

Consists in include the metadata in the same content, as specific HTML tags.

Focused on optimize SEO and discoverability of the resources.

The LRMI Properties v1 (lrmi.net/the-specification)

General Terms (Schema.org):
- title/name
- URL
- description
- image
- about
- created (date)
- Creator
- Publisher
- inLanguage
- useRightsUrl*
- isBasedOnUrl

Educational Terms:
- educationalRole
- educationalUse
- timeRequired
- typicalAgeRange
- interactivityType
- learningResourceType

Competency related:
- educationalAlignment
- educationalFramework
- Alignment Type
- targetDescription
- targetName
- targetURL

* part of LRMI spec but not yet adopted by Schema.org
PREservation Metadata Implementation Strategies

• Dictionary of elements to support long-term digital preservation.

• It allows to describe:
  
  - Relative to the object: Identifiers, integrity, creation context and relation with another.
  
  - Events. Tracking of updates, who and when has been interacting with the object.
  
  - Agents associated with its preservation (people, organization, software, etc)
  
  - Rights associated with the object and the agents.

Más información en:
http://www.loc.gov/standards/premis/v2/premis-2-0.pdf
“...The capacity that presents an information system to

[Gómez Dueñas, 2005]
Semantic Interoperability: searchable digital information across heterogeneous databases.
- To describe: Cataloging rules,
- To index/search: Thesaurus, Ontologies, etc.

Syntax interoperability: Is given by the use of same elements of the metadata schema.

Estructural Interoperability: Common logical models & protocols, in the case of library software and repositories: Z39.50, OAI-PMH, OAI-ORE.
Interoperability relevant topics (according to COAR)

What is COAR?
Confederation of Open Access Repositories. This organization has several focused working groups, specialized in repository interoperability, repository content, I+D infraestructures.

https://www.coar-repositories.org/es/

Figure 1: Areas of Interoperability-Related Services
Nivel 3. OAI-PMH harvesting

OAI compliant repository (Data Provider)

XML-UTF8

Metadata Harvesting

Service Provider

Search

User
The metadata harvesting purpose is to implement a centralized service that could offer several services to the final user, but the main service is to provide a centralized search for all the harvested content.

The centralized services could be classified in:

- Harvesters
- Aggregators
- Centralized Repositories
OpenAire portal / BASE Academic searcher

Agreggates content from Institutional Repositories, Thematic Repositories and primary research data repositories.

73,889,455 Documents of 3,546 Content Sources
LaReferencia.
Agregates content from Institutional Repositories national networks from LatinAmérica

http://lareferencia.info
Centralized repository – World Health Organization

http://apps.who.int/iris
OER aggregator/s (That uses OAI-PMH)

The ARIADNE Foundation is a not-for-profit association that aims to:

Handshake Ecosystem for Educational Contents between Institutional Repository and OER based Repository

by Kazu Yamaji, National Institute of Informatics, JAPAN
OER Catalogue / Repository
(That could aggregate resources via OAI-PMH)

Experts Curate for You
The OER Commons Team designs custom tools and environments to support OER networks.

OER SUPPORT SERVICES

Browse by Topic

Education / Arts / Humanities / Social Sciences / Natural Sciences / Applied Sciences & Technology / Mathematics & Statistics
Setup the network.
Problems that appear after joining metadata from different initiatives
Spanish study about one focused metadata distribution.

<table>
<thead>
<tr>
<th># registros por repositorio</th>
<th>% mal formato</th>
<th>% no es fecha</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-5000</td>
<td>27,97</td>
<td>13,73</td>
</tr>
<tr>
<td>5000-10000</td>
<td>5,44</td>
<td>0,45</td>
</tr>
<tr>
<td>10000-50000</td>
<td>19,09</td>
<td>0,87</td>
</tr>
<tr>
<td>+ de 50000</td>
<td>19,72</td>
<td>0,04</td>
</tr>
</tbody>
</table>

Tabla II. Porcentaje de registros que no cumplen el formato de fecha YYYY-MM-DD vs. porcentaje de registros que no son fechas, en relación al tamaño de cada repositorio.

Como conclusión general, puede afirmarse que los repositorios digitales han experimentado en España un notable desarrollo, adquiriendo una masa crítica considerable. Pero este importante crecimiento en número y tamaño pone de manifiesto la necesidad de reforzar la calidad de los metadatos. Una vez plenamente asentados los repositorios, es tal vez el momento de centrar esfuerzos en ofrecer registros bien descritos, haciéndolos así más accesibles.
The guidelines have been produced to facilitate the metadata exchange (DRIVER or others).

Describe an agreement to standardize the metadata content inside the network, to respect international standards and accomplish semantic interoperability in the harvesting process.

