

INTRODUCCIÓN A OAI-PHM Y SU IMPLANTACIÓN EN EL PORTAL E-REVISTAS

Introducción

- Tecnociencia: <http://www.tecnociencia.es>

Punto de Encuentro entre el mundo de la investigación y el de la empresa que facilita y promueve el intercambio de conocimientos científicos y tecnológicos de forma rápida y eficiente.

- E-revistas: <http://www.tecnociencia.es/e-revistas>

Plataforma digital dentro del Portal Tecnociencia donde se recogen, seleccionan y alojan las revistas electrónicas españolas o latinoamericanas existentes o de nueva creación que cumplan una serie de requisitos de calidad rigurosos.

El acceso a las revistas será abierto.

Objetivos

- Difusión científica y tecnológica a todos los sectores de la sociedad.
- Establecer un sistema de calidad normalizado y validado de las revistas electrónicas.
- Repositorio de todas las revistas que quieran formar parte del portal
- Crear una interfaz al protocolo **OAI-PMH**: múltiples bases de datos y catálogos internacionales.
- Servicios: alertas, personalización, estadísticas, hospedaje de revistas.
- Apoyar open access.

OPEN ACCESS

Antecedentes:

- Internet cambia los sistemas de difusión del conocimiento.
- Científicos no tienen acceso libre a sus propios contenidos.
- Violaciones de copyright (software, repositorios, etc...).
- Barreras impuestas por monopolios editoriales.

Definición

“Disponibilidad gratuita en la Internet pública, para que cualquier usuario la pueda leer, descargar, copiar, distribuir, imprimir, con la posibilidad de buscar o enlazar todos los textos de estos artículos, recorrerlos para indexación exhaustiva, usarlos como datos para software, o utilizarlos para cualquiera otro propósito legal, sin barreras financieras, legales o técnicas, distintas de la fundamental de ganar acceso a la propia Internet.” (Budapest Open Access Initiative 2002)

Surgen iniciativas:

- BioMed Central (BMC, 2001)
- Budapest Open Access Initiative (BOAI, 2002)
- The Public Library of Science (PLOS, 2000)
- The Scholarly Publishing and Academic Resources Coalition (SPARC, 1998)
- Scientific Electronic Library Online (SciELO, 2002)
- Directory of Open Access Journals (DOAJ, 2003)

Objetivos:

- Aumentar la rapidez de distribución de los trabajos por medios electrónicos.
- Reducir los costes de las publicaciones científicas, y aumentar la visibilidad de los trabajos publicados.
- Reducir el fenómeno de "ciencia perdida", es decir aquella que no llega a tener un impacto relevante por su ausencia en las grandes bases de datos y repertorios científicos.
- Acceso a la información para toda la ciudadanía: expert-citizen (colectivos de afectados), activistas, comunidad software libre.

Distintas implantaciones:

- Acceso libre a la totalidad de la información.
- Acceso restringido a determinados sectores (universidades, países subdesarrollados, etc.).
- Acceso libre a los artículos pasados 6 meses de su publicación.
- Acceso libre a las revistas en formato digital y cobro por versión en papel.
- Etc.

Auto-Archivo

- Los propios autores depositan sus documentos en un repositorio de archivos electrónicos abierto.
- Mediante un interfaz web: metadatos y texto completo.
- EL autor es el que solicita a la editorial la inclusión del texto completo pudiendo corregir los datos necesarios en caso de no conseguir el permiso.
- Solución con perspectivas de crecimiento.

Debate:

- Viabilidad: libre distinto que gratis.
- Mismos problemas que en el mundo del software libre. Licencias GPL, copyleft, creative commons.
- Ventajas: difusión, costes bajos, comodidad del formato digital.
- Barreras: intereses económicos, ¿pérdidas de calidad? ¿prestigio?

