

DIGITAL.CSIC: EL DESARROLLO DE UN REPOSITORIO INSTITUCIONAL CIENTÍFICO EN EL CSIC

De Castro, Pablo

Oficina Técnica Digital.CSIC
Unidad de Coordinación de Bibliotecas del CSIC
Consejo Superior de Investigaciones Científicas
pcastro@bib.csic.es

Abstract

En un momento en el que empiezan a materializarse en España políticas institucionales de impulso del acceso abierto a los resultados de la investigación financiada con fondos públicos, resulta particularmente oportuno conocer la evolución y el estado de desarrollo de las infraestructuras de archivo que habrán de dar soporte a dichas políticas. En el caso particular del CSIC, año y medio después de la puesta en marcha del repositorio institucional del Consejo Superior de Investigaciones Científicas Digital.CSIC, <http://digital.csic.es/>, su base de datos almacena ya más de 10.000 trabajos de investigación consultables a texto completo en acceso abierto. Digital.CSIC se encuentra así en pleno proceso de consolidación como memoria digital institucional del CSIC, avanzando en su objetivo de albergar toda la producción científica de la institución en sus diversos formatos documentales. En esta presentación se pretende mostrar la evolución del proyecto Digital.CSIC desde sus inicios, con la doble intención de servir como estudio de caso del desarrollo de un repositorio institucional y de proponer un modelo para otras instituciones que deseen contar con su propio archivo de publicaciones en acceso abierto. En la presentación se examina también el estado de desarrollo actual de Digital.CSIC y sus retos futuros.

At a time when institutional policies promoting Open Access to the results of public-funded research are starting to flourish in Spain, it is particularly interesting to know about the evolution and state of development of the archiving infrastructures that will serve as foundations for those policies. In the case of the Spanish National Research Council (CSIC), the Digital.CSIC Institutional Repository, <http://digital.csic.es/>, is steadily consolidating as one of the institution's main scientific communication tools. A year and a half after its release in January 2008, the Digital.CSIC database already holds over 10.000 open-access full-text documents and keeps growing. The long-term objective of the Digital.CSIC repository is to become a digital institutional memory where the organisation's whole research output is safely kept. In this communication the evolution of the Digital.CSIC Project is presented from its first step –the official signature of the Berlin Declaration by the CSIC back in January 2006– to the present. The case study of the development of a rather large institutional repository is as well intended to serve as an example for other organisations willing to have their own repository.

Keywords

Repositorios institucionales, Acceso abierto, Consejo Superior de Investigaciones Científicas (CSIC), Comunicación científica

Institutional repositories, Open Access, Spanish National Research Council (CSIC), Scientific communication

La adhesión institucional del CSIC el 31 de enero de 2006 a la Declaración de Berlín¹ por el acceso abierto al conocimiento en Ciencias y en Humanidades supuso el primer paso del proyecto Digital.CSIC para el desarrollo y puesta en marcha de un Repositorio Institucional en el CSIC. Promovida por el Presidente del Consejo Superior de Investigaciones Científicas a través de la Vicepresidencia de Ciencia y Tecnología (VICYT), la iniciativa Digital.CSIC adquirió su diseño inicial en las reuniones del Comité Asesor para la Creación del Repositorio Institucional del CSIC. Bajo la responsabilidad de la Unidad de

Coordinación de Bibliotecas (CBIC), el Comité Asesor fue definiendo la estructura que debería tener el futuro Digital.CSIC, los contenidos que debería almacenar y las directrices para la organización de la información, las políticas de archivo en función de los derechos de propiedad intelectual, los formatos de trabajo, etc². En este nivel preliminar del proyecto se definieron también las necesidades de software y recursos humanos que se requerirían para la puesta en marcha de una aplicación. Tras un estudio exhaustivo de las diversas opciones de software –tanto propietario como de código abierto– disponibles para servir de base al desarrollo del repositorio institucional, el Comité Asesor se decantó por utilizar DSpace, un software de código abierto desarrollado por el MIT y HP Labs, por considerar que era la solución más consolidada y la que mejor se adaptaba a las necesidades del proyecto.

