

Palabras clave

Repositorio institucional, Open Access, Consejo Superior de Investigaciones Científicas, Publicación científica

Resumen

Con ocasión de las IX Jornadas TecniMap, celebradas en Sevilla el año 2006, el CSIC presentó una comunicación titulada “Hacia una estrategia Open Access en el Consejo Superior de Investigaciones Científicas”. En dicha comunicación se expresaba la intención de poner en marcha un repositorio institucional que albergara toda la documentación científica generada por los investigadores del CSIC. Esta plataforma constituiría la culminación de una estrategia Open Access de la institución, iniciada en enero de 2006 con la firma por parte del CSIC de la Declaración de Berlín sobre Acceso Abierto al Conocimiento en la esfera de las Ciencias y las Humanidades. Hoy, año y medio después de aquel artículo, podemos anunciar que el repositorio institucional Digital.CSIC es ya una realidad, que nace con la vocación de convertirse en la “memoria científica institucional” del Consejo. A medida que vaya desarrollándose, Digital.CSIC servirá como almacén organizado de toda la producción científica de la institución en formato electrónico, contribuyendo a su preservación y a su mejor difusión en un modelo de acceso libre.

Autores**Clara Giner Durán**

Centro Técnico de Informática CSIC

Julio Pemau Alonso

Centro Técnico de Informática CSIC

Pablo de Castro Martín

Unidad de Coordinación de Bibliotecas del CSIC. Oficina Técnica del Proyecto Digital.CSIC

Luisa Domènech García

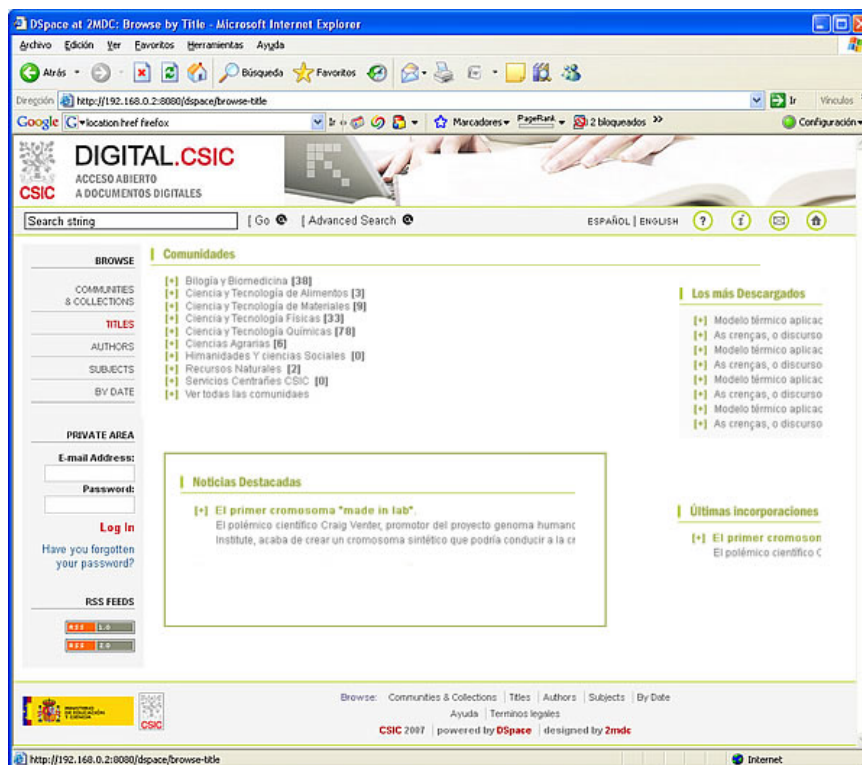
Unidad de Coordinación de Bibliotecas del CSIC. Oficina Técnica del Proyecto Digital.CSIC

El Repositorio Institucional Digital.CSIC: la materialización de la estrategia Open Access del Consejo Superior de Investigaciones Científicas

Con ocasión de las Jornadas TecniMap 2006, celebradas en Sevilla a finales de mayo y principios de junio del pasado año, tres integrantes del área de Servicios Centrales del CSIC presentaron una comunicación titulada “Hacia una estrategia Open Access en el Consejo Superior de Investigaciones Científicas” [RODRÍGUEZ ALCALDE, A. *et al*, 2006].

