



Universidade do Minho
Serviços de Documentação

Interoperabilidade en repositorios digitales de acceso abierto

Eloy Rodrigues

eloy@sdum.uminho.pt





COLECTA DE METADATOS Y REDES DE REPOSITARIOS



Colecta de metadatos y redes de repositorios



- a) OAI-PMH
- b) Directrices DRIVER y SNRD
- c) Directrices OpenAIRE





OAI-PMH



Que es el OAI-PMH?



Universidade do Minho
Serviços de Documentação

The Open Archives Initiative Protocol for Metadata Harvesting

Una forma de interrogar a un archivo sobre su contenido.

Permite la creación de servicios de búsqueda y otras funcionalidades sobre los metadatos de varios archivos.

Versión estable (2.0) en 2002





OAI-PMH: dos componentes

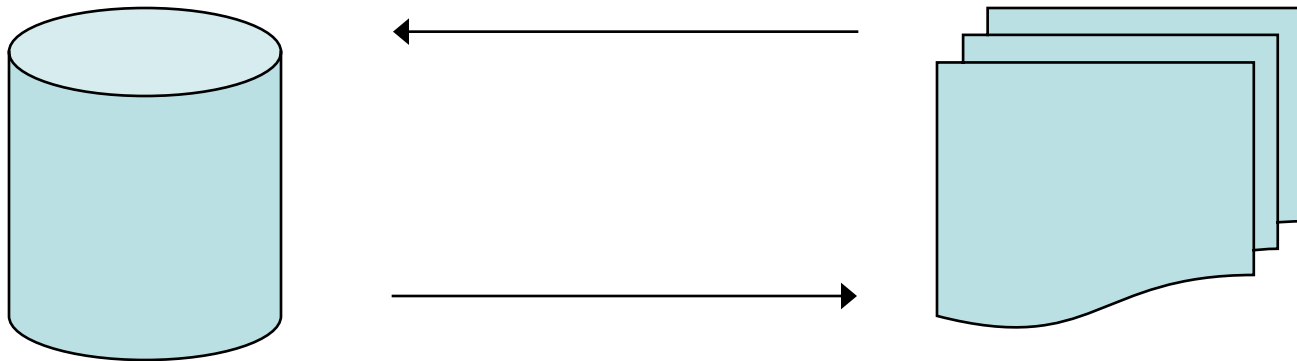
Data Providers – Mantienen un o mas repositorios (servidores Web) que suportan el OAI-PMH como método de exponer los metadatos

Service Providers – Envía solicitudes OAI-PMH a data providers y utiliza los metadatos colectados como base para construir servicios de valor añadido.