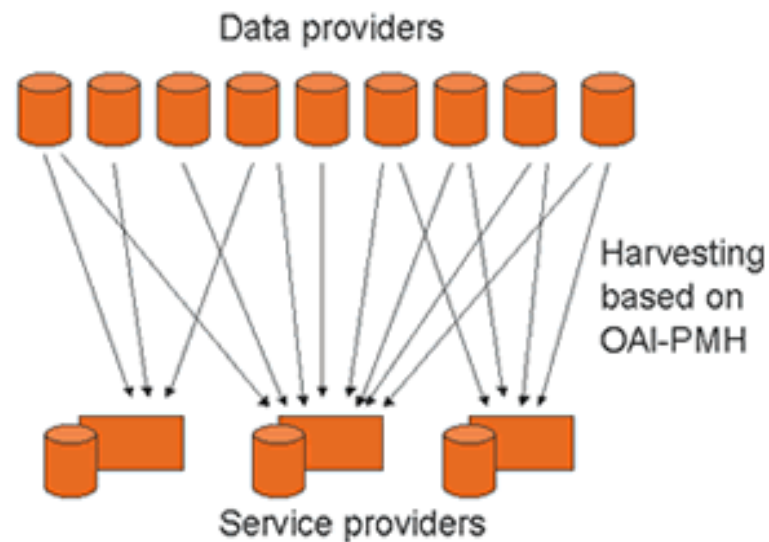
ARQUITECTURA Y PROTOCOLO OAI-PMH

Introducción

- OAI-PHM: Open Archives Initiative Protocol for Metadata Harvesting.
- **Requisitos del sistema:**
 - Integración en los sistemas de archivos abiertos (Open Archives)
 - Mantenimiento de un archivo distribuido
 - Implementación del protocolo OAI-PMH, para el intercambio de metadatos
 - Implementación de un “harvester” para la captura de metadatos
 - Creación de un Data Provider

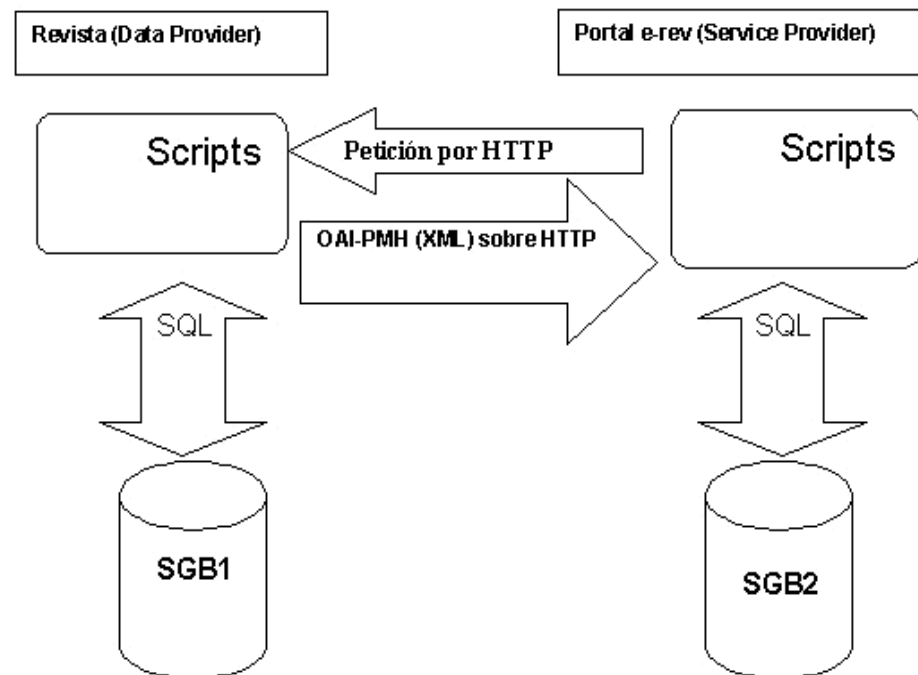
Arquitectura lógica general

- Arquitectura lógica del sistema basada en UPS (Universal Preprint Service) y su derivación propuesta por OAI (Open Archives Initiative)