En dicha comunicación se expresaba la intención de poner en marcha un repositorio institucional que albergara toda la documentación científica generada por los investigadores del CSIC. Esta plataforma constituiría la culminación de una estrategia *Open Access* de la institución, iniciada en enero de 2006 con la firma por parte del CSIC de la Declaración de Berlín sobre Acceso Abierto al Conocimiento en la esfera de las Ciencias y las Humanidades [BDOA, 2003].

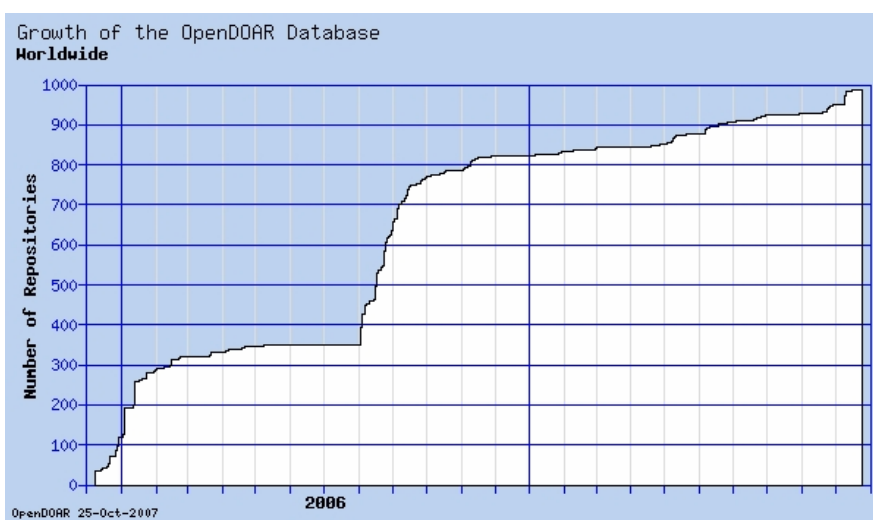
Hoy, año y medio después de aquel artículo, desde el Consejo Superior de Investigaciones Científicas podemos anunciar que el repositorio institucional Digital.CSIC es ya una realidad, que nace con la vocación de convertirse en la “memoria científica” de la institución. Se pretende que, a medida que vaya desarrollándose, Digital.CSIC sirva como almacén organizado de toda la producción científica de la institución en formato electrónico, contribuyendo a su preservación y a su mejor difusión en un modelo de acceso libre.



La puesta en marcha de Digital.CSIC conllevará numerosas ventajas, tanto para la institución como para sus investigadores. Por un lado, la publicación en acceso abierto producirá un cuantioso aumento de la visibilidad de los trabajos científicos del personal investigador del CSIC. Desde el punto de vista de la institución, el repositorio facilitará la organización y gestión de las publicaciones del CSIC, así como su preservación. Eventualmente, Digital.CSIC podría devenir una herramienta de interés en las tareas de evaluación del personal científico del CSIC.

El artículo de TECNIMAP 2006 mencionado anteriormente concluía anunciando la selección del software de código abierto DSpace por parte del CSIC para la construcción de su repositorio institucional. Desde entonces se ha recorrido un largo camino de recogida de información, diseño, desarrollo e implantación técnica de la plataforma por parte de diferentes servicios de la institución, señaladamente la Unidad de Coordinación de Bibliotecas (CBIC) y el Centro Técnico de Informática (CTI) del CSIC. Mucha de esta labor de desarrollo preliminar se ha llevado a cabo en colaboración con otros organismos españoles, sobre todo universidades, que se encontraban igualmente diseñando sus propios repositorios institucionales.