4	22-oct-03	Educación Cultura y Deportes Gobierno de Canarias	Governmental Institution/Council
136	10-oct-05	Universitat Jaume I	University/College/Academy/Higher Education
149	26-ene-06	Universidade da Coruña	University/College/Academy/Higher Education
150	31-ene-06	Consejo Superior de Investigaciones Científicas	Research Institute/Research Organization/National Research Institute
151	10-feb-06	Universitat de Lleida	University/College/Academy/Higher Education
154	08-mar-06	Universitat de Vic	University/College/Academy/Higher Education
155	09-mar-06	Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas	Research Institute/Research Organization/National Research Institute
157	24-mar-06	El Departament d'Universitats, Recerca i Societat de la Informació de la Generalitat de Catalunya	Governmental Institution/Council
160	19-abr-06	Biblioteca de Catalunya	Library/Library Association/Information science organization
168	02-jun-06	Universitat Oberta de Catalunya	University/College/Academy/Higher Education

Tabla 1.- Primeras instituciones españolas firmantes de la Declaración de Berlín de 22/10/03 (fuente: <http://oa.mpg.de/openaccess-berlin/signatories.html>)

En octubre de 2007 se constituyó la Oficina Técnica Digital.CSIC que, con sede en la Unidad de Coordinación de Bibliotecas del CSIC en Madrid, sería la encargada de poner en práctica los criterios de diseño propuestos por el Comité Asesor para desarrollar el repositorio Digital.CSIC. La misión de la Oficina Técnica era diseñar una aplicación basada en DSpace que se adaptase a los requisitos técnicos expresados en la memoria inicial de proyecto, así como almacenar en la base de datos de dicha aplicación un número significativo de trabajos científicos elaborados por investigadores del CSIC con la idea de que fuera posible llevar a cabo su salida pública a Internet a comienzos de 2008.

La presentación oficial de Digital.CSIC tuvo lugar el 18 de enero de 2008 en rueda de prensa ante los medios de comunicación por parte del Presidente del CSIC. En el momento de su presentación, el repositorio institucional Digital.CSIC contaba con un total de 1200 documentos, resultado del trabajo inicial de archivo de trabajos por parte de la Oficina Técnica. No obstante, ya en la memoria inicial del proyecto Digital.CSIC se mencionaba como objetivo el de trabajar para que fuera el propio personal

científico del CSIC quien, una vez conocidos los objetivos del repositorio y sus procedimientos de trabajo, realizara el auto-archivo de sus publicaciones a texto completo en Digital.CSIC. Para alcanzar este objetivo era preciso hacer énfasis en la difusión del repositorio entre sus usuarios finales, los investigadores, y con este fin, nada más concluir la rueda de prensa institucional, se puso en marcha la primera fase del programa de difusión Digital.CSIC. A través de este programa se pretendía explicar en todos los centros e institutos que componen el CSIC (128 a día de hoy) el nuevo modelo de



Fig 1.- Distribución geográfica de los institutos del CSIC

comunicación científica basado en el acceso abierto, presentar el repositorio institucional Digital.CSIC como una de las líneas de promoción del acceso abierto en el CSIC y formar a personal científico y bibliotecas de la institución en el manejo de la nueva herramienta. Dado el elevado número de visitas y sesiones de formación que ello suponía, el programa de difusión se articuló en varias fases organizadas por cuatrimestres³.

La primera actividad interna de difusión del repositorio Digital.CSIC para personal científico del CSIC tuvo lugar el miércoles 23 de enero de 2008 en el Instituto de Matemáticas y Física Fundamental (IMAFF-CFMAC) de Madrid. A esa sesión piloto, a la que asistieron 50 investigadores del Centro de Física Miguel A. Catalán le siguieron decenas de presentaciones en institutos del Consejo en Madrid, Barcelona, Palma de Mallorca, Zaragoza, Sevilla, Granada y Valencia. A lo largo del primer cuatrimestre de 2008 se realizaron 53 presentaciones para un total de 74 institutos CSIC visitados.

Las sesiones de presentación para personal científico se complementaron a lo largo de esta primera fase con sesiones y cursos específicos para el personal de las bibliotecas de los centros visitados, con el fin de que los bibliotecarios de la Red del CSIC pudieran actuar como responsables de la gestión de Digital.CSIC en sus centros. Las bibliotecas asumirían también las tareas de soporte técnico básico en cuanto a la funcionalidad del repositorio de cara a sus investigadores.

