



# Proyecto "Madroño e-Ciencia"

## Memoria descriptiva

<b>Título del Proyecto</b>	Plataforma "Madroño e-Ciencia"
<b>Fecha de inicio</b>	2006
<b>Fecha de finalización</b>	2008
<b>Institución líder del proyecto</b>	Consortio Madroño
<b>Dirección de la gestión</b>	Comisión Técnica del Consortio Madroño
<b>Autora</b>	Alicia López Medina

**Agradecimientos:** la elaboración de este proyecto debe mucha de su inspiración a otros proyectos en él mencionados, especialmente el proyecto DARE, en Holanda, el proyecto ARROW, en Australia y el proyecto ePrints-UK, en el Reino Unido. También a las ideas aportadas por el profesor Jesús M. González Barahona, que lo enriquecieron, y a las siempre oportunas correcciones y propuestas de Luis Zorita, responsable del servicio informático de la Biblioteca de la UNED.

## Índice

1. Sumario
2. Marco del Proyecto
3. Análisis del entorno global
  - Los contenidos digitales
  - El acceso libre a la información científica
  - Software de código abierto
  - Estándares y protocolos abiertos
4. Situación en España
5. Objetivos
6. Organización
7. Modelo del proyecto
8. Funcionamiento
9. Nivel de datos: red de repositorios Madroño
  - Arquitectura

- Software
- Contenidos
- Aspectos legales

10. Nivel de servicios: portal "Madroño e-Ciencia"

11. Evaluación

12. Difusión

13. Colaboración/Financiación

14. Recursos necesarios

15. Planificación

## **1. SUMARIO**

El proyecto "Madroño e-Ciencia" es una iniciativa conjunta del Consorcio de Universidades Madroño cuyo objetivo es crear una infraestructura y unos servicios avanzados para recoger, acceder, preservar y distribuir en soporte digital la producción intelectual generada por la comunidad investigadora y académica de la Comunidad de Madrid.

El fin principal e inmediato que se propone el proyecto es mejorar el acceso y la visibilidad – y por tanto su impacto - de los resultados de la investigación científica de la Comunidad de Madrid.

El proyecto "Madroño e-Ciencia" es un proyecto escalable cuyo puesta en marcha y desarrollo cubrirá el periodo 2006-2008 y tendrá como resultado tres iniciativas relacionadas:

La creación de la **red Madroño de repositorios abiertos**, formada por los repositorios implementados en las Universidades que forman parte del Consorcio Madroño. Se crearán mecanismos de acomodo para acoger los contenidos de investigadores de la Comunidad de Madrid que no tengan acceso a repositorios institucionales.

La creación del **Portal "Madroño e-Ciencia"** que permitirá el acceso, libre y gratuito a través de Internet, desde un único punto de búsqueda común, a los contenidos almacenados en los distintos repositorios de la red Madroño y que ofrecerá una serie de servicios de valor añadido contruidos sobre la base de esos contenidos.

La creación de un **sitio web "Madroño e-Ciencia"** desde el que se difundirá el proyecto.

Los estándares utilizados son robustos en relación con futuros avances y están estrechamente alineados con los desarrollos internacionales, permitiendo un intercambio de información altamente efectivo tanto a nivel nacional como internacional.

## **2. MARCO DEL PROYECTO**

El Proyecto "Madroño e-Ciencia" constituye una de las líneas de actividad propuestas por el Consorcio Madroño en el Programa de Cooperación Bibliotecaria correspondiente al Contrato Programa entre la Administración de la Comunidad de Madrid y el Consorcio Madroño para la regulación del marco de cooperación en el Sistema Regional de Investigación Científica e Innovación Tecnológica IV PRICIT (2005-2008).

Parte, además, de un acuerdo previo entre las Universidades del Consorcio Madroño, constituyendo un objetivo común definido en sus respectivos Planes Estratégicos.

El Proyecto, finalmente, se enmarca en el contexto global de desarrollos similares llevados a cabo a nivel internacional en el ámbito de la publicación y comunicación científica.

## **3. ANÁLISIS DEL ENTORNO GLOBAL**

### **Los contenidos digitales son la clave**

La cuestión se plantea con claridad en la declaración del Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica, 2004-2007<sup>1</sup>: "no se entendería el desarrollo de la Sociedad de la Información si, además de infraestructuras, software y hardware, no se dotara a los ciudadanos de contenidos digitales de calidad, así como de sistemas de fácil acceso y de alta eficacia".

En el "Libro Blanco de la E-Ciencia"<sup>2</sup>, publicado por la FECYT, se reconoce la importancia clave de la información para el avance científico y se advierte que para que la comunidad científica e investigadora española pueda hacer frente a

los grandes retos en las diversas áreas de aplicación de la e-ciencia, se requerirá la existencia y disponibilidad de bancos de datos muy voluminosos y recursos de información accesibles a través de las grandes redes globales de comunicación.

Europa reconocía la importancia clave de los contenidos digitales de calidad para el avance y competitividad de la ciencia europea, el desarrollo de la economía y la mejora de la vida de los ciudadanos en su programa "Econtent 2001-2004", adoptado por decisión del Parlamento Europeo y cuyo objetivo era fomentar el desarrollo y uso de contenidos digitales europeos en internet y la diversidad lingüística en la sociedad de la información.

El reconocimiento de su importancia continúa en la decisión del Parlamento Europeo y del Consejo de 9 de marzo de 2005 que establece para el periodo 2005-2008 un programa comunitario, "EContentplus"<sup>3</sup> – al que destina 149 millones de euros - para "incrementar las posibilidades de acceso, utilización y explotación de los contenidos digitales en Europa, facilitando su creación y difusión, aumentando la calidad del contenido con metadatos bien definidos y reforzando la cooperación entre los actores implicados en los contenidos digitales".

El volumen de contenidos digitales de calidad, su fácil accesibilidad a través de internet y la garantía de su preservación a largo plazo son factores críticos para determinar la capacidad científica y tecnológica en una economía globalizada.

<sup>1</sup> [http://www.madrimasd.org/queesmadrimasd/pricit/PlanNet/\\_documentos/41/PlanIDI/VolII.pdf](http://www.madrimasd.org/queesmadrimasd/pricit/PlanNet/_documentos/41/PlanIDI/VolII.pdf)

<sup>2</sup> <http://www.fecyt.es/documentos/e-Ciencia.pdf>

<sup>3</sup> [http://europa.eu.int/information\\_society/activities/econtentplus/index\\_en.htm](http://europa.eu.int/information_society/activities/econtentplus/index_en.htm)

## **Acceso libre a los contenidos científicos**

El mundo de la comunicación científica está inmerso en un proceso de cambio profundo como consecuencia de las nuevas oportunidades creadas por los desarrollos de las tecnologías de la información y la comunicación. Esta revolución tecnológica permite una mayor difusión de los resultados de la investigación, posibilitando así mejor y más fácil acceso a los mismos, cualquiera que sea su localización geográfica.

Sin embargo, a pesar de las posibilidades que ofrece la tecnología, el hecho es que en este momento el acceso a las publicaciones científicas está siendo sometido a barreras y restricciones impuestas por las grandes editoriales que monopolizan los cauces de distribución de la literatura científica. A causa de este monopolio, las suscripciones a las revistas científicas son cada día más caras, hasta el punto de que muchos organismos no pueden seguir pagando las suscripciones. Y aquí es cuando se produce la siguiente paradoja: los científicos que pertenecen a estas organizaciones, en la mayoría de los casos

financiadas por fondos públicos, no pueden acceder a sus propios trabajos, siendo ellos precisamente los que producen y ofrecen gratuitamente la materia prima de la que se nutren las revistas.

Estas barreras al acceso a la información científica cobran una especial relevancia en el seno de la sociedad del conocimiento, en la que la difusión y el acceso a ese conocimiento, y no sólo su generación, se convierten en un factor crítico para determinar la capacidad científica y tecnológica en una economía globalizada.

En este contexto nace el movimiento "Open Access"<sup>4</sup> cuyo objetivo es poner la literatura científica y académica en Internet y hacerla accesible sin restricciones. Este movimiento cuestiona el monopolio que las grandes editoriales ejercen sobre la distribución de la información científica. Sin embargo, el propósito fundamental del acceso abierto no es castigar o socavar las revistas científicas caras, sino ofrecer una alternativa accesible y aprovechar al máximo todas las ventajas de las nuevas tecnologías -internet-. Por lo tanto, para los propios investigadores, la motivación primordial del acceso abierto no es solucionar el problema del precio de las revistas sino conseguir un acceso mayor y más rápido a la información como lectores y una mayor audiencia e impacto como autores. Para nuestras Universidades, el acceso abierto incrementa la visibilidad de sus investigadores y de la propia institución, reduce sus gastos en revistas y hace avanzar su misión de compartir conocimiento.

El movimiento por el acceso abierto a la literatura científica ha encontrado eco en las políticas científicas de muchos países y en organismos internacionales.

