

# Uso de metadatos en depósitos institucionales españoles: una primera aproximación

Assumpció Estivill,

Ernest Abadal, Jorge Franganillo, Jesús Gascón, José Manuel Rodríguez Gairín  
Departament de Biblioteconomia i Documentació, Universitat de Barcelona

10as Jornadas Españolas de Documentación, Santiago de Compostela, mayo de 2007

## Sumario

1. Antecedentes y contexto de los depósitos institucionales
2. Depósitos institucionales españoles
3. Uso de metadatos en los depósitos institucionales españoles
4. Conclusiones

# 1 Antecedentes y contexto de los DI

- Desarrollo de los depósitos de documentos electrónicos:
  - Impulsados por el movimiento Open Acces.
  - Alternativa a los modelos tradicionales de comunicación científica (S. Harnad, R. Crow) / Complemento a los modelos establecidos de publicación académica (C. Lynch).
  - Inicialmente reúnen la producción intelectual de una disciplina con el objetivo de difundir rápidamente los resultados de investigación (p. ej.: arXiv, 1991- <http://arxiv.org>, Economics Research Network, <http://www.ssrn.com/ern/index.html>, PubMed Central, <http://www.pubmedcentral.nih.gov/>).
  - A partir de 2002, algunas universidades empiezan a crear depósitos institucionales (DI). Este hecho coincide con la disponibilidad de software que facilita la gestión de los depósitos y la asignación de metadatos conformes al estándar OAI-PMH (DSpace, CONTENTdm, Fedora, GNU Eprints, CDS Invenio...).

# 1 Antecedentes y contexto de los DI

- Definiciones de "depósito institucional":
  - "Conjunto de servicios que la universidad ofrece a los miembros de su comunidad para gestionar y difundir los materiales digitales creados por la misma institución y por sus miembros". (C. Lynch, 2003).
  - "Colección organizada de objetos digitales, de alcance institucional, con estructuras consistentes de datos y metadatos para objetos similares, que facilitan su descubrimiento por parte de las comunidades interesadas en aquellos objetos". (G. Payne, 2004).
  - "Depósito permanente, de alcance institucional, de obras digitales diversas producidas localmente (...) que está disponible para uso público y que ofrece sus metadatos a recolectores de metadatos". (ARL, 2006).

# 1 Antecedentes y contexto de los DI

- **Objetivos de los DI:**
  - Gestión y organización de los recursos depositados, incluida su preservación a largo plazo
  - Difusión y acceso
- **Contenidos:**
  - *Preprints* y *postprints* de artículos
  - Trabajos presentados en congresos
  - Tesis electrónicas
  - Recursos de aprendizaje
  - Informes de investigación y técnicos
  - Publicaciones institucionales
  - etc.
- **Los DI no tienen necesariamente todas sus colecciones de acceso abierto.**

Jornadas Españolas de Documentación  
(10as : 2007 : Santiago de Compostela)

5

The screenshot shows the homepage of the Dipòsit Digital de Documents de la UAB. The browser window title is "Dipòsit Digital de Documents de la UAB - Servei de Biblioteques: Pàgina inicial - Windows Internet Explorer". The address bar shows "http://ddd.uab.es". The page header includes the logo "Dipòsit Digital de Documents UAB" and navigation links: "visitant :: sessió :: cistells :: alertes :: identificació".

The main content area is titled "Pàgina inicial Dipòsit Digital de Documents de la UAB - Servei de Biblioteques". It features a search bar with the text "Cercar en 16,443 registres per:" and a search button labeled "Cerca". Below the search bar are "Consells de cerca" and "Cerca avançada" links.