La creación de repositorios ha sido una tarea abordada por multitud de instituciones científicas y académicas en los últimos dos años. Aunque plataformas de naturaleza temática como ArXiv [ARXIV, 2006] datan de quince años atrás, las primeras iniciativas para el desarrollo y puesta en marcha de archivos institucionales comenzaron a desarrollarse como resultado de la Declaración de Berlín sobre Acceso Abierto (22 de octubre de 2003). El registro más exhaustivo de repositorios institucionales y temáticos en la actualidad es el directorio OpenDOAR (*Directory of Open Access Repositories*), un recurso creado en febrero de 2005. En la gráfica de la página siguiente se recoge el número de plataformas registradas en OpenDOAR y su progresión desde finales de 2005. Desde entonces, a medida que se iba expandiendo la filosofía del acceso abierto, se ha venido produciendo un crecimiento considerable del número de repositorios institucionales registrados, hasta alcanzar recientemente una cifra próxima a los 1000 repositorios en todo el mundo.



En las estadísticas de ámbito geográfico de OpenDOAR se observa la presencia de organizaciones de los cinco continentes, aunque las regiones de mayor desarrollo son Estados Unidos y Europa, por albergar un mayor número de instituciones científicas y académicas de primer nivel. España, con dieciocho archivos institucionales recogidos en el directorio a finales de octubre de 2007, presenta una posición relativamente avanzada en las estadísticas de repositorios institucionales en Europa, gracias a la incorporación al registro de numerosas universidades y también, próximamente, del CSIC.

Tabla 1. Repositorios institucionales en España

Archivo Abierto Institucional de la Universidad Rey Juan Carlos	Universidad Rey Juan Carlos, Spain
Archivo Digital UPM	Universidad Politécnica de Madrid. Biblioteca, Universidad Politécnica de Madrid, Spain
e_buah (Biblioteca Digital de la Universidad de Alcalá)	Universidad de Alcalá, Spain
DIALNET OAI Theses	Universidad de La Rioja, Spain
Diposit Digital de la Universitat de Barcelona	Universitat de Barcelona (UB), Spain
DSpace a la Universitat de Girona	The Ferrater Mora Chair of Contemporary Thought, Universitat de Girona, Spain
DSpace.E-prints - Universitat Politècnica de Catalunya	Universitat Politècnica de Catalunya (UPC), Spain
Universitat Politècnica de Catalunya (UPC) - Treballs acadèmics UPC	Universitat Politècnica de Catalunya (UPC), Spain
E-Archivo	Biblioteca, Universidad Carlos III de Madrid, Spain
EPrints Complutense (E-PrintsUCM)	Biblioteca de la Universidad Complutense de Madrid, Universidad Computense Madrid (UCM), Spain
Instituto de Empresa Working Papers	Instituto de Empresa Business School, Spain
Invenia Repository for Technological Innovation	Idhea, Spain
Repositorio Institucional de la Universidad de Alicante (RUA)	University of Alicante, Spain
Research Repository of Catalonia (RECERCAT)	Consorci de Biblioteques Universitàries de Catalunya (CBUC), Spain
RACO (Revistes Catalanes amb Accés Obert)	Consorci de Biblioteques Universitàries de Catalunya (CBUC), Spain
Tesis Doctorals en Xarxa (TDX)	Consorci de Biblioteques Universitàries de Catalunya (CBUC), Spain
Universidade da Coruña UDCDSpace (UDCDSpace)	Universidade da Coruña, Spain
Virtual Library of Historical Press	Ministerio de Cultura, Reino de España, Spain

Las estadísticas sobre repositorios en OpenDOAR no recogen, por el momento, una clasificación por la naturaleza de la institución que permita distinguir los pertenecientes a organismos científicos de los de organismos académicos, entre otras razones porque a menudo resulta complicado establecer las fronteras entre ambos tipos de organización. En todo caso, de una rápida revisión de las instituciones incluidas en el directorio se deduce que aunque en esta iniciativa es mayoritaria la presencia de universidades, son también numerosos los organismos de investigación de ámbito nacional –de naturaleza similar al CSIC- que cuentan con un repositorio institucional más o menos desarrollado. Entre ellos cabe señalar los *Research Councils* en Gran Bretaña, el CNRS francés a través del *Centre pour la Communication Scientifique Directe* (CCSD), la Sociedad Max Planck en Alemania o el CNR en Italia. En los desarrollos más avanzados de este tipo de recurso es característica la creación de redes de repositorios interconectados, en cada uno de los cuales se almacena información de naturaleza diferente. Presentan particular interés los repositorios de los llamados *datasets*, archivos de datos brutos a partir de los cuales se realizan los trabajos de investigación, como por ejemplo el *eCrystals* de la Universidad de

Southampton [ECRYSTALS, 2006], un archivo digital de estructuras cristalográficas obtenidas mediante técnicas de rayos X.