En el año 2003, en el marco del "Sommet Mondial sur la Société de l'Information" (SMSI) tuvo lugar en Ginebra una conferencia "Libre Accès: Vers une Science Libre" con el objetivo de lograr la toma de conciencia de los gobiernos a nivel mundial sobre la necesidad del libre acceso a la información. El resultado de esta cumbre fue una Declaración de Principios en la que los signatarios se declaraban firmemente a favor del Acceso Abierto en todas sus formas como factor esencial para la creación efectiva de la sociedad de la información.

<sup>4</sup> <http://www.soros.org/openaccess/index.shtml>

En octubre del mismo año, la Comisión Europea suscribió la Declaración de Berlín<sup>5</sup>, la cual avalaba el acceso abierto al conocimiento. Más recientemente, el 15 de Junio de 2004, la Comisión Europea decidió hacer un estudio sobre la publicación científica en Europa. El libre acceso está en el centro de este estudio. Estamos esperando los resultados.

El 29/30 de enero de 2004, los ministros de ciencia y tecnología de 34 países, especialmente de la Unión Europea, acordaron que el acceso abierto "contribuye de manera decisiva al progreso de la investigación científica y a la

innovación". La OCDE se compromete a adoptar una serie de líneas directrices en este sentido.

Países como Escocia, Finlandia o Australia han asumido en sus políticas científicas los principios del acceso libre a la literatura científica como mejor medio de servir a los intereses de sus naciones, recomendando a sus investigadores el libre acceso como forma de difusión de sus resultados.

En el año 2004 el Parlamento británico solicitó un informe sobre la publicación científica y el libre acceso cuyo resultado fue el "Scientific Publications: Free for all?", Tenth Report of Session 2003-04/House of Commons Science and Technology Committee, vol I<sup>6</sup>, en el cual se recomendaba que todas las instituciones de educación superior británicas establecieran repositorios institucionales en los cuales su producción científica estuviera accesible en línea libre de cualquier restricción.

El movimiento por el acceso libre a la información científica se materializa fundamentalmente a través de dos tipos de estrategias: la publicación en revistas electrónicas y la creación de archivos o repositorios de contenidos digitales (de carácter institucional o disciplinario); ambas estrategias son complementarias y no competitivas. Ambas tienen como objetivo mejorar y ampliar el acceso a la literatura científica y ambas ofrecen potencial para una mayor eficiencia de la publicación científica.

Los repositorios institucionales son una tendencia emergente entre las Instituciones de Educación Superior. Se fundamentan sobre la práctica habitual entre los investigadores de colocar sus trabajos en línea, en donde entran a formar parte de un sistema global de contenido distribuido. Los repositorios institucionales representan una respuesta estratégica a las oportunidades ofrecidas por un entorno digital interconectado y a los signos de crisis en algunos de sus procesos que está produciéndose en el sistema de publicación tradicional en revistas científicas.

<sup>5</sup> [http://www.zim.mpg.de/openaccess-berlin/berlin\\_declaration.pdf](http://www.zim.mpg.de/openaccess-berlin/berlin_declaration.pdf)

<sup>6</sup> <http://www.publications.parliament.uk/pa/cm200304/cmselect/cmsctech/399/39902.htm>

Un repositorio institucional concretamente es un conjunto de servicios web centralizados, creado para organizar, gestionar, preservar y ofrecer acceso libre a la producción científica, académica o de cualquier otra naturaleza cultural, en soporte digital, generada por los miembros de una institución. Las principales características de un repositorio institucional son su naturaleza institucional, su carácter científico, acumulativo y perpetuo, así como abierto e interoperable con otros sistemas.

Cada vez un mayor número de Universidades e Instituciones y Organismos Científicos están concretando las políticas de acceso abierto mediante la creación de repositorios digitales institucionales que recogen y dan acceso libre a los resultados de su actividad científica tal y como se pone de manifiesto en

el Encuentro Berlin 3<sup>7</sup> "Progress in Implementing the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities" que tuvo lugar en Southampton (UK) en marzo de 2005, al que asistieron delegados de las Instituciones que habían firmado la Declaración de Berlín 2003 y también delegados de instituciones que aún no lo habían hecho. Entre las naciones representadas estaba la India<sup>8</sup>, Japón<sup>9</sup>, EE.UU<sup>10</sup>, Canadá<sup>11</sup>, UK<sup>12</sup> y otros muchos países. Cada Institución informó de sus progresos en la implementación o la provisión del acceso abierto a sus contenidos científicos. Entre las Instituciones más avanzadas están las de ámbito nacional como el CNRS en Francia y en Alemania la Max-Planck Gesellschaft y, a nivel internacional, instituciones como el CERN en Suiza. El CERN tiene una política de acceso abierto que hasta el momento permite el acceso libre al 60% de la investigación producida en sus laboratorios, pero su compromiso es llegar al 100%<sup>13</sup>. Pero también muchas Universidades individuales y consorcios de Universidades están haciendo progresos espectaculares: [La Universidade Do Minho](#) en Portugal<sup>14</sup>, por ejemplo, ha diseñado una determinada política para autoarchivar su producción científica, como también lo ha hecho el Consorcio [SURF/DARE](#) de las 12 principales Universidades holandesas. Las Universidades Escandinavas<sup>15</sup> han estado también muy activas, así como un gran número de Universidades e Instituciones científicas en Italia<sup>16</sup>, Alemania<sup>17</sup> y Francia<sup>18</sup>. Australia<sup>19</sup> no envió delegados esta vez, pero tiene muchas Instituciones muy activas en relación con el desarrollo de políticas y actuaciones de acceso libre a los resultados de su investigación.

Implementar un repositorio institucional requiere un considerable esfuerzo de planificación y de compromiso. Entonces, ¿por qué tantas Universidades están creándolos?. Al contrario de lo que sucede con el sistema de publicación en revistas científicas, el cual limita la lectura y disponibilidad de los materiales así como el formato de la investigación, un repositorio institucional ofrece la difusión más amplia posible de toda la panoplia completa de la producción intelectual digital generada en una institución. Un repositorio institucional puede incrementar la visibilidad y el prestigio de la institución y demostrar su valor para las fuentes de financiación y subvenciones. Aumenta la visibilidad de sus investigadores ampliando la difusión y el uso de sus trabajos. Un repositorio institucional puede estimular la innovación, facilitar un análisis cualitativo del trabajo de sus miembros, apoyar las tareas de enseñanza/aprendizaje, servir de sistema de registro de ideas, ofrecer un catálogo del capital intelectual de la institución. Los repositorios institucionales reconocen y dan cabida a los objetos digitales científicos no incluidos en los canales tradicionales de publicación. Finalmente, los repositorios institucionales pueden mejorar la comunicación científica y hacer avanzar la investigación científica permitiendo a los usuarios localizar y recuperar información relevante más rápida y fácilmente.

<sup>7</sup> <http://www.eprints.org/berlin3/program.html>),

<sup>8</sup> <http://archives.eprints.org/eprints.php?action=home&country=in>

9 <http://archives.eprints.org/eprints.php?action=home&country=jp>  
10 <http://archives.eprints.org/eprints.php?page=all&country=us>  
11 <http://archives.eprints.org/eprints.php?page=all&country=ca>  
12 <http://archives.eprints.org/eprints.php?action=home&country=uk>  
13 <http://info.web.cern.ch/Press/PressReleases/Releases2005/PR04.05E.html#CERN>  
14 <http://archives.eprints.org/eprints.php?action=home&country=pt>  
15 <http://archives.eprints.org/eprints.php?page=all&country=se>  
16 <http://archives.eprints.org/eprints.php?action=home&country=it>  
17 <http://archives.eprints.org/eprints.php?page=all&country=de>  
18 <http://archives.eprints.org/eprints.php?page=all&country=fr>  
19 <http://archives.eprints.org/eprints.php?page=all&country=au>

El rápido crecimiento del número de repositorios digitales de acceso abierto a la información científica ha obligado incluso a los propios editores comerciales a tener en cuenta el fenómeno del acceso abierto. Reed Elsevier anuncia en mayo de 2004 su acuerdo para permitir que los autores publiquen los "post-prints" de sus artículos en libre acceso. Aunque las restricciones de forma que impone pueden no coincidir exactamente con las definiciones públicas de "acceso abierto" estamos de acuerdo con Peter Suber, en SPARC, en que, en cualquier caso, no hay duda de que esta iniciativa supone un avance significativo para el acceso abierto. Springer Verlag, número 2 mundial de la edición científica después de su fusión con Kluwer, ha anunciado en julio de 2004 su modelo de publicación denominado "choice option" que consiste en permitir al autor elegir entre publicar según el sistema tradicional o el acceso abierto, eso sí, con un coste para el autor, en el segundo caso.

## **La tecnología del conocimiento compartido: software libre o de código abierto**

El movimiento denominado "*Free Open Source Software*" (FOSS) o "*Open Source Software*" (OSS), funciona sobre dos principios: la disponibilidad libre del software y un sistema de trabajo en cooperación para su desarrollo.