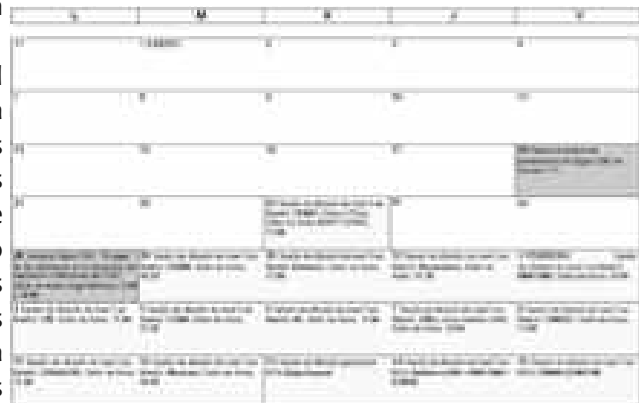


Fig 2.- Cronograma de la primera fase del programa de difusión Digital.CSIC

En el curso de la segunda fase del programa de difusión Digital.CSIC, que comenzó en el mes de junio de 2008, se visitó una serie de institutos CSIC que por necesidades de planificación quedaron fuera de la primera ronda de sesiones de presentación. En esta segunda fase se hizo especial énfasis en visitar todos los institutos de investigación de la Comunidad de Madrid, dado que en la convocatoria CSIC-CAM para la creación y consolidación de grupos de investigación CSIC en la Comunidad de Madrid de 20 de mayo de 2008 ya se mencionaba expresamente la obligatoriedad de depositar los trabajos en Digital.CSIC. Y es que para el progreso del acceso abierto es fundamental el compromiso con esta modalidad de comunicación científica de las instituciones públicas y las agencias financiadoras de la

investigación, que vienen adoptando cada vez en mayor número políticas específicas destinadas a promover la libre disponibilidad de los trabajos científicos en internet a través de los repositorios institucionales. En el CSIC no existe todavía a fecha de hoy una política de mandato institucional que obligue a depositar los trabajos de investigación en Digital.CSIC, pero con ocasión de la salida a público del repositorio, el Presidente del CSIC remitió una carta a todo el personal científico del CSIC informándole de la puesta en marcha de la iniciativa y animándole a tomar parte en ella incorporando su producción científica.



Fig 3.- Cartel anunciador de una sesión de presentación Digital.CSIC para personal científico

En una sesión de presentación Digital.CSIC para personal científico se explican las diferentes iniciativas promovidas por el CSIC a partir de la firma institucional de la Declaración de Berlín para promover el acceso abierto a nivel de todo el organismo. Se hace especial énfasis en el desarrollo y puesta en marcha de Digital.CSIC, pero se mencionan asimismo las iniciativas institucionales para ofrecer acceso abierto a las publicaciones del CSIC, así como el esfuerzo realizado desde la Unidad de Coordinación de Bibliotecas del CSIC para promover la publicación de trabajos en revistas de acceso abierto a través de la firma de acuerdos con las plataformas de publicación que permitan sufragar institucionalmente una parte importante de los costes que conlleva dicha

opción de publicación. En las presentaciones del repositorio se explica también en detalle el procedimiento de auto-archivo de trabajos científicos en Digital.CSIC, con objeto de lograr la implicación de los investigadores del CSIC en la tarea de auto-archivar parte de su producción científica en el repositorio. Finalmente, se hace además especial hincapié en los aspectos relacionados con los derechos de autor, de manera que quede clara la política de permisos de las editoriales científicas comerciales respecto al archivo de artículos en Digital.CSIC, y se conozcan al detalle los procedimientos necesarios para comprobar qué versión de cada trabajo puede archivarse en Digital.CSIC cumpliendo con la legislación sobre propiedad intelectual.