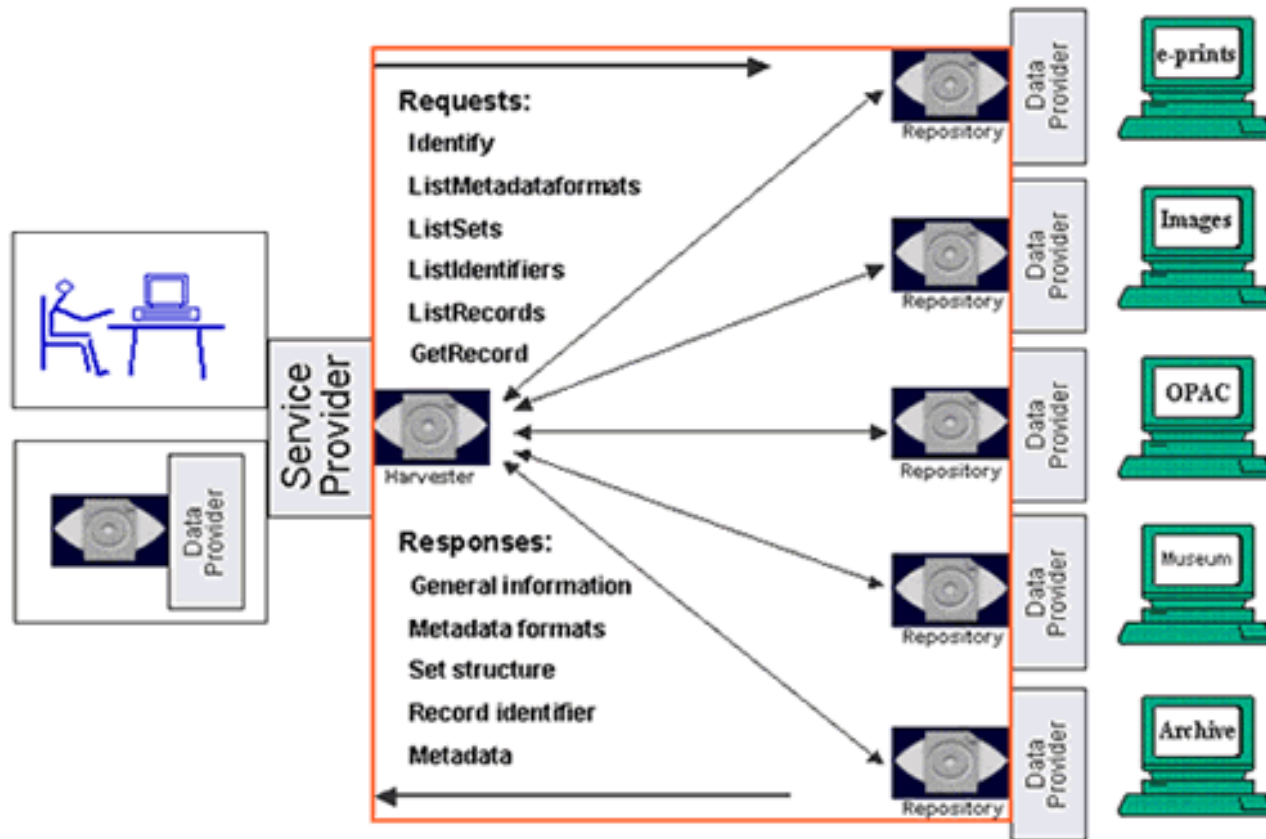


Descripción de los niveles del sistema

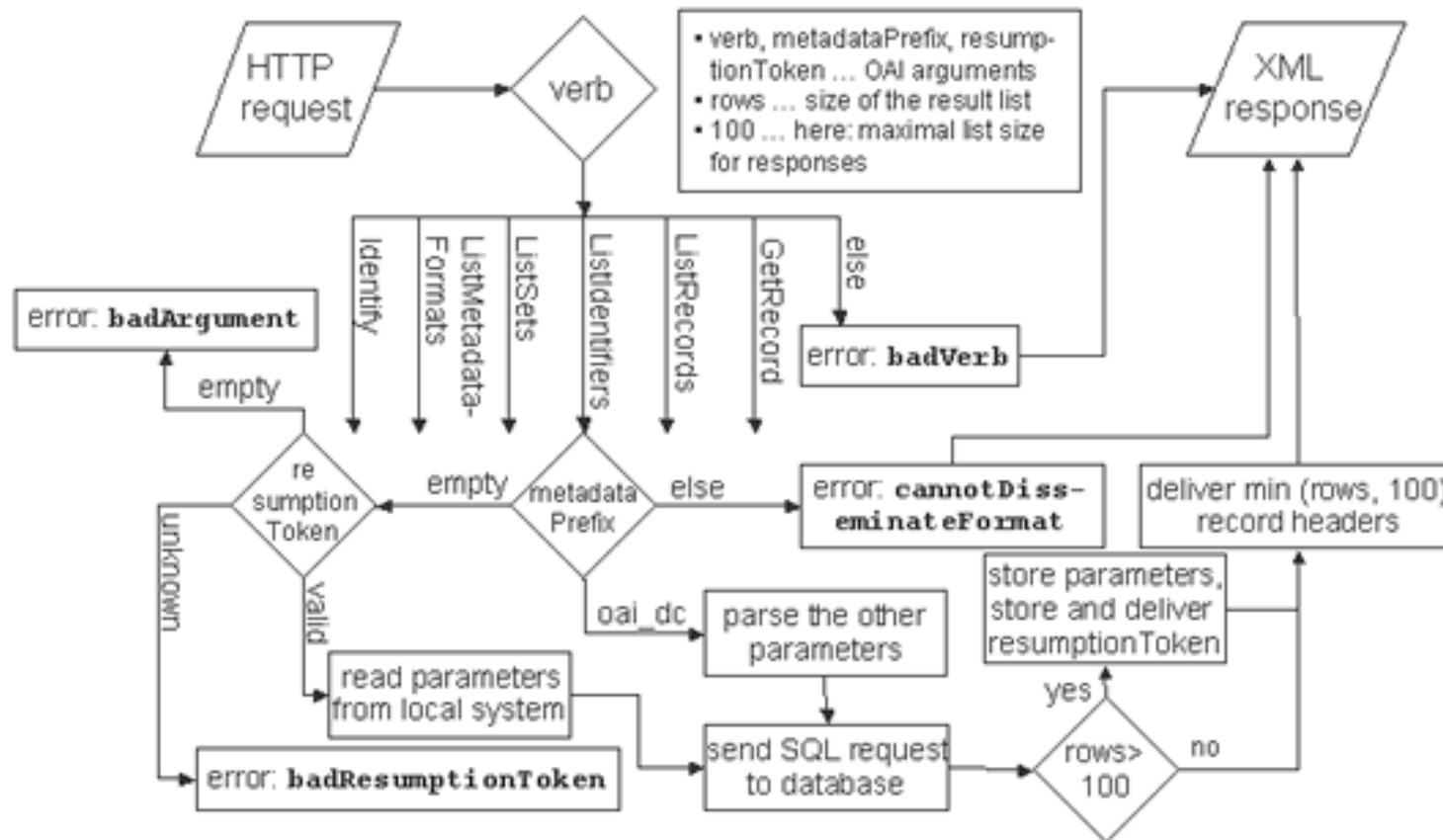
- Nivel de presentación
- Nivel de lógica de presentación
- Nivel de datos



Estructura y flujo del sistema (1)



Estructura y flujo del sistema (2)