Desde el punto de vista del software, "libre" quiere decir, en primer lugar, gratuito, o más concretamente, no sometido a costes de licencia. "Libre" se opone también a "propietario" en el sentido de que el código fuente del programa es público y está abierto a cualquiera que quiera aportar algo a su desarrollo o modificación sin más que indicar su origen. La licencia más generalizada bajo la que se distribuye el software libre es la GNU General Public License.

La cuestión de la comunidad y la colaboración es central para el modelo de desarrollo del software libre. La comunidad del software libre trabaja en

equipo, ofreciéndose consejo, asesoramiento, respuestas, soluciones, intercambiando conocimientos entre personas distribuidas por todo el mundo a través de newsgroup, wikis, blogs, chats, etc... La actividad está en plena ebullición y parece una dirección de desarrollo que se consolida como definitiva.

El último informe publicado por el Observatorio Tecnológico Digicult -enero de 2005<sup>20</sup> -, financiado por el programa "Information Society Technologies" (IST), en el marco del V Programa Marco de la Unión Europea que investiga los desarrollos tecnológicos y recomienda aquellos que pueden usar las instituciones de memoria (bibliotecas, archivos, museos, etc.) para gestionar mejor y hacer más accesibles sus contenidos, ha identificado como una de las tecnologías emergentes clave el software libre o de código abierto.

Esta comunidad ha desarrollado una rica y variada gama de productos alternativos a las soluciones comerciales. Son productos realistas, robustos y, desde el punto de vista funcional, completamente alternativos al software comercial. Los programas de software abierto alivian la opacidad en el diseño del producto que es un obstáculo para el intercambio a través del espacio y el tiempo de objetos digitales creados con productos comerciales cerrados.

La creación de archivos o repositorios digitales se ha visto estimulada en gran parte por la cantidad de programas disponibles de gestión de documentos digitales en código abierto. Parece lógico, en el entorno del movimiento por el acceso libre a la información científica, el apoyo a este tipo de software, también "libre": la guía publicada por el "Open Society Institute" en 2004<sup>21</sup>, nos da una idea del número y diversidad de programas disponibles, cuyas notas comunes para ser incluidos en esta guía son estar disponibles gratuitamente como software libre, ser conformes con la última versión del protocolo OAI-PMH y tener una versión actualizada y disponible públicamente.

Pero la utilización de software libre o de código abierto no sólo se justifica por razones diríamos "ideológicas" o teóricas (en el marco del derecho a acceder y usar la información libremente y sin restricciones), sino que tiene claras ventajas estratégicas desde un punto de vista más práctico.

La dependencia de productos de software propietario no ofrece a la mayoría de las instituciones modelos de costes razonables u opciones sostenibles a largo plazo para los materiales digitales que crean con esos productos. La gratuidad del software libre (aunque no el coste cero) permite la contención de costes y la escalabilidad de las soluciones. Como contrapartida, su utilización va a requerir un fuerte compromiso de actualización a las personas técnicamente implicadas en estos proyectos.

<sup>20</sup> DigiCult Technology Watch Report, january 2005: Core technologies for the cultural and scientific heritage sector  
<http://www.digicult.info/downloads/TWR3-lowres.pdf>

<sup>21</sup> Open Society Institute "Guide to Institutional Repository Software", tercera edición, agosto de 2004:

<http://www.soros.org/openaccess/software/>

Aunque la apuesta por un programa de código abierto conlleva el riesgo de implementar un producto que no está cerrado, el compromiso de la comunidad del software libre con su desarrollo pone a nuestra disposición todo un equipo de expertos trabajando en todo el mundo dispuestos a ayudarnos, trabajando juntos por la mejora del producto sin que existan intereses comerciales en esos desarrollos sino sólo las necesidades de sus usuarios, lo cual convierte la apuesta en una fortaleza frente a productos comerciales cerrados cuyos desarrollos, mejoras y actualizaciones vienen condicionadas por sus intereses de negocio y que además cuestan muy caras y que, en muchas ocasiones, acaban de forma precipitada. De hecho, está apareciendo la figura de la empresa que conoce en profundidad software abierto y asesora sobre posibles soluciones basadas en él.

El compromiso de la comunidad del software libre con el uso de estándares y protocolos abiertos nos garantiza que no estamos invirtiendo en un producto que se quedará anticuado enseguida, que será incompatible con otros sistemas o que no nos permitirá cambiar de programa o reutilizar nuestros contenidos.

La posibilidad de acceder al código significa, por último, mayor facilidad a la hora de desarrollar aplicaciones locales.

### **Protocolos y estándares para la interoperabilidad en un entorno distribuido: "Open Access Initiative" (OAI)**

Las filosofías subyacentes al software de código abierto tienen claras asociaciones con el concepto de estándares y protocolos abiertos, los cuales son vitales para el éxito del intercambio, reutilización y preservación de documentos y datos. Estándares abiertos son aquellos que en virtud de su transparencia y aceptación, ofrecen protección contra la obsolescencia y la inaccesibilidad. Ejemplos de estándares abiertos para formatos son xml, html y adobe pdf. Representan una forma segura de almacenar contenidos para recuperaciones futuras, dado que todos son entendidos, documentados y publicados universalmente. No están asociados a costes de licencias y el uso y distribución de estos formatos es gratis.

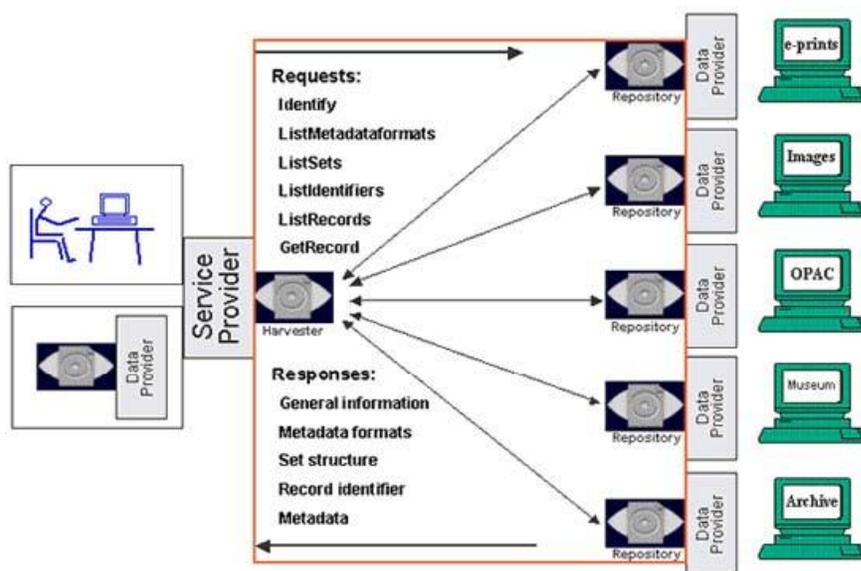
El uso de protocolos abiertos es también una condición para asegurar la interoperabilidad y por tanto el intercambio y transferencia de información. En este sentido, junto con el acceso sin restricciones a la información científica promovida por el movimiento de acceso abierto, se desarrolla otra iniciativa, la "Open Archives Initiative"<sup>22</sup> cuyo objetivo es desarrollar estándares y protocolos que permitan la interoperabilidad entre archivos digitales heterogéneos, de manera que se facilite la difusión eficiente de sus contenidos. La razón para la creación de esta iniciativa es la firme convicción de que la interoperabilidad entre archivos es la clave para incrementar su impacto y establecerlos como alternativas viables al modelo tradicional de comunicación científica.

<sup>22</sup> <http://www.openarchives.org/>

La creación de repositorios institucionales de acceso libre a la información

científica es un paso esencial para mejorar la difusión y el uso de los resultados científicos, pero si estos repositorios no son susceptibles de ser buscados de forma integrada, es decir, si el usuario tiene que encontrar el repositorio para buscar uno a uno, las ventajas de su acceso libre a través de Internet se ven ciertamente limitadas. El protocolo creado en el seno de la iniciativa OAI viene precisamente a solventar este problema, puesto que su implementación hace que los contenidos de todos los repositorios puedan ser buscados de forma integrada desde un único punto de búsqueda. Estos puntos pueden ser desde Google hasta los denominados "proveedores de servicios" según los términos del protocolo: servicios desarrollados por organizaciones que permiten buscar en varios repositorios como si el usuario estuviera buscando en una única colección.

Este protocolo es el protocolo OAI-PMH (*Open Archives Initiative - Protocol for Metadata Harvesting*), cuya arquitectura se basa en dos niveles o funciones: los proveedores de datos y los proveedores de servicios. Los primeros son los archivos o repositorios que proporcionan la información y los segundos son los recolectores o proveedores de servicios que toman los datos, con el objetivo de incorporarles algún valor añadido y presentarlos a los usuarios finales:



**"Proveedores de datos" (data provider):** Es una organización que ofrece acceso a los metadatos que describen sus colecciones a través del protocolo OAI-PMH. El término describe una interfaz -conforme al OAI- de acceso a una base de datos que contiene metadatos sobre objetos digitales. La interfaz es accesible vía una conexión http. Debe ser capaz de responder correctamente a las peticiones que se definen en las especificaciones del protocolo. Un repositorio abierto, en el contexto del protocolo OAI es un servidor, con conexión a Internet que es capaz de responder a las peticiones de un proveedor de servicios.