Modelo OAI-PMH



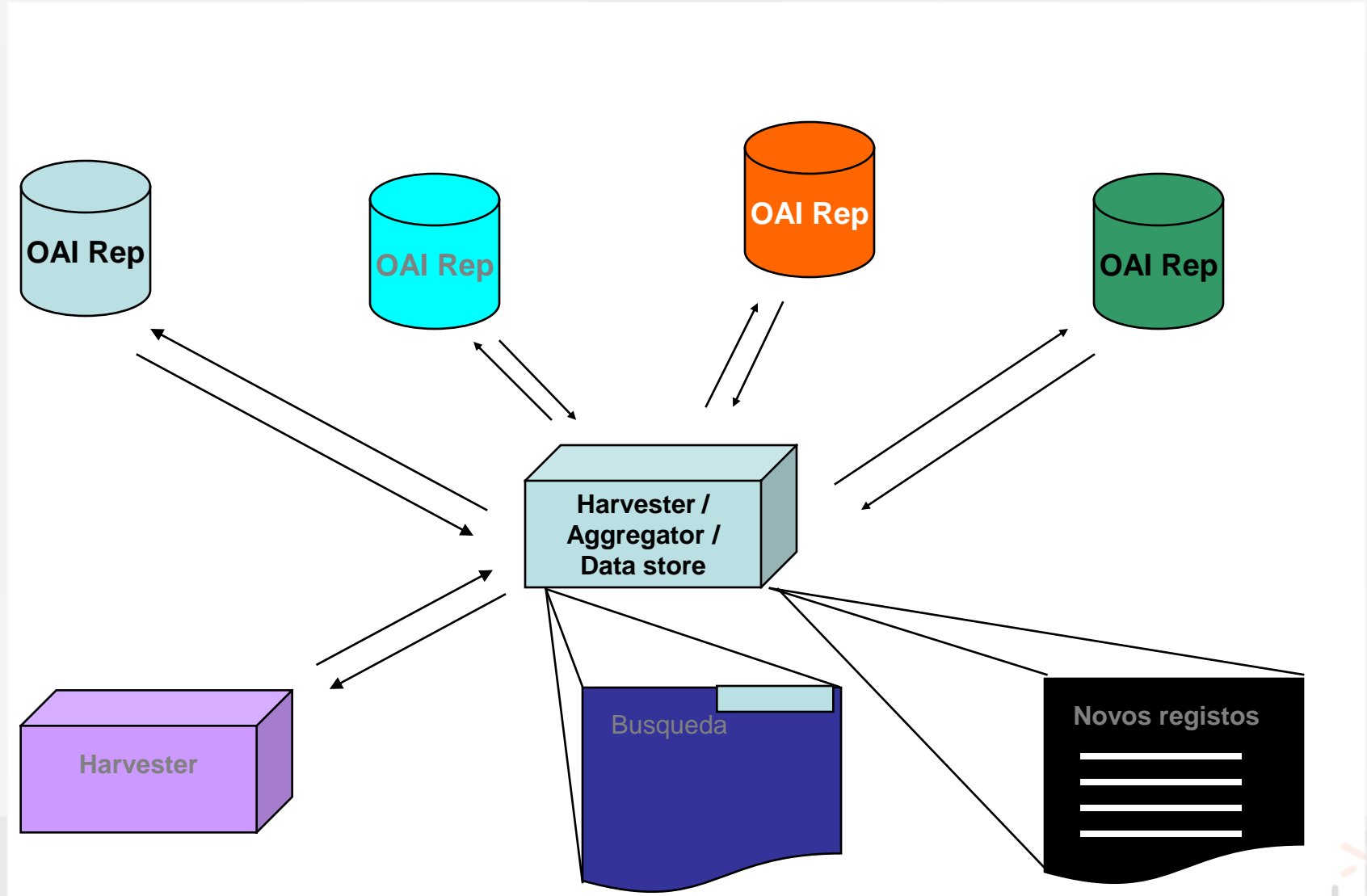
Data providers
exponen metadatos

Service providers
Recogen metadatos
y hacen algo útil
con ellos





Modelo OAI-PMH



Tecnologías



Universidade do Minho
Serviços de Documentação

Construido com tecnologías standard

Protocolo de transporte HTTP

Metadatos codificados en XML

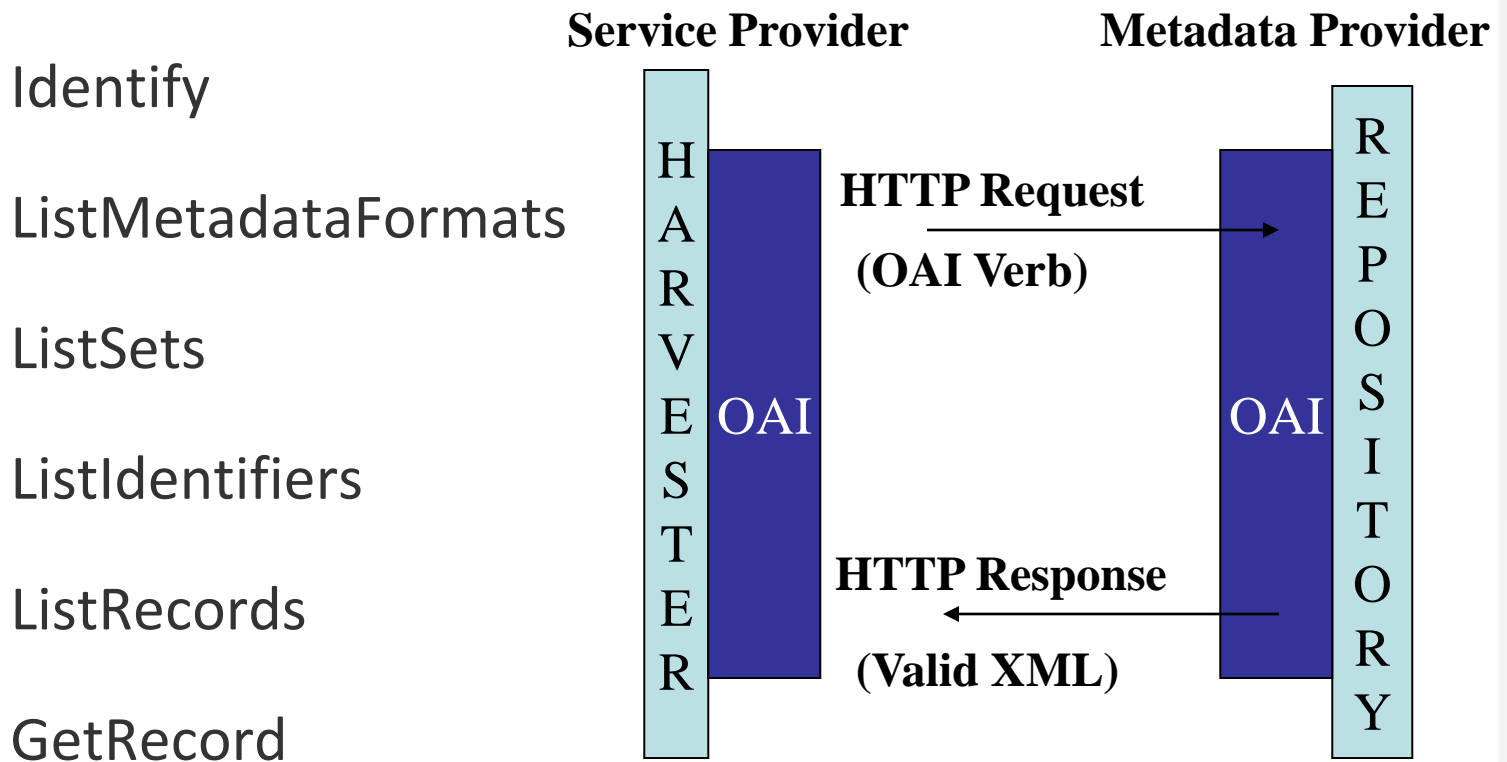
Formato de metadatos Dublin Core





Como Funciona?

OAI "VERBS"



Identificadores Únicos



Universidade do Minho
Serviços de Documentação

Cada registro tiene que poseer un identificador único

Identificadores tienen que ser URI's válidos

Cada identificador debe apuntar para un único registro y siempre para el mismo registro

No se puede reaprovechar identificadores de registros





Datestamps

Necesario en cada registro para suportar harvesting's incrementales

Debe ser actualizado en cada adición / alteración / remoción para garantizar que las alteraciones son propagadas correctamente

Pueden ser en la forma YYYY-MM-DD o YYYY-MM-DDThh:mm:ssZ (debe ser GMT timezone)





OAI-PMH: Resumen

Protocolo basado en HTTP

Argumentos de las solicitudes - GET o POST

Seis tipos de solicitudes

Respuestas codificadas en XML

Suporta cualquier formato de metadatos
(requiere: Dublin Core)

Data Providers y Service Providers



Dublin Core Metadata Element Set



Universidade do Minho
Serviços de Documentação

- Recomendaciones de Dublin Core Metadata Initiative (DCMI):
 - DCMES – Dublin Core Metadata Element Set
 - Un conjunto de 15 elementos de metadatos cuyo objetivo es permitir y facilitar la descubierta de recursos electrónicos. (Woodley, 2001)
 - Norma ISO, ANSI/NISO,
 - ISO Standard 15836-2003 (February 2003):
<http://www.niso.org/international/SC4/n515.pdf>
 - NISO Standard Z39.85-2007 (May 2007):
<http://www.niso.org/standards/resources/Z39-85-2007.pdf>
 - DCQ – Dublin Core Qualifiers
 - Refinamiento de elemento - especifican con mas detalle el significado de un elemento
 - Esquema de codificación – Un esquema que ayuda en la interpretación do valor de un elemento. Incluyen vocabularios controlados y notaciones formales o reglas de *parsing*. (Woodley, 2001)