Como resultado de esta tarea de difusión de la iniciativa en los institutos del CSIC, se ha ido produciendo un crecimiento rápido y constante del número de trabajos disponibles en el repositorio desde su salida a público. Si en enero de 2008 Digital.CSIC contaba con 1200 registros almacenados, trece meses después de la publicación del recurso en Internet, el número de documentos archivados en Digital.CSIC se ha incrementado notablemente, alcanzando los 8500 registros. Las previsiones de desarrollo indican asimismo que para el mes de mayo se habrán superado holgadamente los 10.000 documentos archivados. Este importante crecimiento hace que Digital.CSIC esté convergiendo rápidamente a nivel de contenidos con los grandes repositorios españoles universitarios y de investigación (UPCommons, RECERCAT, RUA, e-Espacio UNED, e-Prints Complutense, por citar algunos de los principales), con la particularidad de que la trayectoria

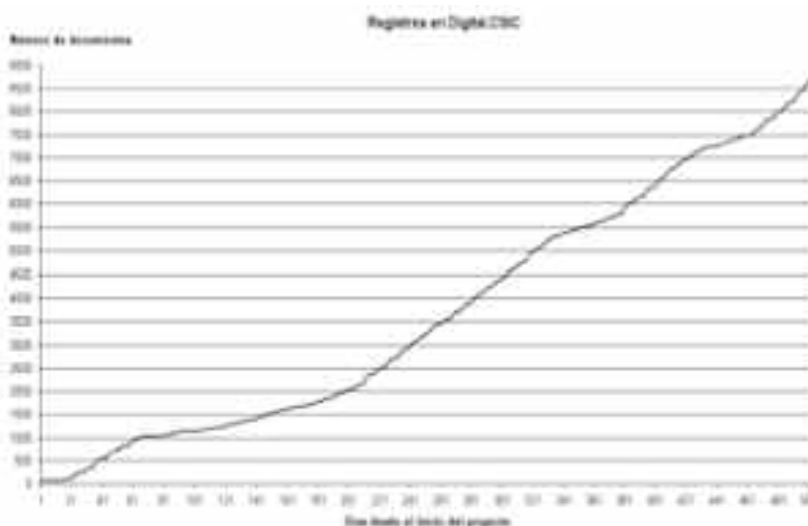


Fig 4.- Curva de crecimiento de contenidos en Digital.CSIC

temporal de éstos es generalmente bastante más dilatada que la de Digital.CSIC.

Una buena muestra de la evolución de Digital.CSIC la constituye su progresión en el Ranking Web de Repositorios Mundiales⁴, una iniciativa del Laboratorio de Cibermetría del Instituto de Estudios Documentales sobre Ciencia y Tecnología (IEDCYT) del CSIC que establece semestralmente una clasificación de los repositorios mundiales en base a parámetros objetivos tales como su tamaño, visibilidad en la web, nivel de riqueza de sus archivos y presencia en Google Scholar. En la tabla adjunta puede verse la evolución de Digital.CSIC en las sucesivas ediciones del Ranking, así como la creciente presencia de repositorios españoles entre el Top 300 del mundo.

Digital.CSIC en el Ranking Web de Repositorios Mundiales http://repositories.webometrics.info/about_rank_es.html		
Fecha de publicación	Posición de Digital.CSIC	Número de repositorios españoles
26/01/2009	74	9
24/07/2008	105	7
26/05/2008	230	5

Tabla 2.- Progresión de Digital.CSIC en el Ranking Web de Repositorios Mundiales

De manera simultánea a la difusión de Digital.CSIC entre el personal de los institutos del CSIC y al esfuerzo por lograr un crecimiento consolidado de contenidos, existe toda una línea paralela de desarrollo progresivo de la aplicación a nivel de software, con el fin de hacerla cada vez más completa en cuanto a funcionalidad y más amigable para sus usuarios investigadores. En este sentido, con la imprescindible colaboración de la Secretaría General Técnica de Informática del CSIC, unos meses después de la salida a público del repositorio se puso a punto el módulo de estadísticas de Digital.CSIC, con el que se pretende poner a punto servicios específicos de valor añadido, permitiendo a los autores conocer el nivel de visibilidad que alcanzan los trabajos que publican en Digital.CSIC. Además, entre los desarrollos recientes en esta línea de desarrollo de software cabe citar asimismo el acceso alternativo de usuario a través de autenticación CAS ante la intranet corporativa del CSIC, la asignación generalizada de permisos de edición que permiten a los autores actualizar y/o modificar la descripción de los trabajos que archiven en Digital.CSIC, y el desarrollo de la versión en inglés de la interfaz que facilita su consulta desde entornos internacionales.