Tabla 2. Repositorios de instituciones científicas recogidos en OpenDOAR

ESRC Social Sciences Repository	Economic and Social Research Council (ESRC), United Kingdom
Marine & Ocean Science ePrints Archive @ Plymouth	National Marine Biological Library (NMBL), United Kingdom
NERC Open Research Archive (NORA)	Natural Environment Research Council (NERC), United Kingdom
Research Findings Register (ReFeR)	United Kingdom Department of Health , United Kingdom
STFC ePublication Archive (STFC ePubs)	Science and Technology Facilities Council (STFC), United Kingdom
eCrystals - Southampton	University of Southampton (Soton), United Kingdom
Archive EduTice	Domaine scientifique de la Doua, Centre pour la Communication Scientifique Directe (CCSD), France
Archive Electronique - Institut Jean Nicod	Institut Jean Nicod (Institut Nicod), France
Archive Ouverte en Sciences de l'Information et de la Communication (@rchiveSIC)	Domaine scientifique de la Doua, Centre pour la Communication Scientifique Directe (CCSD), France
HAL (Hyper Article en Ligne)	Centre pour la Communication Scientifique Directe (CCSD), France
INRIA a CCSD electronic archive server (HAL-INRIA)	Institut National de Recherche en Informatique et en Automatique (INRIA), France
Mémoires en Sciences de l'Information et de la Communication (mémSIC)	Urfist PacaC (Unité Régionale de Formation à l'Information Scientifique et Technique (Provence-Alpes-Côte d'Azur, Corse)), INIST-CNRS (Institut de l'Information Scientifique et Technique du CNRS), France
Thèses en Ligne (TEL::ACCUEIL)	Domaine scientifique de la Doua, Centre pour la Communication Scientifique Directe (CCSD), France
Max Planck Institute for the Study of Societies Publications (MPIfG Publications)	Max-Planck-Institut für Gesellschaftsforschung (MPIfG), Germany
Max Planck Society eDoc Server	Max Planck Gesellschaft (MPG), Germany
CNR Bologna Research Library	Consiglio Nazionale delle Ricerche (CNR), Italy
E-LIS	AePIC (Advanced e-Publishing Infrastructures), CILEA (Consorzio Interuniversitario Lombardo per l'Elaborazione Automatica), Italy
Earth-prints Repository	Istituto Nazionale di Geofisica e Vulcanologia (INGV), Italy
ICTP Open Access Archive (ICTP OAA)	Abdus Salam International Centre for Theoretical Physics (ICTP), Italy

La tipología de documentos que se almacenan en repositorios institucionales como los que aparecen en las dos tablas anteriores es muy variada y depende en gran medida de la naturaleza de las instituciones propietarias. Así, las universidades cuentan generalmente con amplios archivos de tesis doctorales, documentos de trabajo y artículos de las revistas que editan como documentos característicos de su producción institucional, en tanto que los organismos de investigación presentan unos contenidos con presencia predominante de artículos de revistas, trabajos de investigación, patentes o los ya mencionados *datasets*.

Antes de comenzar el desarrollo del repositorio institucional Digital.CSIC, y con el fin de determinar el grado de apoyo del personal científico del CSIC a la publicación en acceso abierto, así como el conocimiento previo que los investigadores pudieran tener sobre esta iniciativa, desde la Unidad de Coordinación de Bibliotecas del CSIC se realizó una encuesta al personal científico de la institución entre mayo y junio de 2007.