Dublin Core - Elementos

Contenido	Propriedad Intelectual	Caso (Instance)
Title	Creator	Date
Subject	Contributor	Format
Description	Publisher	Identifier
Type	Rights	Language
Source		
Relation		
Coverage		





Elementos de refinamiento

- Otros términos asociados al elemento que especifican con mas precisión su significado.
- Por ejemplo, el elemento Date tiene como calificadores de refinamiento de elemento los términos Issued, Created, Available, Modified e Valid.





DIRECTRICES DRIVER





Directrices DRIVER

- Desarrolladas en el ámbito del proyecto DRIVER – Digital Repository Infrastructure Vision for European Research
- Versión actual 2.0 (noviembre de 2008)
- Traducidas en español, japonés e portugués

http://www.driver-support.eu/documents/DRIVER_2_1_Guidelines_Spanish.pdf



Las Directrices DRIVER en varios idiomas



Universidade do Minho
Serviços de Documentação



DRIVER Guidelines 2.0

Guidelines for content providers - Exposing textual resources with OAI-PMH

[November 2008]



[Guidelines for Repository Managers and Administrators on how to expose digital scientific resources using OAI-PMH and Dublin Core Metadata, creating interoperability by homogenising the repository output.]



DRIVER ガイドライン 2.0

コンテンツプロバイダのためのガイドライン
OAI-PMH を使用したテキストソースの公

[2008年11月]



[OAI-PMH と Dublin Core メタデータを使用してリポジトリ出力を均一化し、相互運用性を持たせたデジタル科学リソースの公開方法に関するリポジトリ運営管理者のためのガイドライン]



1/142

status: final 2008-11-01



Digital Repository Infrastructure Vision for European Research

Directrices para proveedores de contenido

Exposición de recursos textuales con el protocolo OAI-PMH

Aplicación piloto

En Español



Versión 1.0

Colaboradores

Martin Faljeu, Wolfram Horstmann, Friedrich Summann, Muriel Foulonneau, Ki Van Godtzenhoven, Patrick Hochstenbach, Paolo Manghi, Gill Hubbard



Digital Repository Infrastructure Vision for European Research

Directrizes para fornecedores de conteúdos

Exposição de recursos textuais com o protocolo OAI-PMH

Aplicação Piloto

Versão 1.0



Colaboradores

Martin Faljeu, Maurice Vanderfaesst, Wolfram Horstmann, Friedrich Summann, Muriel Foulonneau, Karen Van Godtzenhoven, Patrick Hochstenbach, Paolo Manghi, Gill Hubbard

Tradição
Serviços de Documentação da Universidade do Minho

Directrices DRIVER



Universidade do Minho
Serviços de Documentação

- Por qué es necesario directrices?
 - Para guiar los administradores de repositorios a garantizar interoperabilidad y compatibilidad con los servicios DRIVER;
 - Para que los desarrolladores de plataformas de repositorios incorporen nuevas funcionalidades en versiones futuras





- Recursos Textuales(Apartado A)
- Directrices para Metadatos (Apartado B)
- Directrices para OAI-PMH





Recursos Textuales (A)

- Comparar Apartado A con contenido del Repositorio
- Si los contenidos estan de acuerdo con los requisitos de las directrices, no hay que hacer nada...
- Si no, implementar un Set DRIVER



Recursos Textuales - **Obrigatório**



Universidade do Minho
Serviços de Documentação

- El repositorio contiene recursos textuales digitales
 - recurso textual = artículos científicos, tesis doctorales, documentos de trabajo, libros electrónicos y documentos similares resultado de la actividad de la investigación científica
- Los recursos textuales están en formatos ampliamente utilizados y extendidos (PDF, TXT, RTF, DOC, TeX, etc.)
- Los recursos textuales están en acceso abierto, están disponibles directamente en el repositorio para cualquier usuario del mundo, y sin ninguna restricción, como autorización o pago.