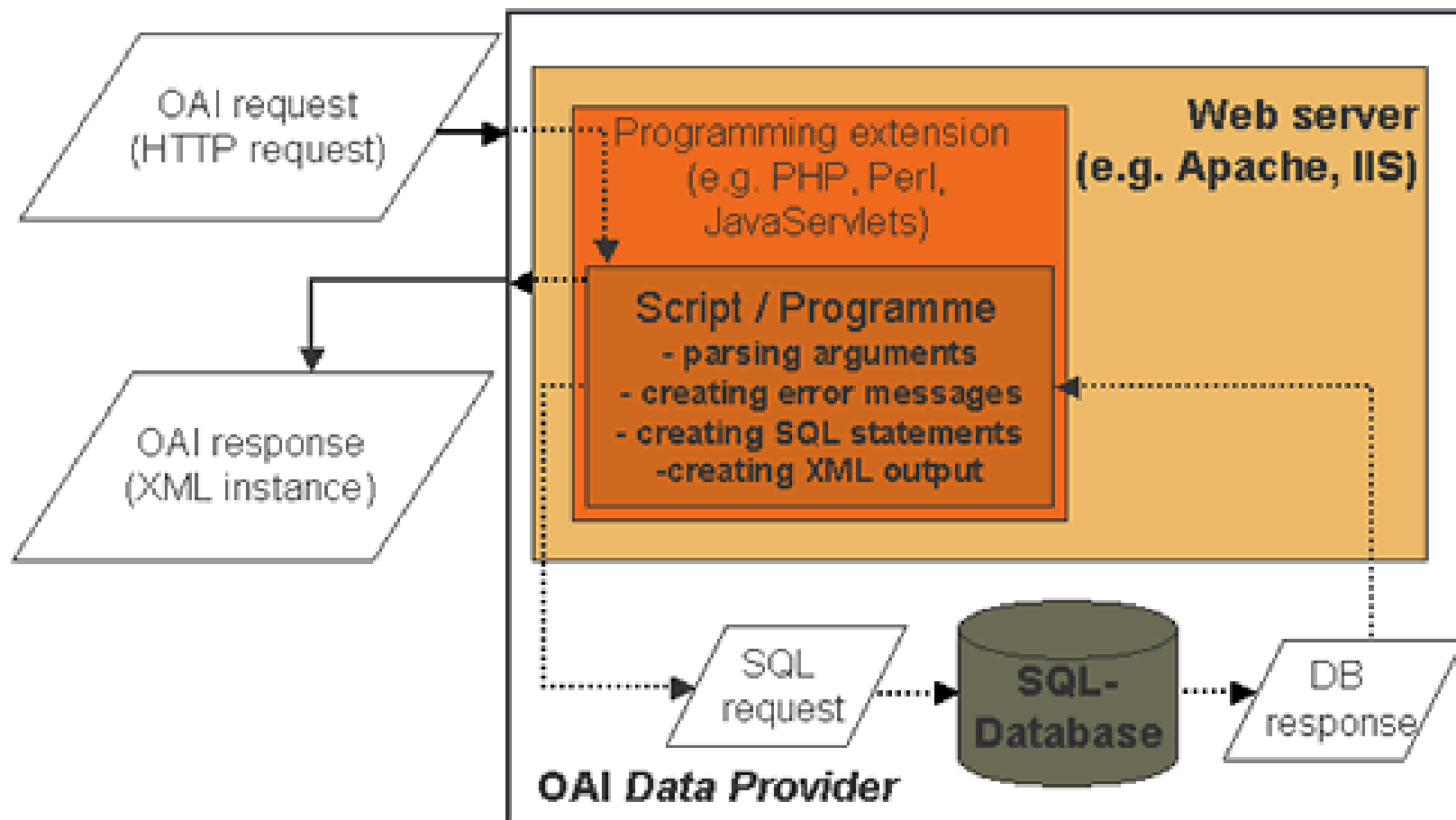


Requisitos del Data Provider

- Soporte de almacenamiento de metadatos
- Soporte Web
- Mantenimiento de una API para la implantación del protocolo
- URL
- URI
- Formato de los metadatos
- Datestamps
- Soporte lógico de una jerarquía de conjuntos (sets)
- Control de flujo (resumption token)

Componentes y arquitectura del data provider

- Analizador para las peticiones
- Generador de errores
- Sistema de consulta
- Generador de respuestas XML
- Control de flujo