Puede consultarse un resumen de los resultados de la misma en http://www.csic.es/cbic/documents/docnoticias/encuesta_OA.pdf

Los resultados de la encuesta [CBIC, 2007] invitan a un moderado optimismo: aunque no son muchos los autores que por el momento han publicado en revistas de acceso abierto, sí hay bastantes que sistemáticamente están archivando sus artículos (*pre-prints* o *post-prints*) en repositorios temáticos, y, sobre todo, en la mayoría de las respuestas se detecta un estado de ánimo favorable a participar en la iniciativa una vez se ponga en marcha la herramienta de almacenamiento apropiada y se les explique de qué manera utilizarla.

Contando con este apoyo de base a la iniciativa de impulsar la estrategia *Open Access* del CSIC, desde la Vicepresidencia de Investigación Científica y Técnica del CSIC se afrontó el diseño y desarrollo de un repositorio institucional. Con este fin se creó un Grupo Asesor en el que, bajo la dirección de la Unidad de Coordinación de Bibliotecas del CSIC, estaban representados la propia Unidad de Coordinación, el Centro Técnico de Informática (CTI), el Servicio de Publicaciones del CSIC, la directora de una revista científica y la mencionada Vicepresidencia. Este grupo de trabajo diseñó la política y la estructura matriz del archivo digital, así como los procedimientos de desarrollo y la tipología de documentos que se almacenarían en el recurso. Una vez seleccionado el programa de *software* de código abierto DSpace como base para la plataforma tecnológica de la aplicación, el desarrollo e implantación del mismo se puso en manos del CTI, que después de varios meses de trabajo estableció las especificaciones técnicas actualmente vigentes.

Aspectos Técnicos de Digital.CSIC

El recurso consta de dos aplicativos:

- Un repositorio de los documentos digitales del CSIC basado en DSpace (software abierto y gratuito <http://www.dspace.org/>) que permite almacenar, gestionar y publicar en la web los documentos digitales.
- Proveedor de metadatos basado en el protocolo OAI-PMH (<http://www.openarchives.org/OAI/openarchivesprotocol.html>) para el intercambio de metadatos. Módulo incluido en DSpace.

Para adaptar el software DSpace a los requisitos del proyecto han sido necesarias algunas modificaciones y nuevos desarrollos que se describen a continuación.

- Nuevo diseño gráfico
- Integración de DSpace en el sistema de autenticación centralizada del CSIC contra un LDAP corporativo. Para ello se utiliza el software Acegi Security con el *Central Authentication Service, CAS*, que es un proyecto de software libre desarrollado inicialmente por la Universidad de Yale y que pasó a ser un proyecto JA_SIG en el año 2004.

Para obtener más información sobre estos sistemas visitar los sitios:

Acegi Security: <http://www.acegisecurity.org/>

CAS: <http://www.ja-sig.org/products/cas/>

JA_SIG: <http://www.ja-sig.org/>

- Carga automática de usuarios con privilegios para depositar documentos, seleccionando el personal científico en el Directorio Corporativo del CSIC (aproximadamente 2600 personas)
- Mantenimiento de PermaLink para cada uno de los documentos almacenados en el repositorio a través de Handle.net (<http://www.handle.net/rfc/rfc3651.html>)
- Desarrollo de un software para gestionar cargas masivas de datos, obtenidas de diversas fuentes a través de documentos XML. El desarrollo se ha realizado en JAVA, basado en la herramienta XMLBeans, y esquemas XSD obtenidos de los documentos XML.
- Posible modificación del módulo de estadísticas de DSpace y utilización de la herramienta WebTrends para la elaboración de estadísticas de acceso personalizadas.

Tanto DSpace como los desarrollos complementarios han sido programados en JAVA.

Los tres entornos (desarrollo, preproducción y producción) están soportados en máquinas virtuales gestionadas con Vmware y con la siguiente configuración:

- Sistema Operativo: Linux RedHat Enterprise
- Gestor de Base de Datos: Oracle 10g
- Servidor web: Apache 2
- Contenedor de Servlets: Tomcat 5.5
- Software del Repositorio: DSpace 1.4.2

Servicios de valor añadido

Desde el comienzo del proyecto se consideró que, aparte de las especificaciones técnicas de carácter general, sería recomendable el desarrollo de aplicaciones de software de desarrollo propio (en forma de *Application Programming Interfaces* o *APIs*) que dotaran al repositorio institucional de servicios de valor añadido para su mejor acogida por parte del personal científico del CSIC. Por su especificidad y por la asincronía de su desarrollo respecto al resto de tareas relacionadas con la puesta en marcha del repositorio, el



diseño y desarrollo de estas herramientas de valor añadido se ha llevado a cabo en colaboración con la empresa 2mdc.com.