**“Proveedores de servicios” (service providers):** Es una organización que solicita metadatos a un proveedor de metadatos y los almacena en su propio servidor. Con estos metadatos puede ofrecer servicios añadidos, como “browsing” y “search”, que no están definidos en las especificaciones del protocolo. La ventaja del proveedor de servicios se basa en el hecho de que puede recolectar metadatos de varios proveedores de datos y combinar los resultados, lo cual le permite hacer que colecciones procedentes de sitios dispares sean presentadas al usuario final de forma integrada, como si hubiera buscado en una única colección. El proveedor de servicios puede ser, a su vez, proveedor de datos para otros proveedores de servicios.

Mediante la implementación de este protocolo, cada uno de los repositorios se convierte en un nodo de una red internacional científica que materializa la visión última que subyace bajo todas estas iniciativas, la creación de una web científica universal de acceso libre, cuyo contenido serán las colecciones de información científica depositadas en archivos abiertos (en términos de interoperabilidad) es decir, un archivo global virtual de la información científica.

En la actualidad muchas Universidades están llevando a cabo iniciativas conjuntas, soportadas sobre el protocolo OAI-PMH, para desarrollar portales de acceso centralizado a su producción científica almacenada en soporte digital en sus repositorios respectivos.

Desde enero del 2004 está disponible el portal de acceso DAREnet<sup>23</sup> a la red de repositorios institucionales de las Universidades holandesas. DAREnet es un portal de acceso, proveedor de servicios, que permite el acceso digital a la producción científica y académica holandesa. En el momento de su puesta en funcionamiento daba acceso a más de 15.000 archivos digitales pertenecientes a 15 instituciones (Universidades e Instituciones asociadas). En Canadá la “Canadian Association of Research Libraries” (CARL) ha puesto en marcha el proyecto “**Institutional Repositories Pilot Project Harvester**”<sup>24</sup>, que constituye la interfaz de búsqueda única de la red de repositorios institucionales implementados en las 12 bibliotecas canadienses participantes. En Australia el “**ARROW National Discovery Service**”<sup>25</sup>, permite la búsqueda desde un único punto de acceso a todo el corpus científico generado por las Universidades australianas depositado en sus respectivos repositorios institucionales, la Biblioteca Nacional de Australia y otras agencias de investigación. En el Reino Unido, el proyecto **ePrints UK**<sup>26</sup> tiene el propósito de desarrollar una serie de servicios a nivel nacional a través de los cuales la comunidad universitaria pueda acceder a la producción científica colectiva disponible en los repositorios abiertos, particularmente aquellos creados en las Universidades británicas. Desde finales mayo de 2004 el servicio está disponible en pruebas<sup>27</sup>.

<sup>23</sup> [http://www.surf.nl/en/themas/index2.php?oid=7&inhoud\\_oid=15](http://www.surf.nl/en/themas/index2.php?oid=7&inhoud_oid=15)

<sup>24</sup> <http://carl-abrc-oai.lib.sfu.ca/index.php>

<sup>25</sup> <http://search.arrow.edu.au/apps/ArrowUI/>

<sup>26</sup> <http://www.rdn.ac.uk/projects/eprints-uk/>

<sup>27</sup> <http://eprints-uk.rdn.ac.uk/>

En junio de este año estaba previsto el inicio del Proyecto "DART-Europe: Digital Access to Research Theses-Europe"<sup>28</sup>, cuyo funcionamiento y estructura se basa en el protocolo OAI-PMH y cuyo objetivo es explorar la creación de un modelo europeo para el depósito, la búsqueda, uso y preservación a largo plazo de las tesis doctorales europeas en el contexto del open access. Todas las instituciones y países con repositorios conformes al protocolo OAI podrán exponer sus tesis a través del portal "Dart Europe E-thesis portal". DART-Europe proporcionará también un repositorio y servicios para permitir a las instituciones y países que carezcan de sus propios repositorios con la infraestructura necesaria para hacer accesibles sus tesis. El propósito del portal es ofrecer una visión europea de su producción científica y la difusión internacional de todas las tesis doctorales europeas, aumentando así su visibilidad e incrementando su impacto científico.

#### **4. SITUACIÓN EN ESPAÑA**

En España, tal y como se manifiesta en las cada vez más numerosas iniciativas que se están poniendo en marcha, existe un creciente reconocimiento de que los contenidos digitales generados por la comunidad científica constituyen un activo institucional y nacional que debe ser gestionado de manera que se maximice su acceso y difusión y se asegure su preservación.

En el contexto universitario, la Red de Bibliotecas Universitarias Españolas (REBIUN) incluye, como uno de sus objetivos estratégicos, "estimular y potenciar la creación de un repositorio de acceso libre a la información científica producida en las Universidades Españolas, cuyo núcleo original sería la implementación de un portal de acceso libre al texto digital completo de las tesis doctorales".

Desde iniciativas regionales, el Consorcio de las Bibliotecas Universitarias Catalanas proporciona acceso libre al texto completo de las tesis doctorales presentadas en sus Universidades y ofrece la posibilidad de que cualquier Universidad española que lo desee pueda incorporarse (recientemente lo ha hecho la Universidad de Cantabria) a su base de datos TDX<sup>29</sup>; por su parte, la Universidad Complutense ya ha puesto en marcha su repositorio institucional<sup>30</sup>, cuyo volumen de contenido es el de mayor relevancia actualmente en España, dando acceso a más de 3.000 tesis doctorales a texto completo.

Fuera del ámbito de los repositorios institucionales, se han llevado a cabo iniciativas importantes en la línea de las revistas electrónicas de libre acceso:

la plataforma digital E-revistas, en el portal Tecnociencia, creada bajo el patrocinio de la FECYT, ofrece acceso libre a revistas digitales en español. DIALNET<sup>31</sup> soporta en la actualidad el estándar OAI-PMH y está registrado como proveedor de datos. España participa en el proyecto SciELO, resultado de la cooperación entre [BIREME](#) (Centro Latinoamericano y del Caribe de Información en Ciencias de la Salud) y [FAPESP](#) (Fundación de Apoyo a la Investigación del Estado de São Paulo). Está siendo desarrollado en España por la Biblioteca Nacional de Ciencias de la Salud ([BNCS](#)), gracias al acuerdo de colaboración establecido entre la [OPS/OMS](#) y el [Instituto de Salud Carlos III](#). El principal objetivo es contribuir al desarrollo de la investigación, aumentando la difusión de la producción científica nacional, y mejorando y ampliando los medios de publicación y evaluación de sus resultados.

<sup>28</sup> <http://www.dartington.ac.uk/dart/index.asp>

<sup>29</sup> <http://www.tdx.cbuc.es/>

<sup>30</sup> <http://www.ucm.es/eprints/>

<sup>31</sup> [http://dialnet.unirioja.es/ayuda/ayuda\\_oai.jsp](http://dialnet.unirioja.es/ayuda/ayuda_oai.jsp)

## 5. OBJETIVOS DEL PROYECTO

Frente al dinamismo internacional en la creación de archivos digitales en el contexto del acceso abierto a las publicaciones científicas, la situación en nuestro país está todavía en una fase incipiente, por lo que el acceso y visibilidad de nuestra producción científica están minimizados. Nuevos repositorios se crean y crecen a un ritmo cada vez más acelerado y muchos países están llevando a cabo proyectos cooperativos para aumentar la visibilidad de sus contenidos creando interfaces y portales de acceso centralizado a su documentación científica. Incorporarse a los desarrollos en este campo, tales como la "Open Archive Initiative", es crucial.

El objetivo del Proyecto Madroño e-Ciencia es crear una infraestructura y unos servicios avanzados para recoger, acceder, preservar y distribuir en soporte digital la producción científica generada por la comunidad investigadora y académica de la Comunidad de Madrid.

Los objetivos específicos del Proyecto Madroño e-Ciencia son:

- Ofrecer una visión centralizada regional de la producción científica generada en la Comunidad de Madrid para incrementar su nivel de impacto y capacidad de influencia en el ámbito nacional e internacional.
- Crear un espacio "open access" para las publicaciones científicas.
- Obtener financiación de fuentes externas, nacionales y europeas.

