

# 4as Jornadas Os-Repositorios

5 de marzo de 2010

## Retos para la integración e intercambio de información: aprendiendo de NEEO

Integration and information exchange challenges:  
learnings from NEEO



Universidad  
Carlos III de Madrid  
[www.uc3m.es](http://www.uc3m.es)

# NEEO (Network of European Economist Online)



- Proyecto del Consorcio Nereus (integrado por bibliotecas especializadas en Economía), para potenciar el acceso abierto y la visibilidad de la investigación en Economía
- Proyecto aprobado por la Comisión Europea a través del programa eContentplus
- Duración: *sep-2007 / feb-2010*

# NEEO. “WP5. Interoperability infrastructure and gateway”

## Objetivos

Ofrecer acceso *abierto e integrado* a contenidos producidos en instituciones líderes (publicaciones y datasets)

Aumentar la visibilidad de la investigación económica

a través de un **portal multilingüe** que permita una infraestructura para ofrecer nuevos servicios

Como búsquedas a texto completo y listas de publicaciones de los investigadores

# NEEO. “WP5. Interoperability infrastructure and gateway”

Aspectos básicos del desarrollo

- **Interoperabilidad:** a nivel de esquemas, registros y repositorios.
- **Estandarización:** uso de estándares abiertos para la comunicación entre sistemas y la descripción de recursos.

# NEEO. “WP5. Interoperability infrastructure and gateway”

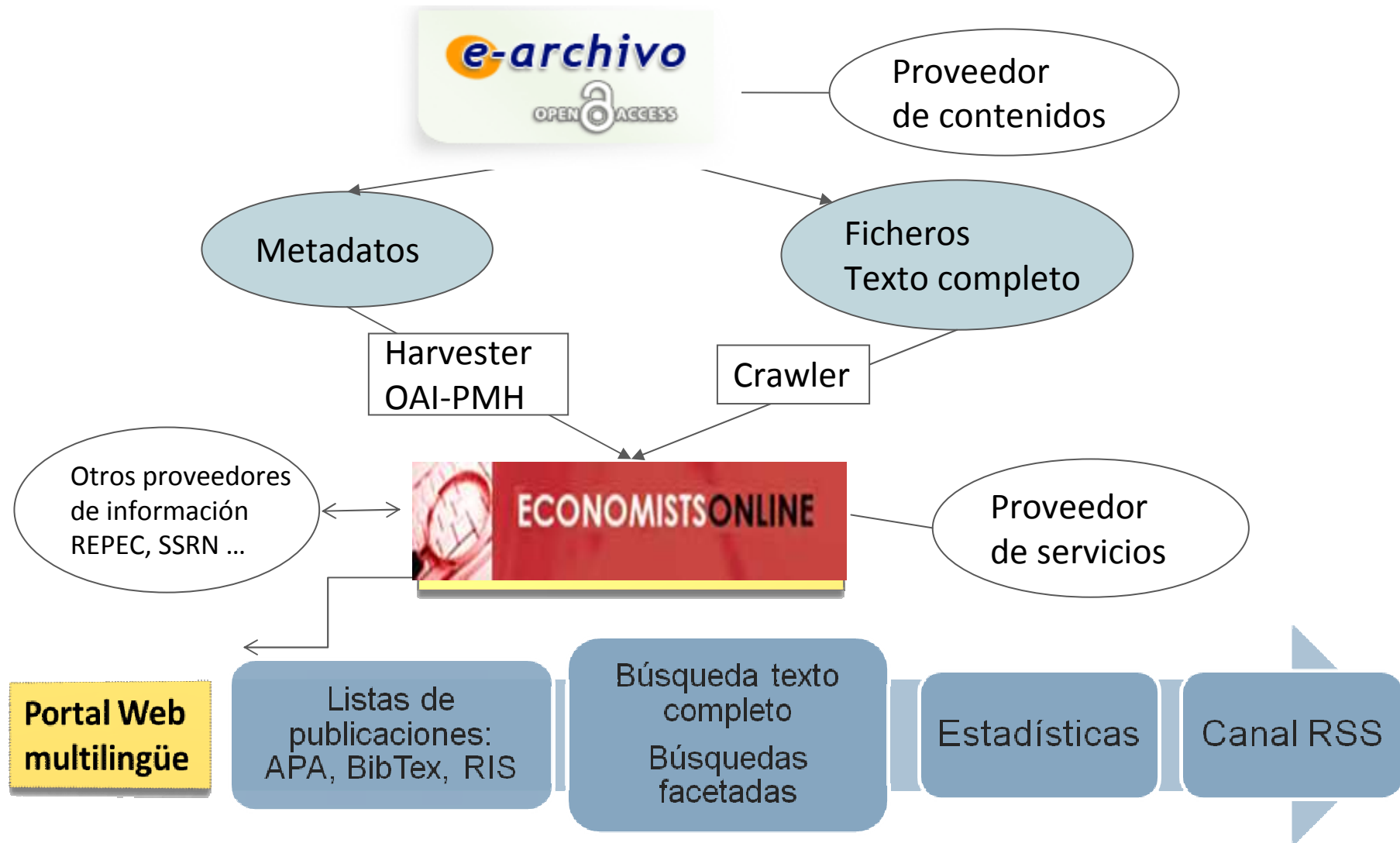
*¿Cómo se obtienen los contenidos del Portal?*