There are two columns of filters:

- Limitar per col·lecció:**
  - Materials de curs** (6,156)
    - Gües de l'estudiant (18) Programes de curs (3,406) Exàmens (2,732)
  - Libres i col·leccions** (12)
    - Libres (12)
  - Publicacions periòdiques** (47)
    - Butlletins (14) Revistes (28) Altres publicacions (5)
  - Articles i informes** (4,257)
    - Articles publicats (4,257)
  - Multimèdia** (313)
    - Àudio i vídeo (56) Exposicions de les Biblioteques UAB (217)
  - Llengua i literatura catalanes** (1,921)
    - Història i crítica de la narrativa catalana contemporània (1,295) Traducció i recepció de les literatures estrangeres a la literatura catalana, 1892-1939 (626)
  - Documentació política** (1,14)
    - Premsa política clandestina (114)
  - Documentació empresarial** (75)
    - Informes anuals (3) Declaracions mediambientals (66) Informes de sostenibilitat (0)
  - Fons personals** (3,548)
    - José Agustín Goytisolo (3,548)
- Enfocat a:**
  - Material de curs UAB** (6,156)
    - Gües de l'estudiant (18) Programes de curs (3,406) Exàmens (2,732)
  - Libres i col·leccions de la UAB** (12)
    - Libres de la UAB (12)
  - Publicacions periòdiques de la UAB** (45)
    - Butlletins (14) Revistes de la UAB (26) Altres publicacions
  - Articles i informes de la UAB** (4,006)
    - Articles publicats a la UAB (4,006)
  - Multimèdia de la UAB** (300)
    - Àudio i vídeo de la UAB (83) Exposicions de les Biblioteques UAB (217)

The footer contains the text "Dipòsit Digital de Documents de la UAB - Servei de Biblioteques :: Cerca :: Llibre :: Personalitza :: Ajuda" and "Aquest lloc està també disponible en els següents idiomes: Català English Español". The taskbar at the bottom shows the system tray with the date "23/07" and the time "23:07".

Dipòsit Digital de Documents de la UAB - Servei de Biblioteques: Resultats de la cerca - Windows Internet Explorer

http://ddd.uab.cat/search?cc=hcnc&sc=1&p=soldevila&f=

Dipòsit Digital de Documents UAB

visitant :: sessió :: castells :: alertes :: identificació

Pàgina inicial > Bases de dades bibliogràfiques > Llengua i literatura catalanes > Història i literatura catalana contemporània > Resultats de la cerca

### Història i crítica de la narrativa catalana contemporània

Cerca: soldevila  qualsevol camp

Consells de cerca :: Cerca avançada

Cerca col·leccions: Història i crítica de la nar...

Ordena'ls per: el darrer primer  - o prioritzar per -

Història i crítica de la narrativa catalana contemporània 29 registres trobats 1 - 10  anar al registre:  La cerca s'ha fet en 0.07 segons.

- Fanny / Verdaguier, Mario**  
(Llibros catalanes)  
*La Vanguardia*, (1930, 20 febrer)  
<http://ddd.uab.es/uab/hicncc/1296.pdf>  
[Registre complet] - [Registres semblants]
- Valentina / Jordana, C.A**  
(Excursions literàries)  
*L'Opinió*, (1934, 17 Juny)  
<http://ddd.uab.es/uab/hicncc/1295.pdf>  
[Registre complet] - [Registres semblants]
- Barcelona i la novel·la / Tasis, Rafael**  
*Una visió de conjunt de la novel·la catalana*, p. 109-123  
<http://ddd.uab.es/uab/hicncc/1268.pdf>  
[Registre complet] - [Registres semblants]
- Carles Soldevila / Tasis, Rafael**  
*Una visió de conjunt de la novel·la catalana*, p. 93-97

Cornell Digital Repository: Communities and Collections - Windows Internet Explorer

http://dspace.library.cornell.edu/community-list

cornell university

Cornell University Library  
Cornell Digital Repository

Home | About Us | Browse | Search | Communities & Collections | Contact Us

Search Us:

Communities and Collections

Shown below is a list of communities and the collections and sub-communities within them. Click on a name to view that community or collection home page.