Also define some criteria for technical interoperability, as the number of records for each page collected, the number of retries, etc.
Validators are tools that will help the authors, the librarians, the repository administrators to accomplish the guidelines and the standards. This tool should help to find errors, to recommend actions, to check the metadata schema for syntactic and content control. It will help to enhance the quality of our repository, to enhance the quality of the centralized service that harvests our content and will help (the most important of all) to the final user to discover our resources.

**LRMI / schema.org validation**

We are currently preparing some examples of LRMI metadata. While these are intended to be informative only, we know that they will affect implementations more than any normative text we could put into a spec—I mean what developer reads the spec when you can just copy an example? So it's important that the examples are valid, and that set me to pulling together a list of tools & services useful for validating LRMI, and by extension schema.org.

**Common things to test for:**

- Simple syntax errors produced by typos, not closing tags and so on.
- That the data extracted is valid schema.org / LRMI.
- Using properties that don't belong to the stated resource type, e.g. educationalRole should be a property of EducationalAudience not of CreativeWork.
- Loose or strict interpretation of expected value types, e.g. the author property should have a Person or Organization as its value, dates and times should be in iso 8601 format?
- Is the data provided for properties from the value space they should be? i.e. does the data provider use the controlled vocabulary you want?
- Check that values are provided for properties you especially require.

**[Hint]** if it is the last two that you are interested in then you’re out of luck for now, but do skip to the “want more” section at the end.
How it works?

<table>
<thead>
<tr>
<th>#</th>
<th>XML Origen</th>
<th>Trasformado</th>
<th>dc:format</th>
</tr>
</thead>
<tbody>
<tr>
<td>6499251</td>
<td>ver metadatos</td>
<td>Sí</td>
<td>ver metadatos, ver validación</td>
</tr>
<tr>
<td>6499252</td>
<td>ver metadatos</td>
<td>Sí</td>
<td>ver metadatos, ver validación</td>
</tr>
<tr>
<td>6499253</td>
<td>ver metadatos</td>
<td>Sí</td>
<td>ver metadatos</td>
</tr>
<tr>
<td>6499254</td>
<td>ver metadatos</td>
<td>Sí</td>
<td>ver metadatos</td>
</tr>
</tbody>
</table>

#### Repositorios detectados
- dspace.unapqitos.edu.pe
- dspace.unifsc.edu.pe
- repositorio.upao.edu.pe
- dspace.ipen.gob.pe
- repositorio.udeach.edu.pe
- cybertesis.unmsm.edu.pe
- preview.openrepository.com
- localhost
- www.repositorioacademico.us
- repository.unap.edu.pe

#### Diagnóstico repositorio: repositorio.udach.edu.pe

**dc:type**
- 4 reg. rechazados

**Válidos**
- Registros válidos

**dc:rights**
- rights:driver: info:eu-repo/semantics/openAccess

**dc:type**
- type:driver: info:eu-repo/semantics/bachelorThesis
- type:status: info:eu-repo/semantics/publishedVersion
- type:status: info:eu-repo/semantics/bachelorThesis

**dc:subject**
- subject:not_empty: 3
- subject:not_empty: 4

**dc:creator**
- creator:not_empty: 12
- creator:not_empty: 5

**dc:format**
- No hay ocurrencias de este campo
Methodology to select a technological solution to implement a Repository
Consider:

- **The nature of the material** *(What we'll deposit in the repository?)*
- **Features** of the platform *(Search, configurable forms, organization for the content?)*
- **Flexibility** to use the desired metadata schema *(We would create an IR with Theses and OERs, and to share each collection with different networks).*
- **Interoperability** features.
- **Community** of the platform. Frequency versions, licence.
- **Our context**: Local IT support, commercial, costs, technology, TCO, Other experiences...
Consider:

- **Escalability.** How my repository will grow? Could the platform support this growth?
- **Security.** For the content, for the accesses, for the service..
- **APIs.** Possibility to connect with another platforms (in the case of OERs for instance with Moodle, with academic systems, with Identity Federations, etc).
Main options:

- **Open source platforms**
  - DSpace
  - Eprints
  - Invenio
  - Fedora

- **Commercial platforms**
  - Commonly associated with a commercial Library Management System or a content provider.

- **SaaS (Software as a service)**
  - Eprints Services:
  - Dspace Direct:
  - Open Repository:
Get ahead on your repository.
Community analysis example

One important element, but not 1
DSpace. Detailed features

- Technologies: Java / Spring / PostgreSQL DB (or Oracle)/ Apache SolR (index)
- Last stable version: 5.2 (± 20 May 2015)
- Main functionalities:
  - Repository organization through a structure of communities and collections (and browse dynamic features). Granularity of rights management from community to file. Different authentication protocols (Shibboleth, LDAP).
  - Responsive UI: JSPUI y XMLUI. Optimization for Google & Google scholar.
  - Faceted search: More like this & Did you mean? Fulltext search.
  - Push information to the user. SID.
  - Deposit protocols SWORD 1.0 y 2.0, RSS, OpenSearch
  - Contribution workflow: Several levels of verification previous to publish.
  - REST API in progress/maturity. RDF export and SPARQL end point (with external tools)
  - Extensible metadata crosswalk model. Several out-of-the box
  - OAI-PMH with metadata transformation and filters. OAI-ORE
DSpace CRIS
Current Research Information System (CRIS) made easy with DSpace

Screenshot
The screenshot below shows some of the main features of the DSpace CRIS module with the sample configuration and basic layout.