- Recolección de los 21 repositorios de las instituciones miembros del consorcio Nereus.
- Intercambio de datos con Repec (Research Papers in Economics)

Febrero 2010: 80.000 referencias bibliográficas / 20.500 publicaciones  
en acceso abierto

*Convive el acceso abierto al texto completo, el acceso restringido a suscriptores,  
y la no disponibilidad del documento.*

# PORTAL Economists Online. Proveedor de Contenidos / Proveedor de Servicios



# E-Archivo: proveedor de contenidos



- Plataforma Dspace 1.5.1
- Esquema de Metadatos Dublin Core Cualificado
- Conforme al protocolo OAI-PMH
- Repositorio Institucional
- Cobertura temática: ciencias sociales, humanidades, ciencia y tecnología.

# Portal Economists Online: proveedor de servicios



## *Directrices*

- NEEO Technical Guidelines (*Documentos textuales*)
  - Annex 1: Use of MODS for Institutional Repositories
  - Annex 2: DIDL Document Specifications for Repositories
- NEEO Technical Guidelines for the Exchange of Usage Metadata
- NEEO Dataset Description Guidelines



# NEEO Technical Guidelines.

*“The NEEO application profile”*: requerimientos específicos

- Uso de los estándares DIDL y MODS (en sinergia con DRIVER):

*“Cómo usar estos estándares para la descripción de publicaciones científicas y su manifestación en objetos digitales complejos, con el fin de crear una infraestructura destinada al intercambio y uso de información”.*