La consideración principal para la oportunidad de este tipo de utilidades es la certeza por parte de la institución de que el mayor reto al que se enfrenta en la implantación de su repositorio es el de conseguir el apoyo de sus investigadores y personal científico. Aunque desde el punto de vista tecnológico y de organización de la información institucional resulta evidente que el proyecto es en sí mismo bueno para la institución, y por tanto indirectamente para sus miembros, cabe esperar que una parte importante de los potenciales usuarios de Digital.CSIC no colabore en el mismo por falta de lo que podrían denominarse “incentivos a corto plazo”.

Aunque la dirección del CSIC podría recurrir a instrumentos para el obligado cumplimiento del archivo del material de carácter científico por parte de sus investigadores, a priori considera preferible optar por estrategias que permitan a los usuarios obtener beneficios directos del archivado de su producción científica (no sólo artículos sino también presentaciones en formato digital, fotografías, ilustraciones, etcétera). Es en este ámbito donde cobra sentido el desarrollo de servicios de valor añadido para los usuarios registrados en Digital.CSIC.

La apuesta pasa por el desarrollo de un *API* que permita la explotación de los datos de Digital.CSIC desde cualquier *website* mediante el copiado y pegado de sencillos *snippets*¹ de código HTML/JavaScript con, entre otros, los siguientes objetivos:

- Permitir que los investigadores incluyan en sus páginas personales un listado de sus artículos con los enlaces pertinentes para su descarga directa, a la manera de un *Currículum Vitae* de publicaciones.
- Incorporar cajas de búsqueda en cualquier página *web* que permitan explorar el repositorio institucional.
- Permitir que los *websites* de Centros, Departamentos y Grupos del CSIC puedan disponer de toda la información, convenientemente filtrada, que sobre los mismos se encuentre en el repositorio. Por ejemplo, podrían incluirse listados de los artículos más recientes, etcétera.
- Facilitar a los usuarios del CSIC la creación de sus propias listas de distribución de artículos para notificar a sus colegas la aparición de nuevos artículos insertados en Digital.CSIC, y que éstos puedan descargarlos directamente desde su cliente de correo.
- Combinar esta API con las de *Google Scholar* o *Thomson-ISI* de forma que puedan encontrarse rápidamente las citas a un determinado artículo, permitiendo la creación de CVs interactivos actualizados en tiempo real.

El desarrollo de esta API podría, además, constituir un elemento de interés en sí mismo para la institución, pues podría distribuirse como código abierto entre el amplio colectivo de repositorios institucionales, dando una visibilidad adicional al CSIC en el marco de la iniciativa *Open Access*.

¹ A la manera en que lo hacen las APIs de Google, Flickr, YouTube,...

Bibliografía

-ARXIV, 2006. arXiv.org. Open Access e-prints archive. Cornell University Library

[<http://arxiv.org/>]

-BDOA. Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. Berlin, 22 de octubre de 2003

[<http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html>]

- CBIC, 2007. Publicaciones científicas y archivo de trabajos en repositorios digitales: encuesta de opinión en el CSIC (23 mayo – 22 junio 2007)

[http://www.csic.es/cbic/documents/docnoticias/encuesta_OA.pdf]

-DSPACE, <http://www.dspace.org/>

- ECRYSTALS, 2006. Crystal Structure Report Archive

[<http://ecrystals.chem.soton.ac.uk/>]

-OPENDOAR. Directory of Open Access Repositories.

[<http://www.opendoar.org/>]

-RODRÍGUEZ ALCALDE, A., PONSATI OBIOLS, A., CALA RIVERO, C., 2006. Hacia una estrategia Open Access en el Consejo Superior de Investigaciones Científicas.

[http://www.csi.map.es/csi/tecnimap/tecnimap_2006/03T_PDF/hacia%20una%20estrategia%20open.pdf]