Entre los objetivos que se están abordando actualmente cabe destacar el esfuerzo por automatizar la incorporación de registros al repositorio institucional, con la definición de estándares de transferencia automática de registros desde las bases de datos de publicaciones de las que disponen diferentes institutos del CSIC. Este desarrollo, que incrementará de manera cuantiosa el ritmo de crecimiento de contenidos archivados en Digital.CSIC y que simplificará también notablemente la colaboración de los investigadores para depositar sus trabajos, se complementa con la reciente puesta a punto de un formulario de comunicación Digital.CSIC para que los autores puedan remitir sus publicaciones recientes a sus bibliotecas para que sean archivadas en el repositorio en la modalidad de archivo delegado, esto es, a través de representantes autorizados de los propios investigadores.

Como parte de esta línea de desarrollo se pretende también lograr una creciente integración de Digital.CSIC con otros recursos corporativos del CSIC y de su Red de Bibliotecas que pueda dar lugar a sinergias beneficiosas para su desarrollo. Entre los más recientes avances destacados en este campo puede citarse la instalación del servicio de enlaces SFX, así como la futura integración del repositorio Digital.CSIC con el metabuscador MetaLib de la Red de Bibliotecas del CSIC.

Finalmente, una vez consolidada una plataforma sólida de trabajo a nivel de contenido y de uso, se plantea el objetivo de establecer una política de incremento de la visibilidad de Digital.CSIC con su presentación en foros nacionales e internacionales sobre acceso abierto, así como de fomentar la integración de Digital.CSIC en las redes europeas e internacionales de repositorios científicos. En relación con estos últimos objetivos ya se han dado algunos pasos iniciales como la publicación en

revistas internacionales de artículos de difusión de la iniciativa Digital.CSIC⁵ y la presentación del repositorio en foros de ámbito más generalista como la pasada Feria del Libro de Madrid (con la colaboración del Área de Cultura Científica del CSIC) u otras ferias de carácter científico como el Euroscience Open Forum (ESOF 2008) de Barcelona.

Institutos CSIC con mayor presencia de trabajos en Digital.CSIC (15/02/2009)	
Instituto de Análisis Económico (IAE – Barcelona)	331
Instituto de Geología Económica (IGE – Madrid)	331
Museo Nacional de Ciencias Naturales (MNCN – Madrid)	318
Instituto de Filosofía (IFS – Madrid)	293
Escuela Experimental de Aula Dei (EEAD – Zaragoza)	286
Instituto de Ciencias de la Tierra "Jaume Almera" (ICTJA – Barcelona)	286
Centro de Biología Molecular "Severo Ochoa" (CBMSO – Madrid)	238
Instituto de Estructura de la Materia (IEM-CFMAC – Madrid)	233
Instituto de Física Interdisciplinar y Sistemas Complejos (IFISC – Palma)	229
Instituto de Física Corpuscular (IFIC – Valencia)	225
Instituto de Acústica (IA – Madrid)	224
Instituto de Políticas y Bienes Públicos (IPP – Madrid)	224
Instituto de Historia (IH – Madrid)	213
Centro Nacional de Biotecnología (CNB – Madrid)	173
Instituto "Eduardo Torroja" de Ciencias de la Construcción (ETCC – Madrid)	157

Fig 5.- Institutos CSIC con mayor número de contenidos en Digital.CSIC

En relación con la participación de Digital.CSIC en plataformas comunes de desarrollo del acceso abierto, es de destacar que desde el inicio del proyecto se ha venido produciendo una estrecha colaboración con la plataforma e-ciencia de la Comunidad de Madrid⁶ en la definición y aplicación de estándares de trabajo cooperativos para garantizar que el desarrollo del repositorio institucional del CSIC se armonizase con otros proyectos en la misma línea promovidos por las universidades públicas de Madrid. Además de esta participación directa en la plataforma del Consorcio Madroño, Digital.CSIC ha asumido también desde su inicio las directrices técnicas establecidas por el Proyecto DRIVER (Digital Repository Infrastructure Vision for European Research) de la Unión Europea, cuyo objetivo prioritario es la creación de una infraestructura paneuropea de repositorios digitales que sea a la vez cohesiva, robusta y flexible⁷.