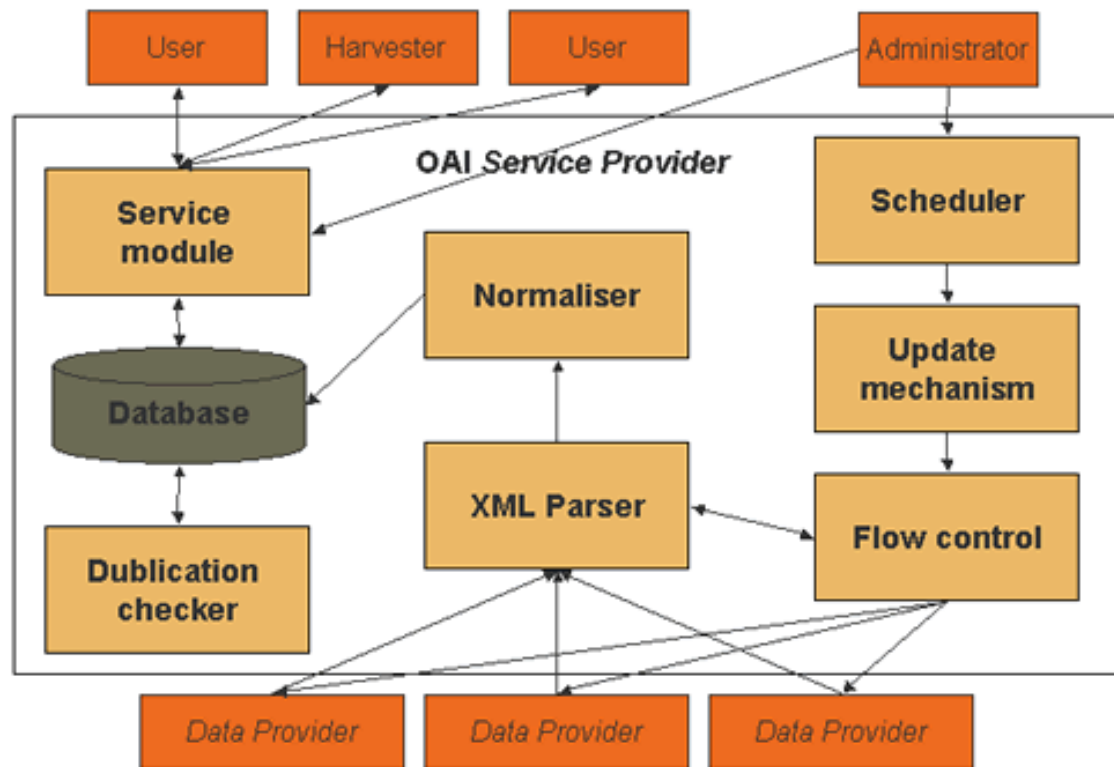
Requisitos del Service Provider

- Soporte de almacenamiento de metadatos
- Soporte Web
- Mantenimiento de un API para la implantación del protocolo

Componentes y arquitectura del Service Provider

- Gestión del archivo
- Creador de peticiones
- Temporizador (Scheduler)
- Control de flujo
- Mecanismo de actualización
- Analizador XML
- Herramienta de normalización
- Base de datos
- Comprobador de redundancias (duplication checker)
- Módulo de servicios

Componentes y arquitectura del Service Provider



Descripción del procolo OAI-PMH

- Basado en HTTP
- Las respuestas son en XML
- Peticiones por métodos GET y POST
- Peticiones con la forma clave=valor
- Existen 6 peticiones que se pueden realizar
 - Identify
 - GetRecord
 - ListIdentifiers
 - ListRecords
 - ListSets
 - ListMetadataFormats

DUBLIN CORE

Introducción

- Formato principal de intercambio de registros mediante el protocolo OIA-PMH.
- La Iniciativa de Metadatos Dublin Core DCMI es una organización dedicada a la promoción y difusión de normas interoperables sobre metadatos y el desarrollo de vocabularios especializados en metadatos para la descripción de recursos que permitan sistemas de recuperación más inteligentes.

Definición:

Esquema XML (XSD) con los siguientes campos:

- **Título** (Title)
- **Autor o Creador** (Creator)
- **Claves** (Subject)
- **Descripción** (Description)
- **Editor** (Publisher)
- **Colaborador** (Contributor)
- **Fecha** (Date)
- **Tipo del Recurso** (Type)
- **Formato** (Format)
- **Identificador del Recurso** (Identifier)
- **Fuente** (Source)
- **Idioma** (Language)
- **Relación** (Relation)
- **Cobertura** (Coverage)
- **Derechos** (Rights)

- Los elementos poseen nombres descriptivos que pretenden transmitir un significado semántico a los mismos.
- Cada elemento es opcional y puede repetirse. Además, los elementos pueden aparecer en cualquier orden.
- Clasificación de elementos:
 - Elementos relacionados principalmente con el contenido del recurso: Title, Subject, Description, Source, Language, Relation, Coverage.
 - Elementos relacionados principalmente con el recurso cuando es visto como una propiedad intelectual: Creator, Publisher, Contributor, Rights
 - Elementos relacionados principalmente con la instanciación del recurso: Date, Type, Format, Identifier.

XSD: <http://dublincore.org/schemas/xmls/simpledc20021212.xsd>

```
<element name="dc" type="oai_dc:oai_dcType"/>

<complexType name="oai_dcType">
  <choice minOccurs="0" maxOccurs="unbounded">
    <element ref="dc:title"/>
    <element ref="dc:creator"/>
    <element ref="dc:subject"/>
    <element ref="dc:description"/>
    <element ref="dc:publisher"/>
    <element ref="dc:contributor"/>
    <element ref="dc:date"/>
    <element ref="dc:type"/>
    <element ref="dc:format"/>
    <element ref="dc:identifier"/>
    <element ref="dc:source"/>
    <element ref="dc:language"/>
    <element ref="dc:relation"/>
    <element ref="dc:coverage"/>
    <element ref="dc:rights"/>
  </choice>
</complexType>
```

XSD:

Cada uno de los elementos es de tipo "String" con un atributo adicional de lenguaje:

```
<xs:complexType name="elementType">
  <xs:simpleContent>
    <xs:extension base="xs:string">
      <xs:attribute ref="xml:lang" use="optional"/>
    </xs:extension>
  </xs:simpleContent>
</xs:complexType>
```