- Agriculture & Life Sciences**
  - Agricultural Ecology Program: Upper Susquehanna River Basin**
  - Agroecology**
    - Agroecology - Preprints [2]
    - Agroecology - Publications [0]
  - Animal Science**
  - Applied Economics and Management**
  - Biological Statistics and Computational Biology**
  - Biological and Environmental Engineering**
    - BEI - Papers, Research and Monographs [1]
  - Biomedical Sciences**
  - Biotechnology Program**
  - Communication**
    - Communication - Monographs, Research and Papers [4]
  - Crop and Soil Sciences**
    - Bio-char [5]
    - Cornell Waste Management Institute [19]
    - General Publications [1]
    - Geographic Information Systems [1]
    - Manuscripts and Water Quality Data [1]
    - What's Cropping Up? Newsletter
  - Events**
    - Mann Library Chats in the Stacks [1]

Sign on to:

# 1 Antecedentes y contexto de los DI

## ■ Algunos estudios:

- Alma Swan, Sheridan Brown (2005). *Open access self archiving: an author study*. Truro, UK: Key Perspectives.  
<http://eprints.ecs.soton.ac.uk/10999/>.
  - 100% de crecimiento de la práctica de depósito por parte de los autores en depósitos institucionales o temáticos entre 2003 y 2004.
  - 1.296 respuestas de todas las partes del mundo, de una población parcialmente experimentada en las prácticas de depósito.
- Association of Research Libraries (2006). *Institutional repositories: [executive summary]*. Washington: ARL, July 2006. (Spec kit ; 292). <http://www.arl.org/bm~doc/spec292web.pdf>.
  - 78% de las bibliotecas universitarias encuestadas en enero de 2006 tienen ya un DI en marcha o lo tendrán en 2007.

# 1 Antecedentes y contexto de los DI

## ■ Algunos estudios:

- Clifford A. Lynch, Joan K. Lippincott (2005), "Institutional repository deployment in the United States as of early 2005", *D-lib magazine*, vol. 11, no. 9. <http://www.dlib.org/dlib/september05/lynch/09lynch.html>.
  - 40% de las universidades encuestadas tienen un DI en funcionamiento; 88% de las que no lo tienen están en la fase de proyectarlo.
- G. van Westrienen, Clifford A. Lynch (2005) "Academic institutional repositories: deployment status in 13 nations as of mid 2005". *D-lib magazine*, vol. 11, no. 9. <http://www.dlib.org/dlib/september05/westrienen/09westrienen.html>.
  - Alemania, Noruega y los Países Bajos: 100% de las universidades.
  - Australia: 95%.
  - Suecia, Bélgica y Dinamarca: entre el 65 y el 50%.
  - Reino Unido e Italia: 22%.
  - Con la excepción de Bélgica, la mayoría de los depósitos tienen unos pocos centenares de documentos depositados.

# 1 Antecedentes y contexto de los DI

## ■ Algunos estudios:

- Philip M. Davis, Matthew J.L. Conolly (2007), "Institutional repositories: evaluating the reasons for non-use of Cornell University's installation of DSpace", *D-lib magazine*, vol. 13, no. 3-4. <http://www.dlib.org/dlib/march07/davis/03davis.html>.
  - Poco uso del DI de Cornell University —y también de otros DI de universidades americanas— por parte de los investigadores. Razones:
    - Poca visibilidad —depósitos aislados
    - Software con pocas funcionalidades
    - Preferencia por los depósitos temáticos
    - Normas culturales de cada disciplina: preferencia por los depósitos temáticos, donde deposita documentos la comunidad de la disciplina; reticencia a compartir resultados antes de su publicación formal; publicar en determinadas revistas da una impronta que no se consigue con la mera diseminación de trabajos.

# 1 Antecedentes y contexto de los DI

## ■ Algunos estudios:

- José Antonio Frías, Tania Santiago. "Las universidades españolas y las iniciativas OAI". *Repositorios institucionales: una vía hacia el acceso, la visibilidad y la preservación de la producción científica. Jornadas Científicas y Talleres* (Zaragoza, 14 y 15 de diciembre de 2006).  
<http://redsicura.iata.csic.es/xarxa/ocs/papers/Frias-Zaragoza.ppt>.
  - Referencia depósitos en funcionamiento o en fase de proyecto en 53 universidades españolas.
  - Valoración global: desarrollo incipiente de los DI españoles en relación con sus contenidos y visibilidad.