Por lo que respecta a plataformas integradoras de recursos en acceso abierto, Digital.CSIC está incluido en los más importantes directorios internacionales de repositorios como ROAR⁸ (Registry of Open Access Repositories, University of Southampton, UK) u OpenDOAR⁹ (Directory of Open Access Repositories, University of Nottingham, UK). El repositorio institucional del CSIC participa asimismo como proveedor de datos en Recolecta¹⁰, una iniciativa conjunta de la Fundación Española para la Ciencia y la Tecnología (FECYT) y la Red de Bibliotecas Universitarias (REBIUN) para desarrollar un proveedor de contenidos unificado para todos los repositorios de ámbito nacional. Finalmente, los contenidos de Digital.CSIC son periódicamente recolectados de modo sistemático por los grandes agregadores de ámbito internacional como OAIster¹¹ (University of Michigan, EE.UU.) o Scientific Commons¹² (Universidad de St-Gallen, Suiza).

Resta aún mucho camino por recorrer hasta hacer de Digital.CSIC un recurso de información verdaderamente útil para los investigadores del CSIC y para la institución en su conjunto. Son muchas las funcionalidades que permanecen todavía a nivel de diseño que enriquecerán notablemente el servicio que Digital.CSIC puede ofrecer a la institución y a su personal. No obstante, la trayectoria recorrida hasta el momento ha sido considerablemente fructífera y es de esperar que si el desarrollo de la aplicación continúa al ritmo actual, el CSIC podrá contar pronto con un repositorio institucional al nivel de los de otras instituciones internacionales de similar relevancia como el CNRS¹³ francés o la Max Planck Gesellschaft¹⁴ en Alemania.

Bibliografía

¹ Declaración de Berlín por el acceso abierto al conocimiento en ciencias y en humanidades, http://www.zim.mpg.de/openaccess-berlin/berlin_declaration.pdf.

² Giner Durán, Clara et al, *El Repositorio Institucional Digital.CSIC: la materialización de la estrategia Open Access del Consejo Superior de Investigaciones Científicas*. TECNIMAP 2007, Nov 27-30th, 2007, Gijón, Spain, <http://hdl.handle.net/10261/2222>.

³ Castro, Pablo de, *Digital.CSIC: un repositorio institucional para el Consejo Superior de Investigaciones Científicas*. 2^{as} Jornadas OS-Repositorios: “La ecología de los repositorios institucionales” (Gijón, 14 diciembre 2007), http://www.csic.es/cbic/servicios/docs/PPT_Gijon.pdf.

⁴ Ranking Web de Repositorios Mundiales, http://repositories.webometrics.info/about_rank_es.html

⁵ Ponsati, Agnès y Castro, Pablo de, *Working for a new scholarly communication model at the Spanish National Research Council (CSIC): the institutional Repository Digital.CSIC*. Research Information 37: 16-17 (2008), <http://digital.csic.es/handle/10261/6747>.

⁶ Proyecto e-ciencia: Buscador de Archivos Abiertos (Comunidad de Madrid), <http://www.madrimasd.org/informacionidi/e-ciencia/proyecto/default.asp>.

⁷ DRIVER: Digital Repository Infrastructure Vision for European Research, <http://www.driver-repository.eu/>

⁸ ROAR: Registry of Open Access Repositories, <http://roar.eprints.org/>

⁹ OpenDOAR: Directory of Open Access Repositories, <http://www.opendoar.org/>

¹⁰ RECOLECTA: Recolector de ciencia abierta, <http://www.recolecta.net/buscador/index.jsp>

¹¹ OAIster, <http://www.oaister.org/>

¹² Scientific Commons, <http://www.scientificcommons.org/>

¹³ Hal: Hyperarticles en Ligne (CCSD-CNRS), <http://hal.archives-ouvertes.fr/>

¹⁴ Max Planck Society eDoc Server, <http://edoc.mpg.de/>