For a more exhaustive tour of the DSpace CRIS capability please take a look to the institutional repository of The University of Hong Kong, The HKU Scholars Hub.

Researcher profile | Faceted Search & Browse | Administration UI

Researcher profile
Some specialized distributions of DSpace HR information & specialists CV integrated into the repository and related with their production.
Eprints software platform

Technologies : Perl / MySQL development based.
Last stable version : 3.3.14 (20 Apr 2015)
Main features :

- Repository organization through dynamically defined structures as subject, institutional divisions.
- Google scholar optimization.
- Plugins gallery : IRStats2, EdScholar, MePrints, ORCID, etc. Some of them have been reported to have issues between versions (or only ready for old versions).
- Deposit protocol SWORD v.1.3
- Self archive.
- API for building new plugins.
- OAI-PMH protocol.
- Metadata quality control. Massive metadata operation tools..
- More than one repository in same deployment.
Bienvenido a RICABIB: el Repositorio Digital Institucional del Centro Atómico Bariloche e Instituto Balseiro

utilizar los servicios del repositorio.

RICABIB

Visualizando una gama completa de campos. Utilizar el campo de búsqueda en el top de la página para una búsqueda

edificios

edificios

edificios

edificios

edificios

edificios

edificios

edificios

edificios

edificios

edificios

edificios

edificios

edificios

edificios

edificios
CDS Invenio

Technologies : Python/Cpython MySQL.
Last stable version : 1.21 (2015 May)
Main Features :
  - MARC21 cataloging
  - Scalability & replication. More than 1,000,000 DO in CERN deploy.
  - Static & virtual collections.
  - Index specific technology. Results cluster by collection. SolR plug-in accepted.
  - Self Archive features
  - OAI-PMH.
  - Open source project started in 2002. High maturity degree.
One example of integrating (and developing) tools on top of a common repository platform

http://www.jorum.ac.uk/

Our Background

Jorum is a Jisc funded Service for UK Further and Higher Education, to collect and share Open Educational Resources (OER), allowing their reuse and repurposing. Jorum is the first port of call for 1000s of resources, all shared and created by those who teach or have been inspired in the FE and HE and professional skills by Joram is provided by Mimas, part of the Jisc Digital Resources division; and is the UK's largest OER repository collections growing daily.

All of the resources shared are created under a Creative Commons licence.

(And, in case you were wondering, the word 'Jorum' is of Biblical origin and means a collecting (or drinking

Contact Us

Implementing the Learning Resource Metadata Initiative (LRMI) in Jorum

By Dr Ben Ryan, published July 7, 2014

Jorum and the LRMI

Jorum was delighted to receive grant funding from Creative Commons to participate in the Learning Resource Metadata Initiative (LRMI) project. Across the last 12 months, in parallel with 8 other funded US organisations, Jorum has been developing and testing features which will enhance both the resource discovery system and interoperability of Jorum content.

First of all ... What is LRMI?

The Learning Resource Metadata Initiative (LRMI) is an extension of schema.org that adds education-specific mark-up to the existing mark-up that schema.org defines. So what does schema.org do? Put simply it allows the addition of "structured data" to web pages that machines can understand without altering the way a web page works.
Test, test and more test...

- Making a Pilot.
  - Try to obtain a **sample of all the material types** that the repository should store and preserve.
  - Try to input the samples manually using the repository input form. Try to adapt the input forms. Document the problems found.
  - Functionality. Try to modify the search form, add criteria, modify the browse criteria according to your need. Document the problems found.
  - **Integration**. If you have an user identity provider, try to configure it and test with several users. Try to define different rights for the material (if the repository will restrict some of the content).
  - Test the tools for **import from previous/legacy systems** to the platforms to be evaluated. Consider results and if is there some things that should be arranged manually.
  - Put the things that were documented (as unsolved) in the technical support mail list. Evaluate time of response, quality, number and priority of the items that were solved.

**Why is not Fedora here?**

Probably because it will take another complete presentation. Fedora is built on top of more advanced, scalable and robust services (In my honest opinion) than currently any other IR software that we could use. It could be integrated with another tools as Islandora and is natively RDF enabled. But, in fact, out of the box, the solution is not “ready to go”.
How many initiatives are around the world using DR? (All repositories types) - 2014

http://repository66.org
Questions?
Thank you very much

Emiliano Marmonti
emarmonti@gmail.com
IAEA – June 2015