- Ampliar y mejorar el acceso a la producción científica propia.
- Reunir la producción intelectual de nuestras instituciones, la cual de otra forma permanecería dispersa bajo el sistema tradicional de publicación científica.
- Ofrecer un alto valor añadido a los contenidos digitales generados por la actividad científica y académica en nuestra Comunidad.
- Ofrecer a los investigadores de la Comunidad de Madrid un modelo complementario de publicación científica que incremente el impacto de sus trabajos en el entorno del acceso abierto y que proteja sus derechos individuales a la propiedad intelectual en el entorno de los derechos de autor y copyright.
- Asegurar la preservación y permanencia a largo plazo de los contenidos digitales generados por la comunidad científica de la Comunidad de Madrid.
- Modernizar la gestión de la información científica y académica de nuestras instituciones en el entorno de los desarrollos de las TIC`s.
- Crear nuevos medios de difusión que maximicen la visibilidad y por tanto el impacto de la ciencia de Madrid.
- Incentivar la creación y el uso de la producción científica propia.
- Proporcionar un entorno que favorezca y estimule la transferencia y el intercambio de experiencia y conocimiento entre los participantes en el proyecto
- Construir un sistema robusto, escalable, modular, interoperable, basado en estándares y protocolos abiertos, que explore la gestión de los derechos digitales y que tenga en cuenta cuestiones de archivo y preservación.
- Conseguir una masa crítica de contenidos digitales de calidad en castellano y/o de científicos españoles.
- Facilitar y estimular la formación en conocimientos críticos como metadatos, uso y aplicaciones de xml, desarrollos de software abierto, servicios web, etc. con el fin de crear una comunidad de expertos que asegure la continuidad del proyecto y que puede ser utilizada como un recurso nacional.
- Identificar y difundir guías de "mejores prácticas" y ofrecer servicios de asesoramiento técnico.

El proyecto Madroño e-Ciencia responde también a los objetivos generales contemplados en el IV Plan Regional de Investigación Científica e Innovación Tecnológica de la Comunidad de Madrid:

- Promover la creación de capacidades regionales colectivas en interacción con otros entornos regionales o nacionales, dirigidas a la generación de conocimiento y su transformación en riqueza y bienestar, impulsando la democratización de la ciencia y la tecnología.
- Facilitar el acercamiento de la ciencia y la tecnología a los ciudadanos en condiciones de eficiencia con el fin de mejorar el bienestar y la competitividad territorial.
- Consolidar Madrid como nodo importante en la red europea y global de "regiones de conocimiento" impulsando el desarrollo de la ciencia y la tecnología.
- Implicar a la sociedad madrileña en el sistema regional de ciencia y tecnología.
- Promover el desarrollo de una industria regional de investigación y conocimiento.
- Impulsar la creación de una imagen de marca en torno a la investigación madrileña.
- Consolidar un mercado regional de conocimiento.
- Integrar Madrid en las redes globales de conocimiento.
- Cooperar, integrar y complementar las actuaciones en IDI a todos los niveles, el interregional, con la Administración General del Estado y con las redes globales del conocimiento internacional.
- Utilizar intensivamente las tecnologías de acceso a la información y al conocimiento.
- Divulgar la política regional en ciencia y tecnología.

Además, el proyecto pretende conseguir las siguientes metas:

- Liderar el movimiento por el acceso abierto a la información científica ("Open Access") en España y erigirse en uno de los líderes a nivel mundial.

- Ser referencia en el campo de los archivos abiertos de información científica a nivel mundial.
- Constituirse en centro de experiencia en estos temas con capacidad para asesorar a otras instituciones interesadas.
- Constituirse en el núcleo original de un portal de acceso abierto a la información científica española y en español en todo el mundo.
- Crear un entorno de cooperación abierto y constante con el sistema nacional de investigación y con el Espacio Europeo de investigación y del conocimiento.

## **6. ORGANIZACIÓN**

El proyecto propone tres tipos de participación:

### **Grupo de Gestión**

Este grupo corresponde al Consorcio Madroño que será el responsable del Proyecto en su conjunto, dirigiendo y coordinando su desarrollo y dirección estratégica y tomando las decisiones políticas y operacionales.

Su estructura será la siguiente:

#### Consejo de Gobierno

Compuesto por los Rectores o personas en quien deleguen de cada una de las Entidades Consorciadas.

#### Comisión Técnica

Formada por el Presidente del Consorcio, los Directores de las Bibliotecas del Consorcio y el Director Técnico del Consorcio

#### Grupo de trabajo

Responsable de investigar, evaluar, identificar y proponer soluciones técnicas relativas al proyecto, presentando sus informes a la Comisión Técnica para ayudarle en la toma de decisiones. Este grupo estará en contacto y trabajará en colaboración con el Grupo de Actividad.

Participan en este Grupo bibliotecarios de la Universidad Carlos III de Madrid, La Universidad Complutense, la UNED y un documentalista del Consorcio Madroño.

## Grupo de Actividad

Este grupo incluye personas y organizaciones, que pueden ser tanto locales como internacionales, que pueden contribuir al desarrollo del proyecto llevando a cabo determinadas tareas. Estas tareas pueden consistir, por ejemplo, en asesoramiento o proporcionar desarrollos parciales de software, etc.

El proyecto solicitará la participación en este grupo a profesores, departamentos y expertos así como a comunidades científicas directamente comprometidas que proporcionen evaluación, materiales, control de calidad, etc., que ayuden a encauzar el proyecto y que den solidez técnica y científica al mismo.

## Participantes Asociados

El proyecto invitará a participar en este grupo a las Instituciones interesadas que no pertenezcan al Consorcio Madroño, de manera que puedan aportar sus ideas en las sucesivas etapas estratégicas del proyecto y al mismo tiempo estén informados de su desarrollo con vistas a su posible participación en el mismo.

## 7. MODELO DEL PROYECTO

El modelo propuesto para el desarrollo de la plataforma "**Madroño e-Ciencia**" es el mismo adoptado internacionalmente en proyectos similares y que se deriva de la elección del protocolo OAI-PMH como protocolo de intercambio de información en el contexto del acceso abierto a la información científica.

Se propone un modelo distribuido, basado en una arquitectura abierta, soportado sobre tecnologías estandarizadas y de amplia aceptación en el ámbito de desarrollo de esta iniciativa.

La arquitectura del modelo se organiza en dos niveles:

- Un **nivel de datos**, en el que la infraestructura se implementa y se mantiene. Cada institución establece un repositorio que recoge los contenidos generados por la actividad intelectual de su comunidad y la mantiene disponible para su uso de acuerdo con estándares internacionales. Este nivel corresponde a la **Red de Repositorios Madroño**.
- Un **nivel de servicio**, en el que se utiliza el material del nivel de datos para desarrollar servicios de valor añadido para los científicos,

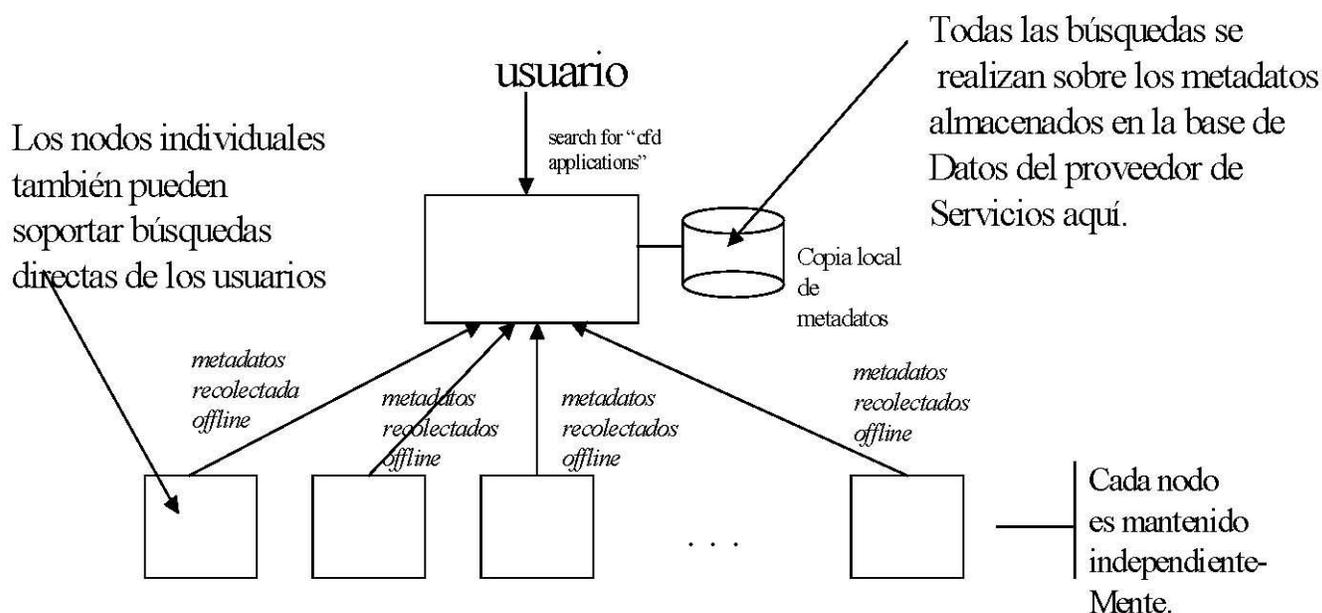
estudiantes, o cualquier usuario interesado. El principal es la posibilidad de acceder a los contenidos de esos repositorios distribuidos desde un único punto de búsqueda como si se tratara de una única colección. Además, las posibilidades son numerosas: desde un servicio de búsqueda sencilla hasta la creación de portales temáticos, canales de exportación RSS, servicios de generación de bibliografías, etc. Este nivel corresponde al **Portal Madroño e-Ciencia**.