Recursos Textuales - **Obrigatório**



Universidade do Minho
Serviços de Documentação

- Los recursos textuales están descritos mediante registros de metadatos
- Los recursos textuales y de metadatos se vinculan entre sí de tal modo que un usuario final pueda acceder al recurso textual mediante el identificador (normalmente una URL) del registro de metadatos.
- La URL de un recurso codificada en el registro de metadatos siempre se puede localizar y nunca se cambia ni se reasigna.
- Un identificador único identifica el registro de metadatos y el recurso textual (no hay punteros a sistemas externos, como un sistema bibliotecario nacional o un editor)





Recursos Textuais - Recomendado

- Verificación transparente de la integridad de un recurso textual.
- Medidas de control de calidad (del contenido científico) de los recursos textuales expuestos para limitarlos a, por ejemplo, los recursos textuales incluidos en el informe científico anual (o equivalente).
- La URL de un recurso codificada en el registro de metadatos se basa en un esquema de identificadores persistentes, como DOI, URN, ARK, etc.





Passo 1 – Directrices Generales (A)

- Comparar Apartado A con contenido del Repositorio

- Si los contenidos están de acuerdo con los requisitos de las directrices, no hay que hacer nada...

- Si no, implementar un Set DRIVER

Si no



Passo 1 – Directrices Generales (A)



Universidade do Minho
Serviços de Documentação

- Comparar Apartado A con contenido del Repositorio
- Si los contenidos están de acuerdo con los requisitos de las directrices, no hay que hacer nada...
- Si no, implementar un Set DRIVER



Implementar un set DRIVER



Universidade do Minho
Serviços de Documentação

- Una forma de seleccionar y filtrar los registros relevantes para el set.

Quando se implementa um set DRIVER?

Quando el repositorio contiene documentos que no corresponden a la definición de “recurso textual”:

- contenidos que no sean literatura científica
- Contenidos con acceso restringido, embargado o controlado
- Contenidos en formatos no comunes



Directrices para metadatos



Universidade do Minho
Serviços de Documentação

- Las directrices han sido producidas para facilitar el **intercambio de metadatos** entre los proveedores de contenidos y los servicios (DRIVER o otros).
- Describen la conversión de los formatos internos para el estándar *Unqualified Dublin Core* para facilitar la colecta (*harvesting*).

▪



Directrices para metadatos-

Requisitos mínimos



- Los metadatos se estructuran según la norma Dublin Core no cualificado (ISO 15836:2003).
- Los elementos individuales de DC se utilizan según lo dispuesto las directrices del presente apéndice.
- Es obligatorio utilizar codificación Unicode.



Directrices DRIVER – Uso de los elementos



Universidade do Minho
Serviços de Documentação

El **uso de elementos** puede ser:

Obrigatório = el elemento siempre debe estar presente en el registro de metadatos;

Obrigatório si procede = si se puede obtener el elemento, debe agregarse al registro de metadatos;

Recomendado = se recomienda el uso del elemento

Optativo = el hecho de utilizar o no el elemento no tiene relevancia





Directrices DRIVER – Uso de los elementos

<i>Elemento básico</i>	<i>Estado</i>	<i>Esquemas de codificación</i>
Title	M	Ninguno, texto libre
Creator	M	Estilo bibliográfico APA como en una lista de referencia Sintaxis: apellidos, iniciales (nombre) [http://en.wikipedia.org/wiki/Apa_style#Reference_list]
Subject	MA	Se puede elegir cualquier palabra clave y clasificación (preferiblemente en inglés), que esté definido por un esquema URI (preferiblemente info:eu-repo/classification).
Description	MA	Ninguno, texto libre La práctica recomendada es incorporar un resumen en inglés. “Abstract” (Resumen) es la interpretación predeterminada al valor de <code>dc:description</code>
Publisher	R	Ninguno



