

# Portal “e-ciencia”

VI SEMANA DE LA CIENCIA MADRID 2006

## MESA REDONDA

*Iniciativas “open access”. Los repositorios científicos institucionales al servicio de la sociedad*

# Marco del Proyecto

Constituye una de las líneas de actividad del “Programa de Cooperación Interbibliotecaria” enmarcado en el Contrato Programa entre la CAM 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).

# Participantes

- La Comunidad de Madrid
- Universidad Carlos III de Madrid
- Universidad Autónoma de Madrid
- Universidad Nacional de Educación a Distancia (UNED)
- Universidad Politécnica de Madrid
- Universidad Juan Carlos I de Madrid
- Universidad de Alcalá
- Universidad Complutense de Madrid

# Contexto

Es una iniciativa que lleva a la práctica los principios del acceso abierto a la información científica de la Declaración de Berlín 2003, firmada por todas las Universidades participantes en el proyecto.

Es un proyecto abierto, cuyo germen son las Universidades del Consorcio Madroño, pero que está abierto a la participación de todas las Instituciones y Organizaciones de investigación de la Comunidad de Madrid.

# Definición

Una **plataforma digital**

- Que de soporte a las iniciativas de publicación en acceso abierto a los investigadores de la Comunidad de Madrid,
- Que permita almacenar, gestionar, organizar, tener disponibles y preservar los contenidos digitales generados por la actividad científica de la Comunidad de Madrid y
- Que ofrezca unos servicios fundamentales para incrementar su visibilidad, difusión, accesibilidad y uso en abierto a través de Internet.

# Arquitectura

Soportada sobre el protocolo OAI-PMH:

- **Nivel de datos**: infraestructura de adquisición, organización, gestión, almacenamiento y disponibilidad de los contenidos: **red de repositorios institucionales + 1 repositorio independiente**
- **Nivel de servicios**: implementación de servicios de valor añadido sobre la agregación de los contenidos disponibles en la infraestructura de datos: **Portal “e-ciencia”**

# Nivel de datos

Cada Institución es responsable de su propio repositorio. Hemos limitado las acciones conjuntas a aquellas que aseguren la interoperabilidad entre los repositorios de la red en términos tecnológicos, de políticas y organizativos.

Características:

- Arquitectura distribuida
- Software
- Contenidos + certificación científica
- Cuestiones legales

# Arquitectura distribuida

¿Por qué no hemos creado un único repositorio para todos?

- No existe una única solución que sirva para todos.
- Existe el protocolo OAI-PMH.

Además...

# 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

# Internacionalidad

El modelo distribuido es el subyacente a la tecnología abierta de la web y sigue la dirección de la mayoría de los proyectos similares que se están desarrollando en el mundo:

DareNet (Holanda)

Arrow (Australia)

E-prints UK (Reino Unido)

# Integración

La arquitectura distribuida facilita la integración de otros nodos a la red que pueden no tener la naturaleza institucional, por ejemplo los archivos abiertos temáticos de comunidades científicas o pertenecer a Instituciones con otras características, por ejemplo hospitales, que no se adaptarían bien a un repositorio institucional centralizado.

Permite, además, a cada nodo de la red su integración independiente en otras redes o proyectos.

# Otros fines de los repositorios institucionales

El “open access” no es el único fin por el que una Institución crea un repositorio.

- Dar visibilidad y difusión a los resultados de su actividad científica y académica con una cierta imagen de “marca”: ser indicador de prestigio.
- Asegurar la preservación de su producción intelectual
- Garantizar un almacenamiento digital estructurado fiable de su propia propiedad intelectual
- Ofrecer a su comunidad una herramienta de gestión de sus contenidos digitales.
- Permitir la reutilización de sus contenidos en otros escenarios: Dialnet, portal e-ciencia, su OCW, proyectos internacionales como DRIVER, etc.

# Auto-archivo



Es más fácil si estamos cerca de donde se crean y se auto-archivan los contenidos.

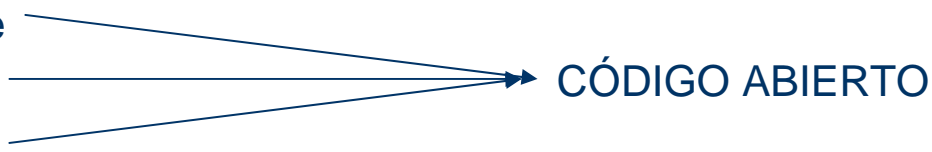
# Tecnología

El proyecto no establece ninguna imposición a los repositorios respecto a la tecnología utilizada salvo:

- Que sean conformes con el protocolo OAI-PMH

# Software

Sigue funcionando la regla de que no hay una única solución válida para todos:

- Dspace
  - Eprints
  - Fedora
- 
- CÓDIGO ABIERTO

\* Digitools

- Evolucionan más rápida y orgánicamente: muchos individuos y equipos trabajando en paralelo e intercambiando conocimiento
- Responde mejor a las mejoras y desarrollos locales (no responden a intereses comerciales)
- Nuevas versiones más frecuentes
- La seguridad se mejora porque el código está expuesto al mundo.
- Ahorra costes (no es gratis, pero permite diferentes tipos de decisiones presupuestarias)
- Podemos desarrollar nuevas funcionalidades a nuestro propio ritmo

# Contenidos

- Todos los contenidos que se generen como resultado de la actividad científica, académica y cultural de nuestras Instituciones. La garantía de su carácter científico o académico viene avalada por la Institución.
- Todo tipo de documentos: textuales, multimedia, software, simulaciones, gráficos, etc.
- Distinguiremos entre aquellos documentos depositados que han sido revisados por pares y aquellos que no lo han sido, para garantizar a los autores y usuarios la certificación científica de los contenidos.
- Los contenidos estarán almacenados en los servidores de los repositorios respectivos.
- El sistema permitirá el acceso al texto completo de los contenidos digitales de los repositorios.

# Mecanismos de carga de contenidos

- Auto-archivo (el más lento y complicado)
- Cargas de colecciones digitales de la biblioteca u otras ya existentes en la Universidad.
- Procesos de digitalización retrospectiva masiva. Hemos empezado por las tesis y revistas.

# Cuestiones legales

- **Licencia de depósito:** dictámenes de las respectivas asesorías jurídicas/licencias por defecto en los propios programas. El acuerdo entre la Institución y el autor para la preservación, organización y distribución a través del repositorio de los contenidos.
- **Licencia de uso:** el acuerdo entre el autor y el usuario final para el uso de los contenidos: “creative commons”

# Nivel de servicios

- El “**harvester**” + servidor para el proveedor de servicios donde se almacenan los metadatos recolectados, alojado y mantenido por el Consorcio Madroño
- Un **portal web** de acceso a la interfaz de búsqueda del proveedor de servicios + valores añadidos, creado, mantenido y accesible por la Comunidad de Madrid en su espacio Madri+d

# Por qué un nivel de servicios

Los repositorios deben ser considerados dentro de un contexto de información más amplio. Crear una pequeña escala de silos de información dentro de los repositorios no es una gestión de la información muy estratégica en la era “Google”

La promesa de que los repositorios institucionales servirán de soporte a la realización del “open access” y mejorarán la comunicación científica depende de la existencia de proveedores de servicios externos que exploten el conjunto de redes de repositorios institucionales a escala global.

# El harvester

- Hemos utilizado el software ARC –código abierto- (es el mismo que utiliza el servicio OAISTER)
- El responsable de Madroño trabaja directamente en el desarrollo de mejoras y nuevas propuestas del software con su creador Xiaoming Liu.
- El proveedor de servicios recolectará los metadatos de los repositorios participantes y, en fases sucesivas, podrá recolectar aquellos repositorios disponibles que considere oportuno incluir en el portal.
- El proveedor de servicios podrá ser también proveedor de datos para otros proveedores de servicios.
- Formato de metadatos soportado: Dublin Core

# Servicios primera fase

- Búsqueda simple (ya lo ofrece ARC)
- Búsquedas avanzadas (por campos) (ya lo ofrece ARC)

El punto fuerte que queremos ofrecer es la definición de SETS para permitir la navegación organizada por subcolecciones de materias y tipos de documentos.

Los sets permitirán dar una muestra significativa del peso específico de una determinada disciplina en el conjunto de la ciencia de la Comunidad de Madrid.

# Servicios fases sucesivas

- Búsqueda a texto completo
- Ofrecer el servicio en español, inglés y francés como mínimo.
- Integrar Herramienta de estadísticas
- Integración de gestores bibliográficos
- Integrar Herramienta de citas
- Sindicación de búsquedas
- Corrector ortográfico para ayudar a chequear que los documentos tienen la ortografía y codificación adecuada
- Permitir ordenar los resultados por distintos campos
- Permitir a los investigadores "crear una cuenta" para acceder más directamente a los datos de sus documentos
- Permitir personalizar el contenido, guardar la última configuración de búsquedas, el aspecto,...

# Portal web de acceso Madri+d

- Acceso a cada uno de los repositorios individuales participantes + repositorio independiente
- Asesoramiento sobre los procedimientos de “auto-archivo”
- Acceso a la interfaz de búsqueda común del proveedor de servicios + servicios de valor añadido
- Novedades sobre el desarrollo del proyecto
- Información sobre el movimiento OA a nivel mundial
- Actividades de difusión y formación: [blog “open access” en madri+d](#)
- Asesoramiento jurídico
- Foro de consultas

# Portal e-ciencia

---

¡Estará disponible este mes!

**¡Muchas gracias por su atención!**

Alicia López Medina

[alopezm@pas.uned.es](mailto:alopezm@pas.uned.es)