## **8. FUNCIONAMIENTO**

Los usuarios (autores de los documentos científicos) se registran y autoarchivan una copia electrónica de su trabajo en el repositorio apropiado. Estos archivos permiten al mismo tiempo disponer de mecanismos de control de acceso y calidad y la posibilidad de incluir diferentes tipos de documentos. Estos repositorios contienen, además de los documentos, los metadatos asociados, los cuales son recolectados por el harvester del proveedor de servicios usando el protocolo OAI-PMH y luego son pasados a un servicio web externo que crea los servicios de valor añadido sobre esa información.

# Recolección de metadatos

- Extracción de metadatos desde fuentes distintas
- Construcción de servicios sobre las copias locales de metadatos
  - Recursos permanecen en los repositorios remotos



El proveedor de servicios (el Portal Madroño e-Ciencia) lanza su "harvester" (o programa de interrogación a un proveedor de datos, según el protocolo OAI-PMH) contra un nodo y recolecta sus metadatos (no el documento en sí, que permanece en el servidor del repositorio recolectado) que almacena en su propia base de datos. Sobre esa base de datos construye la interfaz de búsqueda y los servicios de valor añadido a los que accederá el usuario final, el cual podrá buscar desde un único punto, de forma transparente para él, en todos los contenidos de los distintos repositorios recolectados por el proveedor de servicios, como si se tratara de una única colección.

Cuando un usuario final encuentra un documento en el Portal del proveedor de servicios puede acceder al documento completo, almacenado en el servidor del repositorio al que pertenece, a través de la url que le ofrece uno de los

metadatos del registro.

La conformidad de cada nodo con el protocolo OAI-PMH significa que sus metadatos, y por tanto sus contenidos, estarán accesibles no sólo para el proveedor de servicios Madroño, sino para cualquier proveedor de servicios del mundo que los solicite, incluidos buscadores de Internet como Google Scholar o portales temáticos o educativos.

A su vez, el Portal Madroño e-Ciencia podrá recolectar no sólo los metadatos de la red de repositorios Madroño sino los de cualquiera de los repositorios conformes con el protocolo OAI-PMH que existen en el mundo; sobre esta base se construiría el portal de acceso a la investigación española y en español disponible en repositorios distribuidos por todo el mundo.

## **9. NIVEL DE DATOS: RED DE REPOSITARIOS MADROÑO**

Para conseguir un impacto significativo del proyecto es necesario reunir una masa crítica de contenidos de calidad. Por ello, parte de los recursos disponibles estarán dedicados a la implementación de repositorios (proveedores de datos).

### **Arquitectura**

Se opta por una arquitectura distribuida en la que cada repositorio es independiente y puede ser contemplado como un nodo de una red regional de repositorios, capaz de integrarse a su vez en otras redes, nacionales e internacionales.

La arquitectura abierta de la red permite la integración en la misma de cualquier Institución de la Comunidad de Madrid que implemente un repositorio conforme con las especificaciones establecidas por el Proyecto en relación con los repositorios de la red Madroño.

Igualmente podrán integrarse los repositorios de carácter disciplinario (archivos abiertos que almacenan contenidos no por su origen institucional sino por su contenido temático, especializados en una disciplina o área de investigación concreta) desarrollados y mantenidos por o con la participación de comunidades científicas de la Comunidad de Madrid.

El proyecto ha optado por una solución distribuida en lugar de centralizada por las siguientes razones:

- **Escalabilidad**

Crear un único repositorio centralizado no es una arquitectura viable, ya que no escalaría bien dado el tamaño y número potenciales de objetos digitales que se acumularán con el tiempo.

- **Reutilización de contenidos**

La arquitectura distribuida debe permitir a cada participante reutilizar sus contenidos en otros sistemas, proyectos o servicios sin tener que crearlos, almacenarlos o duplicarlos cada vez.

- **Internacionalidad**

El modelo distribuido es el subyacente a la tecnología abierta y sigue la dirección de la mayoría de los proyectos internacionales en este ámbito.

- **Institucionales**

El modelo distribuido permite a cada nodo o repositorio de la red la definición de sus propias políticas locales en relación con la creación, uso y difusión de sus contenidos, y ofrece a cada institución una herramienta básica para garantizar un almacenamiento digital estructurado fiable de su propia propiedad intelectual. Esto es un factor importante en la era digital.

Un repositorio de la red Madroño es una prestación consistente en hardware, software, data y procedimientos que:

- contiene objetos digitales que representan cualquier documento científico y académico generado por una Institución científica o Académica.
- asegura la identificación adecuada de los objetos digitales mediante metadatos y un identificador único y persistente
- proporciona funcionalidades de gestión y archivo de los objetos digitales.

Los acuerdos básicos de partida son:

- Todos los repositorios serán conformes con el protocolo OAI-PMH.
- Ofrecerán los documentos a texto completo.
- Permitirán el acceso a los mismos sin ningún tipo de restricción. El acceso a los contenidos será libre y gratuito. El proyecto Madroño e-Ciencia suscribe los principios del movimiento por el acceso libre a la información, entendiendo como tal la definición propuesta por la "Budapest Open Access Initiative"<sup>32</sup> como "disponibilidad libre en internet, permitiendo a cualquier usuario leer, bajar, copiar, distribuir, imprimir, buscar o enlazar los textos completos de los documentos, o

usarlos con cualquier otro propósito legal, sin ninguna barrera económica, legal o técnica más allá de aquellas inseparables de conseguir el acceso a la propia internet. La única limitación a la reproducción y distribución, y el único papel del copyright en este ámbito debe ser dar al autor el control sobre la integridad de su trabajo y el derecho a ser reconocido y citado apropiadamente”.

## Software

El proyecto Madroño e-Ciencia está comprometido con el fomento del software de código abierto para la creación de repositorios institucionales. Los más utilizados por otros proyectos similares son **Dspace**<sup>33</sup>, desarrollado en la MIT, Estados Unidos, **Eprints**<sup>34</sup>, desarrollado en el Reino Unido y con más implantación en Europa), y **Fedora**<sup>35</sup> (Flexible Extensible Digital Object and Repository Architecture), desarrollado conjuntamente por la Biblioteca de la Universidad de Virginia y la Universidad de Cornell en Estados Unidos.

<sup>32</sup> <http://www.soros.org/openaccess/read.shtml>

El proyecto Madroño e-Ciencia no establece ninguna imposición respecto al software a utilizar en cada repositorio, excepto la condición – esencial para el desarrollo del proyecto - de que soporte el protocolo OAI-PMH.

La utilización de software de código abierto es una estrategia para contener los costes de implementación y garantizar la escalabilidad del Proyecto.

El harvester del proveedor de servicios Madroño será capaz de interrogar diferentes implementaciones de software de proveedores de datos, entre otras cosas porque también va a recolectar repositorios que no pertenecerán a la red Madroño.

Sin embargo, la posibilidad de que cada institución pueda implementar su repositorio con el programa que considere más adecuado está en consonancia con el modelo del proyecto, puesto que permite a la institución planificar y gestionar de forma independiente el uso que quiere dar a sus contenidos más allá del proyecto Madroño e-Ciencia, por ejemplo, su interoperabilidad con su entorno e-learning, con su sistema de gestión bibliotecaria, las necesidades específicas de sus usuarios, etc.

## Contenidos:

El proyecto “Madroño e-Ciencia” suscribe la definición de “contenido” adoptada por SPARC (The Scholarly Publishing & Academic Resources Coalition)<sup>36</sup>: no hay distinción entre la literatura publicada y la denominada “literatura gris”.

Esto quiere decir que el contenido es:

- Erudito: tanto de carácter científico como académico.
- Producido, enviado o patrocinado por un miembro de la comunidad científica o cualquier otro agente autorizado por la política de la Institución.
- No efímero; el trabajo debe estar completo y preparado para su distribución
- Licenciado a perpetuidad; el autor debe ser capaz y tener la voluntad de ceder, no en exclusiva, a la institución el derecho a preservar y distribuir su trabajo a través del repositorio.