Directrizes DRIVER – Uso de los elementos



Universidade do Minho
Serviços de Documentação

Contributor	O	Estilo bibliográfico APA como en una lista de referencia Sintaxis: apellidos, iniciales (nombre) [http://en.wikipedia.org/wiki/Apa_style#Reference_list]
Date	M	Fecha ISO 8601 W3C-DTF - “Published” (Publicado) es la interpretación predeterminada al valor de dc:date
Type	M	Se puede elegir cualquier tipo de publicación y tipo de versión (preferiblemente en inglés), que esté definido por un esquema URI (preferiblemente info:eu-repo/semantics).
Format	R	Lista registrada de IANA de tipos de medios de Internet (tipos MIME) - [http://www.iana.org/assignments/media-types/]
Identifier	M	Esquema URI, vínculo a identificador persistente (URN, handle, DOI), documento a texto completo o página de transición.
Source	O	Directrices de codificación de información de citas bibliográficas de metadatos Dublin Core [http://dublincore.org/documents/dc-citation-guidelines/] como en dcterms:bibliographicCitation
Language	R	ISO 639-3
Relation	O	Ninguno
Coverage	O	“Period” (Período) es la interpretación predeterminada al valor de dc:coverage Codificación: Período DCMI [http://dublincore.org/documents/2000/07/28/dcmi-period/] Para obtener más información sobre esquemas de codificación, véase el Capítulo 5 Uso de vocabularios y semánticas.
Rights	R	Ninguno
Audience	O	Ninguno “Education level” (Nivel educativo) es el valor predeterminado de dc:audience.





Directrizes DRIVER – Elemento DC_Type

El elemento ‘type’ de DC se utiliza con tres propósitos:

1. Obligatorio: Tipo de publicación (controlado): para indicar el tipo de publicación basado en el vocabulario controlado del tipo de publicaciones DRIVER.
2. Optativo: Tipo de publicación (libre): para indicar el tipo de publicación basado en un vocabulario del repositorio local.
3. Recomendado: Versión (controlado): para indicar el estado del proceso de publicación.





Directrizes DRIVER – Elemento DC_Type

- info:eu-repo/semantics/article
- info:eu-repo/semantics/bachelorThesis
- info:eu-repo/semantics/masterThesis
- info:eu-repo/semantics/doctoralThesis
- info:eu-repo/semantics/book
- info:eu-repo/semantics/bookPart
- info:eu-repo/semantics/review
- info:eu-repo/semantics/conferenceObject
- info:eu-repo/semantics/lecture
- info:eu-repo/semantics/workingPaper
- info:eu-repo/semantics/preprint
- info:eu-repo/semantics/report
- info:eu-repo/semantics/annotation
- info:eu-repo/semantics/contributionToPeriodical
- info:eu-repo/semantics/patent
- info:eu-repo/semantics/other

<https://issue.guidelines.driver.research-infrastructures.eu/wiki/UseOfVocabulariesAndSemantics#Publicationtype>





- Seguir la versión 2.0 del protocolo.
- Registros eliminados: se deben declarar. Utilizar la opción “transient” o “persistent” (conservar temporalmente o indefinidamente la información sobre el registro eliminado).
- ResumptionToken (testigos de reanudación) al menos conservados durante 24 h.
- Lotes de recolección entre 100 y 500 registros.
- Correo electrónico de administración disponible.
- Comunicar cambios en la URL base, sets, formatos de metadatos
- Set Driver
 - Solo si es necesario.
 - Agrupa los registros textuales y de acceso abierto.
 - setName: Open Access DRIVERset; setSpec: driver





DIRECTRICES SNRD



Directrices SNRD



Universidade do Minho
Serviços de Documentação

- Las Directrices SNRD se basan en las “Directrices DRIVER 2.0”
- Para ser compatible con las Directrices SNRD, todos los aspectos de las Directrices DRIVER son válidos, con algunas excepciones



Directrices SNRD
Directrices para proveedores de contenido del
Sistema Nacional de Repositorios Digitales
Ministerio de Ciencia, Tecnología e Innovación
Productiva

Buenos Aires, Abril de 2012





SET SNRD

Los registros que integren el set snrd deben cumplir con las siguientes reglas:

- Los registros deben respetar el esquema de metadatos propuestos por el SNRD y debe prestarse especial atención a:
 - Aportar una URI accesible (que enlace en forma directa al objeto digital descrito o a la ficha de acceso al mismo)
 - Incluir sólo valores de las listas controladas definidas en el esquema de metadatos (dc:type, dc:rights). (Ver el Anexo I).
 - Respetar los esquemas de codificación definidos en las directrices, en especial campos como fechas, derechos e idioma (dc:date, dc:Rights, dc:language).