## 2 Depósitos institucionales españoles

- Los 16 DI españoles que constan en el estudio se han identificado en los registros de depósitos y recolectores de metadatos siguientes:
  - Open Archives Initiative. *Registered data providers*, <http://www.openarchives.org/Register/BrowseSites>.
  - *Registry of open access repositories (ROAR)*, <http://roar.eprints.org/>.
  - *OpenDOAR: directory of open access repositories*, <http://www.opendoar.org/>.
  - *OAIster*, <http://www.oaister.org/>.
  - *Directorio y recolector de recursos digitales (DR)*, <http://roai.mcu.es/es/inicio/inicio.cmd>.

## 2 Depósitos institucionales españoles

Depósito	Software	Nº de registros
RACO: revistes catalanes d'accés obert — CBUC	Open Journal Systems	31.585
Dipòsit digital de documents — Universitat Autònoma de Barcelona	CDS Invenio	16.508
TDX: tesis doctorals en xarxa — CBUC (contiene tesis de 17 universidades)	ETD-db	4.375
E-Prints — Universidad Complutense de Madrid	GNU Eprints	3.968
RECERCAT: dipòsit de la recerca de Catalunya — CBUC	DSpace	3.293
E-SpacioUNED — Universidad Nacional de Educación a Distancia	Fedora	2.041

## 2 Depósitos institucionales españoles

Depósito	Software	Nº de registros
DSpace revistas — Universitat Politècnica de Catalunya	DSpace	1.706
Treballs acadèmics — Universitat Politècnica de Catalunya	No consta	1.387
Dialnet — tesis (contiene tesis de 6 universidades)	No consta	792
Revistas electrónicas — Universidad de Navarra	DSpace	671 (11 revistas)
E-archivo — Universidad Carlos III de Madrid	DSpace	617
DSpace E-prints — Universitat Politècnica de Catalunya	DSpace	357

## 2 Depósitos institucionales españoles

Depósito	Software	Nº de registros
UDCDSpace — Universidade da Coruña	DSpace	339
DSpace a la Universitat de Girona	DSpace	132
E_buah: biblioteca digital — Universidad de Alcalá	DSpace	181
DIGIBUG — Universidad de Granada	Consulta mediante el catálogo general	No consta



## 2 Depósitos institucionales españoles

### ■ Características de los DI españoles:

- Contenidos:
  - Depósitos que contienen una única tipología de recursos (p. ej., tesis, artículos de revista, etc.).
  - Depósitos que contienen diversas tipologías de documentos (p. ej., artículos de revista, *preprints*, comunicaciones en congresos, publicaciones institucionales, recursos docentes, colecciones especiales, etc.). Son los más numerosos.
- Organización:
  - Diversos DI por institución según el tipo de recurso (p. ej.: UPC)
  - Un solo DI:
    - Organizados por colecciones que responden a la tipología documental.
    - Organizados por comunidades y colecciones —como es el caso del software DSpace.
  - Participación en DI colectivos (p. ej.: RACO, TDX, Recercat, Dialnet—Tesis).

## 2 Depósitos institucionales españoles

### ■ Características de los DI españoles:

- Software de gestión:
  - DSpace es el software dominante (8 instalaciones en los DI examinados).
  - Otros programas utilizados: GNU Eprints (UCM), ETD-db (TDX, CBUC), Fedora (UNED), CDS Invenio (UAB). Este último tiene el formato MARC como formato de base y trabaja con MARC 21, MARC XML i DC).
- Presencia en registros, directorios, proveedores de servicios OAI:
  - Representados en OAI-PMH, OAIster, Open DOAR i ROAR: 9 DI.
  - Representados en OAI-PMH, OAIster i ROAR: 1 DI.
  - Representados en OAI-PMH i ROAR: 1 DI.
  - Representados en el Directorio y recolector del Ministerio de Cultura: 12 DI.
  - Sólo 4 DI han optado por no darse de alta en los servicios anteriores.
- Implantación continuada de nuevos DI (p. ej.: Universidad Rey Juan Carlos, Universidad de Alicante, Universitat de Barcelona).