<sup>33</sup> <http://libraries.mit.edu/dspace-mit/index.html>

<sup>34</sup> <http://www.eprints.org/>

<sup>35</sup> <http://www.fedora.info>

<sup>36</sup> SPARC Institutional Repository Checklist & Resource Guide:

[http://www.arl.org/sparc/IR/IR\\_Guide.html#pub\\_material](http://www.arl.org/sparc/IR/IR_Guide.html#pub_material)

Los contenidos que puede albergar un repositorio son:

- Equivalentes a los impresos: artículos de revistas, ponencias, tesis, documentos de trabajo, capítulos de libros, informes técnicos, monografías, etc.
- Contenidos audiovisuales y multimedia
- Objetos de investigación: "data sets", simulaciones, software, etc...
- Objetos de aprendizaje: cualquier contenido, software o acontecimiento utilizado durante una sesión de aprendizaje soportada sobre tecnología "e-learning".

Los contenidos digitales que estarán disponibles en los repositorios serán, inicialmente, las tesis doctorales de las Universidades participantes. Muchas de estas tesis nacen ya digitales y se depositarán directamente en el repositorio a partir de su puesta en marcha; sin embargo, para lograr un contenido crítico suficiente lo antes posible, las Universidades acometerán además un proceso de digitalización retrospectiva de las tesis doctorales que están en otros formatos.

El objetivo es lograr que, una vez creados los repositorios, sean los propios

creadores de los contenidos los que los depositen en los mismos. El proyecto llevará a cabo actividades de información y difusión de la cultura del "autoarchivo" como forma alternativa de publicación de sus trabajos en el entorno del acceso abierto a la información científica.

La importancia de poder acceder en soporte digital a los contenidos generados por la actividad científica de las Universidades de nuestra Comunidad se pone de manifiesto en el hecho de que éstas, tal y como se menciona en el IV Plan PRICIT 2005-2008, generan el 45% de su producción científica.

Respecto a la calidad de los contenidos depositados en los repositorios institucionales es responsabilidad de cada Universidad. No se ha detectado ningún sistema de control editorial en ninguno de los proyectos similares investigados; el procedimiento parece autorregularse, puesto que se da por supuesto que a ningún investigador o departamento le interesa difundir lo más ampliamente posible cualquier cosa. El hecho de estar respaldado por la Universidad, así como delimitada la capacidad de depositar documentos a las personas autorizadas por las políticas de cada Institución, parece ofrecer suficiente garantía de que los contenidos depositados tienen el carácter científico y académico que se requiere.

Existe, sin embargo, una preocupación entre los profesores en cuanto a la categoría científica de los contenidos de los repositorios institucionales. En este sentido, los repositorios de la red Madroño establecerán una distinción muy clara entre los artículos que han sido sometidos a la revisión por pares ("peer reviewed") y aquellos que no lo han sido.

Se promoverá la proyección de contenidos en el idioma castellano, pero también de otros (especialmente el inglés).

## **Aspectos legales**

### **Derechos de gestión, distribución y preservación otorgados a la Institución responsable del repositorio.**

Para permitir a la Universidad la gestión, distribución y preservación del material depositado en el repositorio institucional, es necesario que cada autor que contribuya ceda a la institución un derecho no exclusivo e irrevocable para distribuir el contenido de su trabajo, migrar su formato con fines de preservación y mantenimiento del contenido a perpetuidad. A pesar de que no es posible establecer un único acuerdo o licencia de depósito, dada la independencia que se pretende dar a las decisiones locales de cada Universidad en cuanto a sus políticas en relación con los repositorios institucionales, el proyecto "Madroño e-Ciencia" creará un acuerdo plantilla que sirva de guía y orientación para la elaboración de los respectivos acuerdos de depósito y que asegure que los principios generales comunes estarán cubiertos en todos ellos.

La firma de este acuerdo o licencia de depósito no supone, en ningún caso, la pérdida o cesión de los derechos de propiedad intelectual y de copyright del autor, el cual sólo ha hecho una cesión no exclusiva, lo que significa que retiene el control total sobre su trabajo para usarlo y distribuirlo en otros contextos.

El problema surge cuando se trata de depositar artículos que se quieren publicar en una revista científica o que ya se han publicado. En este caso, la mayoría de las veces las editoriales exigen la cesión exclusiva de los derechos de autor sobre el artículo al investigador, lo cual significa, en muchos casos, la imposibilidad de depositar la versión publicada en la revista en el repositorio institucional. Ya hemos comentado, sin embargo, cómo esta situación está empezando a cambiar, puesto que son ya frecuentes los casos de editoriales que permiten a los autores el depósito de sus artículos también en los repositorios de su Institución. El proyecto SHERPA<sup>37</sup> ofrece un servicio de información sobre las distintas políticas en este sentido de varias importantes editoriales internacionales.

Respecto a la asignación de números de control bibliográfico (ISSN, ISBN, etc.) será decisión de cada Universidad. La asignación de estos identificadores no supondrá, en cualquier caso, ninguna restricción al acceso libre y gratuito a los contenidos de los repositorios.

<sup>37</sup>

<http://www.sherpa.ac.uk/romeo.php>

El proyecto estudiará las implicaciones legales en relación con la creación de un repositorio independiente para alojar los contenidos de los investigadores que no tengan acceso a ningún repositorio digital institucional.

### **Derechos de uso de los contenidos por parte del usuario final.**

El Proyecto estudiará y evaluará el uso de la licencia "Creative Commons"<sup>38</sup>, cuya versión española ha sido ya incorporada.

## **10. NIVEL DE SERVICIOS: PORTAL "MADROÑO e-Ciencia"**

El portal "Madroño e-Ciencia" constituye el punto de acceso desde el cual los usuarios finales podrán acceder, de forma transparente, libre y gratuita a través de Internet, a los contenidos digitales almacenados en los repositorios de la red Madroño y a los servicios de valor añadido construidos sobre la base de esos contenidos.

El sistema funciona sobre la arquitectura del protocolo OAI-PMH: se creará un servicio central que recolectará y actualizará de forma periódica los metadatos disponibles (conformes al protocolo OAI-PMH) en los distintos repositorios de la red Madroño. Estos metadatos serán almacenados en el servidor de este servicio central. La implementación, gestión y mantenimiento de este servicio central serán responsabilidad del personal del Consorcio Madroño.

El sistema que ha recolectado los metadatos (harvester) hará disponibles estos datos al usuario final mediante un sistema de búsqueda que además posibilitará el acceso al texto completo del documento mediante la url donde lo haya guardado la correspondiente institución.

Con objeto de poder hacer accesible esta información a dispositivos distintos de los navegadores de internet, tales como móviles, PDA's, etc., el resultado de la búsqueda estará disponible en formato xml (siguiendo el paradigma MVC de la independencia de los contenidos y su presentación).

El proyecto prevé una primera etapa de examen y evaluación del funcionamiento del sistema antes de iniciar la recolección de metadatos de otros repositorios que no pertenezcan a la red Madroño.

El proveedor de servicios Madroño es, a su vez conforme con el protocolo OAI-PMH, es decir, los metadatos que ha almacenado en su servidor están disponibles para ser recolectados por otros proveedores de servicios o portales nacionales e internacionales como por ejemplo OAISTER<sup>39</sup> o el portal "DART-europe" de tesis doctorales europeas, y otros, incrementando así la exposición de los contenidos de los repositorios de la red Madroño.

<sup>38</sup>

<http://creativecommons.org/worldwide/es/>

El sistema de recolección de metadatos no sólo permite al proveedor de servicios ofrecer un punto de acceso único a todos los contenidos distribuidos en repositorios distintos sino que le ofrece la posibilidad de proporcionar un valor añadido a esos metadatos.

En fases sucesivas del proyecto, el proveedor de servicios Madroño desarrollará e incorporará al Portal Madroño e-Ciencia los siguientes servicios:

- Exportación de canales Atom, Rdf, RSS, etc.
- Estadísticas
- Asignación de metadatos según un sistema de clasificación basado en las áreas académicas de investigación.
- Implementación de un servidor Z39.50
- Exportación en XML
- Generación de bibliografías
- Generación de presentación de contenidos personalizada
- Herramientas para la creación colaborativa de documentos
- Sistema de búsqueda y recuperación de la información multilingüe.
- Foros y soporte para comunidades científicas de discusión

El proyecto estudiará la implementación de un servicio web de análisis de citas. Se investigará el proyecto "Open Citation"<sup>40</sup> de la Universidad de Southampton, que a su vez está siendo implementado en el proyecto "e-print UK" del Reino Unido.