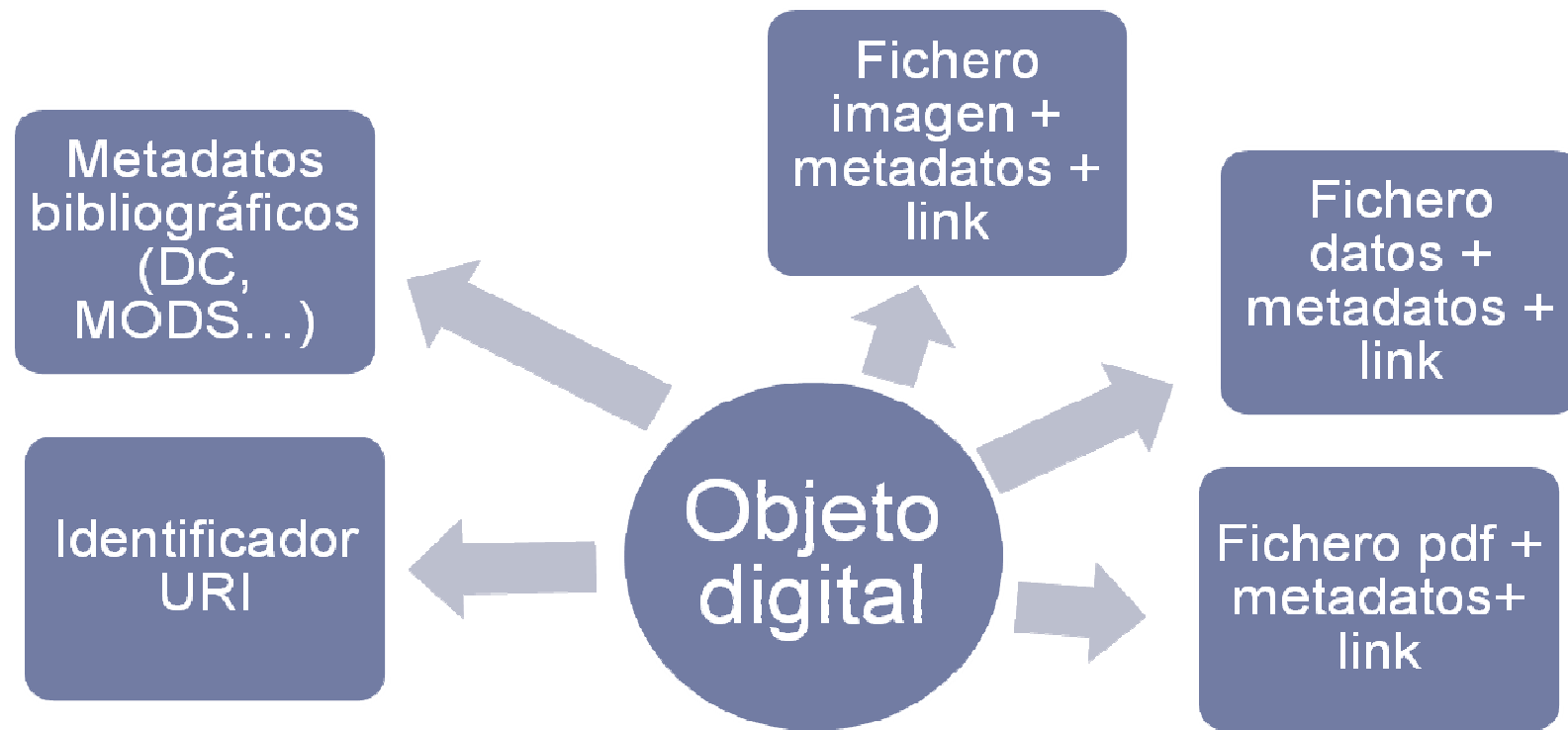
(conjunto de metadatos + documentos textuales + imágenes + ficheros de datos. Se requiere la descripción e identificación de cada uno de los ficheros individualmente)

\* By Benoit Pauwels (ULB)

# NEEO Technical Guidelines. DIDL

*Representación de un objeto digital mediante DIDL-XML*

(Digital Item Declaration Language)



# NEEO Technical Guidelines. MODS

## *Representación de los metadatos bibliográficos*

**MODS:** formato de intercambio de metadatos entre repositorio y portal EO

Características:

- Estándar internacional de la LC que permite describir publicaciones con un alto nivel de detalle, y construir servicios añadidos de calidad.
- Estructura jerárquica: posibilidad de establecer relaciones y enlaces entre elementos.
- Granularidad : elementos, subelementos y atributos.
- Estructura modular que favorece la escalabilidad del sistema.

# NEEO Technical Guidelines.

## *Representación de los metadatos bibliográficos*

### *Solución en E-Archivo:*

- Se mantiene el esquema original de metadatos QDC, creando subcampos dentro de cada elemento de metadato, que denotan una entidad de información propia.
- Se define una lista de tipos de documentos en un xml, susceptibles de ser servidos al portal EO. Se establecen subcampos para cada tipo.