- Los objetos descritos son de acceso abierto.



Directrices SNRD



Universidade do Minho
Serviços de Documentação

Título

Nombre	title
Definición	Nombre que se da al recurso. Normalmente, el título es un nombre por el cual el recurso es conocido formalmente.
Uso	Obligatorio
Instrucciones	<p>Transcribir el título del original, en orden y ortografía del recurso. Utilizar mayúsculas únicamente para la primer palabra del título y los nombres propios.</p> <p>Los subtítulos deben separarse del título mediante dos puntos, precedidos y seguidos por un espacio.</p> <p>Quando el recurso tenga el título en más de un idioma, colocar cada título en instancias separadas del elemento.</p>
Correspondencia Campos MARC	130, 210, 240, 242, 243, 245, 246, 730, 740
Esquema de codificación	Ninguno



Directrices SNRD



Universidade do Minho
Serviços de Documentação

Primera instancia	Segunda instancia (Términos acordados por el SNRD)	Tercera instancia (Versión permitida)	Descripción
article	<ul style="list-style-type: none">artículo	accepted / published /updated	Trabajo presentado para ser publicado en una revista científica
book	<ul style="list-style-type: none">libro	accepted / published /updated	Libro o publicación monográfica
bookPart	<ul style="list-style-type: none">parte de libro	accepted / published /updated	Parte o capítulo de libro
conferenceObject	<ul style="list-style-type: none">documento de conferencia	accepted / published /updated	Todo tipo de documento publicado en un congreso, conferencia, etc.
doctoralThesis	<ul style="list-style-type: none">tesis doctoral	accepted / published /updated	Nivel más alto de una tesis
masterThesis	<ul style="list-style-type: none">tesis de maestría	accepted / published /updated	Nivel intermedio de una tesis
bachelorThesis	<ul style="list-style-type: none">tesis de gradotrabajo final de grado	accepted / published /updated	Nivel más bajo de una tesis
patent	<ul style="list-style-type: none">patentemarcamodelo industrialmodelo de utilidaddocumento legal	draft / submitted / accepted / published /updated	Patente
review	<ul style="list-style-type: none">reseña artículorevisión literaria	accepted / published /updated	Recensión de libro o de artículo
workingPaper	<ul style="list-style-type: none">documento de trabajo	draft / submitted	Trabajo preliminar de carácter científico o técnico publicado en una institucional





DIRECTRICES OPENAIRE





Directrices OpenAIRE

“Las Directrices OpenAIRE 1.0, orientarán a los administradores de repositorios para definir e implementar sus políticas locales de administración de datos cumpliendo con las demandas de la Comisión Europea sobre Acceso Abierto. Además, cumplirán con los requisitos técnicos de la infraestructura de OpenAIRE, diseñada para soportar y monitorizar la implementación del piloto 7PM OA”

Las directrices OpenAIRE complementan y se construyen sobre las **Directrices DRIVER**





Directrices OpenAIRE

Para ser compatible con las OpenAIRE, todos los aspectos de las Directrices DRIVER son válidos, con las excepciones siguientes:

1. Recursos textuales

No es obligatorio Acceso Abierto en OpenAIRE. Los recursos pueden estar en acceso abierto, bajo embargo o tener acceso restringido

2. Uso de los elementos dc:relation (relación) y dc:rights (derechos)

Para cumplir las Directrices de OpenAIRE, el uso de los campos dc:relation y dc:rights es más específico que el expresado en las Directrices DRIVER 2.0





Directrices OpenAIRE

OpenAIRE use of oai_dc

The use of oai_dc is based on the usage by DRIVER, as expressed in the DRIVER Guidelines 2.0, pp. 52-82.