### 3 Uso de metadatos en los DI españoles

- Análisis inicial basado en la observación. Posteriormente se hará un estudio estadístico más sistemático.
- Se han considerado los 15 elementos básicos del DC sin calificar más el identificador específico <bibliographicCitation>.
- En general, los DI españoles usan un amplio número de metadatos en la descripción:
  - 11 DI usan, como mínimo, 8 elementos del DC —13 si se consideran usos más esporádicos de <subject> y <description>.
  - Elementos más frecuentes:
    - <title>, <creator> y <identifier>, presentes en todos los DI
    - <date> y <type>, en 14 y 11 DI respectivamente
    - <rights>, en 10 DI
    - <subject> y <contributor>, en 9 DI
    - <publisher>, <format> y <language>, en 8 DI
    - <description>, en 7 DI (14 si se consideran usos más esporádicos)
    - <relation> sólo aparece en algunos registros de *E-prints* de la UCM
    - <coverage> y <source> no se utilizan en ningún DI.

### 3 Uso de metadatos en los DI españoles

- Los DI de universidades españolas contienen más metadatos que los mostrados en estudios anteriores (Ward, 2004) → 3 años son tiempo suficiente para justificar esta mejora.
- Los metadatos de los distintos depósitos son bastante equiparables en cuanto a los elementos que contienen → predominio del software DSpace.
- En general están conformes al OAI-PMH.
- Sin embargo:
  - Los contenidos de los elementos no están normalizados.
  - Sólo 4 DI incluyen, de manera bastante sistemática, los 8 elementos que se consideran básicos (Shreeves, 2005): <title>, <creator>, <subject>, <description>, <date>, <format>, <identifier> y <rights>.
  - La mayoría de los DI incluyen algunas directrices elementales sobre políticas de colección, pero entre su documentación pública no se encuentran directrices que guíen la asignación de metadatos, el uso de lenguajes controlados, etc.

### 3 Uso de metadatos en los DI españoles

- Características que definen la calidad de los metadatos:
  - Bruce y Hillmann (*Metadata in practice*, 2004):
    - Completos y correctos
    - Control de la procedencia (fuente primera y modificaciones ulteriores)
    - Conformes a las expectativas del perfil de aplicación y de los usuarios
    - Documentación de las decisiones
      - Directrices para la asignación de metadatos y ejemplos de buenas prácticas
      - Documentación y disponibilidad de los vocabularios utilizados
    - Consistencia y coherencia en el depósito, pero también metadatos consistentes con los de otras colecciones → búsquedas federadas
    - Puntualidad y actualización de los metadatos
    - Accesibilidad.
  - Tennant (*Bitter Harvest*, 2004?)
    - El DC simple es demasiado simple:
      - Ofrecer metadatos más completos a los recolectores (DC calificado)
      - Ofrecer metadatos en otros formatos (MARC XML, MODS, etc.)

### 4 Conclusiones

- ¿Objetivos i usuarios de los DI?
  - Usuarios internos de la institución tanto desde el punto de vista de la preservación, como de la difusión:
    - ¿Qué tipos de documentos hay que archivar?
    - ¿Son conocidos los DI por los investigadores? ¿Los utilizan? ¿Cómo se promueve su uso? ¿Cómo se evalúa?
    - ¿Tiene interés/utilidad que los *preprints/postprints* del Dep. de Lingüística General convivan con los del Dep. de Biología Vegetal?
    - ¿Cuál es el nivel de interoperabilidad semántica de los metadatos con otros depósitos de la institución? ¿Y con el catálogo?
  - Uso público:
    - La mayoría de los DI aseguran que sus contenidos estén presentes en los buscadores de Internet y en los recolectores de metadatos.
    - Falta de normalización de los contenidos de los metadatos, diversidad lingüística de los metadatos, incapacidad de los proveedores de servicios de eliminar duplicados, de agrupar autores, versiones, etc.