\* Software original by Benoit Pauwels (ULB)

# NEEO Technical Guidelines

## *Representación de los metadatos bibliográficos*

*Solución en E-Archivo. Artículos*

Metadato DC	Subcampos
title	a title
Contributor.author	a authorlastname   a authorfirstname   =[ID author]
Identifier.bibliographicCitation	a journal title   v volume   i issue   d date   p pages
publisher	a name   c location

# NEEO Technical Guidelines

## *Representación de los metadatos bibliográficos*

*Solución en E-Archivo. Capítulos libros*

Metadato	Subcampos
title	a chapter title
Contributor.author	a authorlastname   a authorfirstname   =[ID author]
Identifier.bibliographicCitation	a title book   h author of book   u publisher   d date   p pages   w series name   v series number
type	a chapter of book

# NEEO Technical Guidelines.

## *Ficheros enriquecidos*

- Formatos (MIME Media Types)
- Metadatos de descripción del fichero
  - Tipo (ej: mimeType="application/pdf")
  - Fecha de la última modificación
  - Descripción general del contenido (ej: Capítulo 1)
  - Versión (vocabulario controlado): preprint, postprint, publisher version
  - Acceso (vocabulario controlado): "acceso abierto", "acceso restringido"

# NEEO Technical Guidelines.

## *Ficheros enriquecidos*

*Solución en E-Archivo:* Se definen subcampos en la tabla de descripción del fichero

Subcampos	Valor
d fecha del depósito o de la última modificación	aaaa-mm-dd
s accesibilidad del fichero	INTERNET, INTRANET, INVISIBLE
t fecha de fin de embargo	aaaa-mm-dd
v versión del documento	PREPRINT, POSTPRINT, PUBPRINT



# NEEO Technical Guidelines.

## *Identificador de autor DAI (Digital Author Identifier)*

- Único (URI)
- Persistente (no puede cambiar)
- Cada repositorio elige su propio formato

Servicios ofrecidos a partir del DAI:

- Generación de listas bibliográficas por autor

*Solución en E-Archivo:*

dc.contributor.author

| agonzalo, | Jesús  
| =[jgonzalo]

# NEEO Technical Guidelines.

*Protocolo OAI-PMH: directrices específicas para el traslado de los objetos digitales*

- **Crosswalks** (*conversión de esquemas de metadatos*):
  - Desarrollo de un crosswalk que convierta los metadatos bibliográficos del esquema implementado en cada repositorio al esquema MODS-XML.
  - Desarrollo de un crosswalk que defina la representación del objeto digital en un documento DIDL-XML.

# NEEO Technical Guidelines.

*Protocolo OAI-PMH: directrices específicas para el traslado de los objetos digitales*

- **SETS** (*agrupación de items*)
  - EO recolectará sólo aquellos registros de metadatos que estén relacionados con su área temática de interés.
  - Para hacer posible la recolección selectiva, el proveedor de datos debe crear un “set” específico que agrupe los registros objeto de recolección.

# NEEO Technical Guidelines.

*Protocolo OAI-PMH: directrices específicas para el traslado de los objetos digitales*

- El portal EO recolectará registros de los proveedores de datos:
  - utilizando el '*metadataPrefix*': DIDL
  - usando una combinación de las peticiones OAI: '*listRecords*' y '*getRecord*'

# NEEO Technical Guidelines.

## *Protocolo OAI-PMH: directrices específicas para el traslado de los objetos digitales*

### *Soluciones en E-Archivo:*

- **Crosswalks:** Primero se convierten los metadatos QDC a MODS; después otro crosswalk genera el paquete DIDL que contiene además de los metadatos bibliográficos en MODS, el objeto digital completo, y es el que se ofrece a través de la petición OAI.
- **Sets:** Se crea el set “Economist Online” y se mapean los registros susceptibles de ser recolectados por EO.