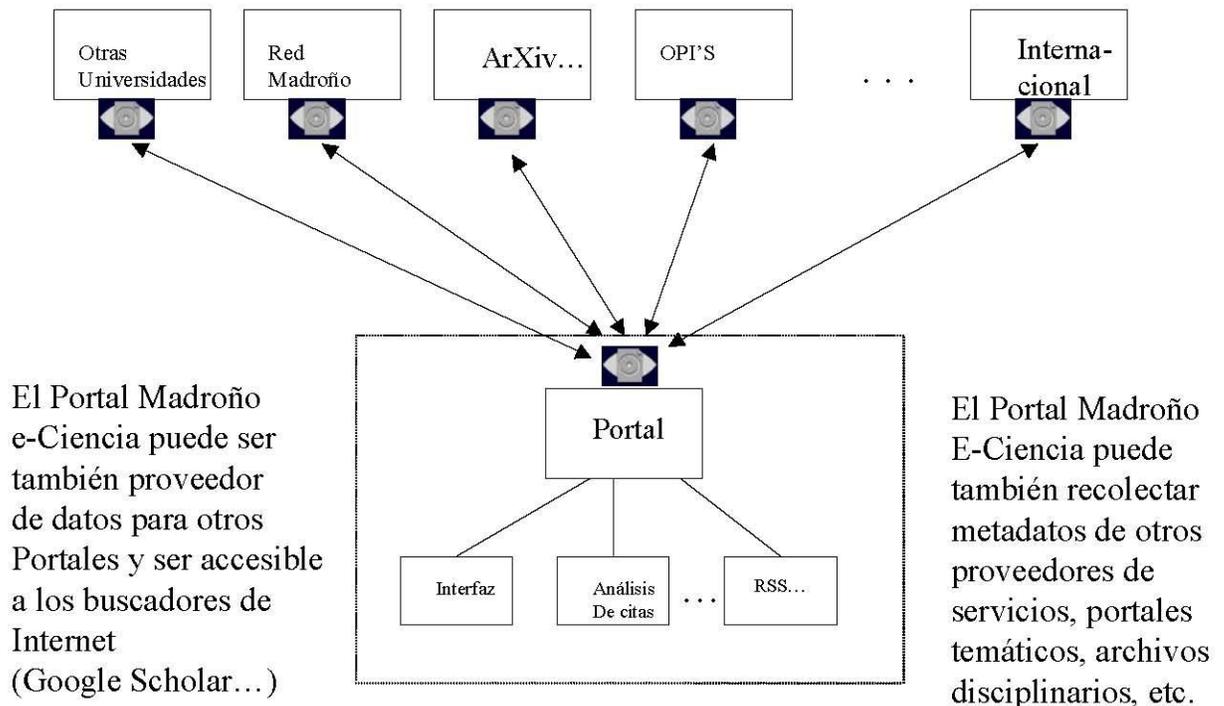
Para el estudio y desarrollo de estos servicios se identificará y contactará con los departamentos y comunidades científicas que estén relacionados con esos desarrollos y estén interesados en el proyecto, para solicitar su colaboración.

El portal Madroño e-Ciencia será accesible desde las páginas web de las Universidades del Consorcio, en el sitio web del Consorcio, en el sitio web IDI de la Comunidad de Madrid y desde el propio sitio web del "Proyecto Madroño e-Ciencia", así como desde cualquier lugar en la web en la que se ponga un enlace al mismo.

<sup>39</sup> <http://oaister.umdl.umich.edu/o/oaister/>

<sup>40</sup> <http://opcit.eprints.org/>

## Arquitectura Portal Madroño e-Ciencia



### 11. EVALUACIÓN

El proyecto llevará a cabo un proceso de evaluación continua del servicio, incluyendo todos sus aspectos, culturales, económicos, organizativos y tecnológicos. Un objetivo clave de este proceso de evaluación es asegurar la capacidad del proyecto para resolver problemas, corregir y rediseñar su planificación y comprobar su viabilidad en el futuro.

Algunos de los indicadores que se utilizarán para la evaluación del servicio son:

- Número de repositorios puesto en marcha
- Número de documentos digitalizados
- Número de documentos accesibles desde el Portal
- Número de consultas

- Número de descargas
- Número de asistentes a las Jornadas

Como parte del proceso de evaluación se generarán estadísticas y se remitirán periódicamente informes internos a los participantes; se emitirá un informe con los resultados de la evaluación al final de cada etapa del proyecto.

## 12. DIFUSIÓN

El proyecto difundirá sus resultados lo más ampliamente posible, tanto en nuestro país como internacionalmente. Esta labor de difusión se llevará a cabo de forma continua a lo largo de toda la duración del desarrollo del proyecto.

Canales específicos de comunicación serán:

**Sitio web "Madroño e-Ciencia":** este sitio web será regularmente actualizado para ofrecer las últimas noticias en relación con el desarrollo del proyecto. Todos los documentos públicos relacionados con el proyecto estarán accesibles en la página web. Además, el sitio web del proyecto ofrecerá enlaces a sitios relacionados.

**Anuncios:** importantes acontecimientos o etapas del desarrollo del proyecto serán comunicados en listas de discusión relevantes, tanto nacionales como internacionales, así como vía RSS.

**Publicaciones:** se enviarán publicaciones sobre el proyecto a conferencias y revistas relevantes.

**Presentaciones:** se utilizarán Conferencias relevantes como oportunidad para la difusión del Proyecto.

**Otras comunidades:** el proyecto hará un esfuerzo especial en difundir sus resultados entre la comunidad investigadora.

**Jornadas:** el proyecto organizará Encuentros/Jornadas para discutir y promocionar el acceso abierto a la información científica y promover un cambio de cultura en el ámbito de la publicación científica en el entorno de la e-ciencia.

## 13. COLABORACIÓN/FINANCIACIÓN

El proyecto Madroño e-Ciencia se declara convencido de que la única garantía de éxito del mismo es la colaboración, la cual, además es una estrategia clave para la obtención de fuentes de financiación externa.

Por ello, desde el principio, el proyecto Madroño buscará la colaboración de aquellas personas y organizaciones, tanto públicas como privadas, relacionadas con el ámbito de interés del Proyecto.

Como organismo clave en España, el Proyecto establecerá contacto con el Consejo Superior de Investigaciones Científicas. También contactará con otras iniciativas nacionales como el proyecto TDX del Consorcio de Bibliotecas de la Comunidad Autónoma de Cataluña, el proyecto e-revistas del portal Tecnociencia, el proyecto Dialnet) e iniciativas europeas similares.

La sostenibilidad del proyecto a largo plazo se basará en diferentes vías de financiación.

Por una parte se cuenta con las aportaciones de los miembros y participantes en el Proyecto y los fondos aportados por la Comunidad de Madrid.

El proyecto, como recurso de interés nacional, investigará programas de financiación del Ministerio de Educación y Ciencia y de la Secretaría de Estado de Telecomunicaciones y para la Sociedad de la Información (Ministerio de Industria, Turismo y Comercio)

Dado que el proyecto comparte objetivos y procedimientos con varias iniciativas en marcha en otros países, se considera apropiado investigar la participación del Proyecto Madroño e-Ciencia en programas financiados por la Unión Europea.

La colaboración con organismos privados será valorada como posible fuente de financiación.

#### **14. RECURSOS NECESARIOS**

<b>Para la creación de la red de repositorios Madroño</b>
---

**Recursos humanos:** se requerirá personal bibliotecario y analistas; las instituciones participantes aportarán su propio personal para la implementación de sus respectivos repositorios.

**Software:** a pesar de que existe software libre ya disponible para la implementación de repositorios institucionales, se requerirá instalar, configurar y adaptar el software elegido a las características particulares de cada institución y a los requerimientos del proyecto.

**Hardware:** básicamente cada institución requerirá un servidor capaz de alojar tanto el software como los datos a gestionar (metadatos y documentos).

### **Para la creación del Portal Madroño**

**Recursos humanos:** el Consorcio Madroño aportará su personal para la implementación del Portal.

**Software:** existe software libre para implementar un proveedor de servicios. El proyecto estudiará y evaluará los programas disponibles y se asegurará en su elección que se alinea con desarrollos similares internacionales.

**Hardware:** el proveedor de servicios Madroño requerirá un servidor cuya capacidad será determinada teniendo en cuenta que sólo almacenará metadatos, ya que los textos completos permanecerán alojados en los servidores de sus respectivas instituciones.

### **Para la creación del repositorio independiente que aloje los contenidos de los investigadores que no dispongan de acceso a un repositorio institucional**

**Recursos humanos:** dado que este repositorio estaría gestionado por el Consorcio Madroño como tal, sería conveniente estudiar, en la fase correspondiente del proyecto, la contratación de una persona.

**Respecto al hardware,** se podría utilizar el mismo servidor del proveedor de servicios Madroño, en cuyo caso habría considerar una mayor capacidad del mismo para que pueda albergar también los textos completos y no sólo los metadatos.

En esta escalabilidad que se prevé para el proyecto y llegados a este punto de desarrollo, sería imprescindible contar con el apoyo informático, tanto hardware, como software y personal de Madrid I+D, ya que el proyecto abarcaría al conjunto de la comunidad investigadora de la CM.

**Software:** se estudiarán y evaluarán los siguientes programas: E-prints, Dspace.

### **Otros recursos necesarios**

**Proceso de digitalización retrospectiva:** será necesario destinar parte de los recursos disponibles a la digitalización masiva de las tesis doctorales no disponibles en formato digital.

**Difusión:** también será necesario destinar parte de los recursos a las actividades de difusión del proyecto así como de desarrollo de los contenidos entre la comunidad investigadora y docente, como organización de Jornadas, viajes a Congresos, etc.