In three cases the use of the DC elements is specific to OpenAIRE, opposed to a more general use in DRIVER:

projectID is needed to connect project information to the publication in the OpenAIRE information space. It equals to the Grant Agreement number as found in all documentation/correspondence between the EC and the researcher/coordinator. - info:eu-repo/grantAgreement/EC/FP7

accessRights will define the type of access to the publication

When the value of accessRights is “embargoedAccess” **embargoEndDate** will define the date of the embargo period.

The (generic) namespace can be used to express project information from other funders than the EC.





Directrices OpenAIRE

Version 1.0 – Julio 2010

Version 1.1 – Noviembre 2010

Correcciones de nombres y referencias.
Inclusion de informacion de version y licencia.

Version 2.0 – Octubre 2012 –

Compatibilidad con agregadores (portales).

Extension del Namespace para identificacion de proyectos

[info:eu-repo/grantAgreement/Funder/FundingProgram/ProjectNumber/Jurisdiction/ProjectName/ProjectAcronym/](http://info.eu-repo/grantAgreement/Funder/FundingProgram/ProjectNumber/Jurisdiction/ProjectName/ProjectAcronym/)

Version 3.0 – Diciembre 2012

Cambio de nombre del OpenAIRE OAI set de **ec_fundedresources** para **openaire**.

Nuevos elementos “relation” para indicar identificadores externos, referencias y conexión de datasets





OpenAIRE

Directrices OpenAIRE



Universidade do Minho
Serviços de Documentação

História de una serie de directrices

- DRIVER Guidelines v2 (2008)
 - Normalización del uso de elementos DC y de del protocolo OAI-PMH
 - OAI-Set "driver" -> registros de publicaciones en acceso abierto
 - Introducción de vocabularios (uso del namespace info:eu-repo/semantics)
- OpenAIRE Guildelines v1 (2010)
 - OAI set "ec_fundedresources"
 - extensión de vocabularios (access rights, embargo period)
 - Esquema de identificación para proyectos (limitado a recursos financiados por EC)
- OpenAIRE Guidelines v2 (2012)
 - Soporte a agregadores OAI
 - Esquema de identificación para proyectos que se puede aplicar a cualquiera financiador, esquema o proyecto





Directrices OpenAIRE



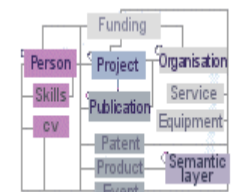
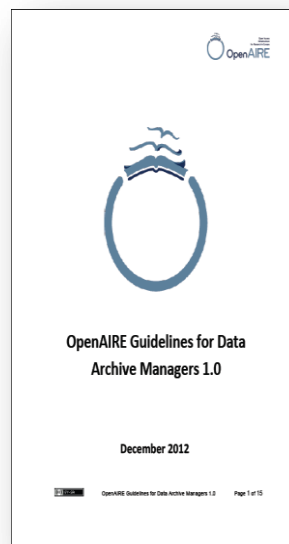
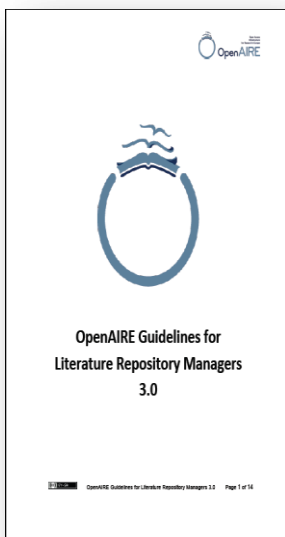
Universidade do Minho
Serviços de Documentação

OpenAIRE

1 Guidelines for Literature Repository managers version 3.0

2 Guidelines for Data Archive Managers

3 Guidelines for CRIS - CERIF-XML



Directrices OpenAIRE



Repositorios de publicaciones



Universidade do Minho
Serviços de Documentação

OpenAIRE

Lo novedoso...

- La versión 3.0 introduce tres alteraciones principales:
 - El nombre del set OpenAIRE OAI pasó de `ec_fundedresources` para **openaire**.
 - Nuevos elementos para indicar identificadores alternativos, relaciones con otras publicaciones (referencias) y relaciones con datos primarios (datasets).
 - Las recomendaciones sobre el uso de Dublin Core fueran heredadas y incorporadas de las directrices DRIVER.