# NEEO Technical Guidelines.

*Protocolo OAI-PMH: directrices específicas para el traslado de los objetos digitales*

- Petición:

[http://e-archivo.uc3m.es:8080/dspace-oai/request?metadataPrefix=didl&verb=ListRecords&set=hdl\\_10016\\_3012](http://e-archivo.uc3m.es:8080/dspace-oai/request?metadataPrefix=didl&verb=ListRecords&set=hdl_10016_3012)

- Respuesta: fichero xml con los registros.

<http://e-archivo.uc3m.es:8080/oai/request?verb=GetRecord&metadataPrefix=didl&identifier=oai:e-archivo.uc3m.es:10016/811>

# NEEO Technical Guidelines for the Exchange of Usage Metadata

## *¿Qué servicio queremos ofrecer?*

- Los datos típicos:
  - Datos de descarga / vista (item, comunidad, institución)
  - Bestsellers (item del IR, o del Departamento, o del Área, ha sido el más descargado)
  - Evolución en el uso del Repositorio
  - Rankings ....
- Servicios avanzados:
  - Creación de una red de publicaciones relacionadas
  - Recomendaciones por materias

# NEEO Technical Guidelines for the Exchange of Usage Metadata

- Control sobre co-autores → duplicidades (la misma publicación está disponible en múltiples repositorios institucionales o en los de las editoriales) → se consideran todos los accesos desde cualquiera de estos repositorios.

(Ej. EO, RePEc, Ideas, EconPapers)

- Estándares usado para la recolección :
  - Information Exchange (OpenURL ContextObject, OAI-PMH)
  - Metadata Format (SWUP Scholarly Works Usage Community Profile)



# NEEO Technical Guidelines for the Exchange of Usage Metadata

- *NEEO-SWUP ContextObject*

Atributo	Descripción
Identifier *	Identificación del evento
Timestamp *	Fecha del evento

Entidades	Descripción
Referent *	Publicación
ReferringEntity	Institución donde fue generado el evento
Requester	Recurso que generó el evento
ServiceType	Tipo de evento (view / download)
Resolver *	Recurso que ofrece el servicio solicitado
Referrer	Recurso desde el que se generó el evento

# NEEO Technical Guidelines for the Exchange of Usage Metadata

## *Otras cuestiones a tener en cuenta*

- Anonimato (MD5 encryption)
- Información geográfica (GeoIP API's y BD de MaxMind Inc.)
- Eliminación de “double clicks” (son filtrados por EO Gateway)
- Eliminación de “robots”(filtrado en los IR bajo las mismas especificaciones, lista de robots/spiders/linkcheckers/...)
- Registro mínimo (Identifier, Timestamp, Referent and Resolver)
- Recomendación NEEO (Identifier, Timestamp, Referent, ReferringEntity, Requester, Resolver and Referrer)

# NEEO Technical Guidelines for the Exchange of Usage Metadata

## @MIRE development

- Basado en Apache Solr
- Genera índices a partir de los logs de apache, tomcat o dspace
- Construye el XML correspondiente a la petición realizada por EO (según SWUP)

[http://e-archivo.uc3m.es/oai-  
neeo/request?verb=ListRecords&metadataPrefix=swup](http://e-archivo.uc3m.es/oai-<br/>neeo/request?verb=ListRecords&metadataPrefix=swup)

# Conclusiones

- El nivel de integración exigido por NEEO es novedoso y avanzado en relación con el estado actual de los repositorios en Europa. La tecnología y los estándares se encuentran en continua evolución.
- Dicha implementación requiere una alta cualificación profesional: expertos desarrolladores que conozcan las aplicaciones
- La UC3M ha requerido los servicios de la empresa @mire, la cuál reúne las condiciones arriba expresadas.

# 4as Jornadas Os-Repositorios

5 de marzo de 2010

Muchas gracias por vuestra atención



Universidad  
Carlos III de Madrid  
[www.uc3m.es](http://www.uc3m.es)