



UNIVERSIDAD DE  
COSTA RICA

# Metadatos e Interoperabilidad: No toda publicación en línea es visible

**Guillermo Esteban Murillo Goussen**

[guillermo.murillogoussen@ucr.ac.cr](mailto:guillermo.murillogoussen@ucr.ac.cr)



# Agenda

- Metadatos
  - Definición
  - Beneficios
  - Meta-etiquetas
  - *Dublin Core*
  - Críticas
- Visibilidad
  - Imágenes, Flash
- Interoperabilidad
  - *OAI-PMH*



# METADATOS



# Metadatos - Definición

- Literalmente: “sobre datos”.
- Datos que describen otros datos.
- Descriptores de un “recurso”.
- Datos estructurados que describen información, contenido, calidad, condición y demás características de los datos.



# Metadatos – Informática

- Etiquetas que constituyen un puente sobre el “intervalo semántico”.
- Información no relevante para el usuario final pero sí de suma importancia para el sistema que maneja los datos.



# Metadatos - Beneficios

- Organizan y contribuyen con el mantenimiento de la colección - acervo de información –repositorio de datos.
- Utilizar información adicional los resultados del buscador resultan más precisos y el usuario se ahorra filtraciones manuales complementarias.
- Proporcionan información útil a herramientas de interpretación y procesamiento automático de datos.



# Meta-etiquetas HTML

- Etiquetas con información sobre el propio documento web: autor, editor, codificación, etc.
- La sintaxis del elemento META es la siguiente:
- `<meta name="nombre-propiedad" content="valor-propiedad" scheme="esquema-datos" lang="idioma" dir="dirección-del-texto">`
- `<meta http-equiv="nombre-encabezado-http" content="valor-encabezado">`



## Meta-etiquetas HTML (cont.)

- Metadatos: *Description, Keywords, Author, Copyright, Robots, Dublin Core*
- Los elementos *META* forman parte del contenido del elemento *HEAD* del documento.

```
<html xmlns=http://www.w3.org/1999/xhtml>
```

```
<head>
```

```
<meta name="Author" content="Pedro Villar Martos" lang="es">
```

```
<meta http-equiv="Content-Type" content="text/html;  
charset=ISO-8859-1">
```

```
</head>
```





# Meta-etiquetas *Dublin Core*

```
<meta name="DC.Title" content="HTML con Clase - Metadatos">  
<meta name="DC.Creator" content="Juan Ramón Pozo">  
<meta name="DC.Subject" content="Metadatos">  
<meta name="DC.Description" content="Los elementos META de HTML:  
tipos, funciones, etc.">  
<meta name="DC.Publisher" content="ConClase.net">  
<meta name="DC.Date" content="2001-06-16">  
<meta name="DC.Type" scheme="DCMIType" content="Text">  
<meta name="DC.Format.Medium" content="text/html">  
<meta name="DC.Relation.isPartOf" content="http://www.conclase.net/">  
<meta name="DC.Identifier"  
content="http://html.conclase.net/articulos/metadatos">  
<meta name="DC.Language" content="es">  
<meta name="DC.Rights" content="(c) 2001 by Juan R. Pozo. All rights  
reserved.">
```



# Dublin Core

- Sistema de 15 definiciones semánticas descriptivas.
- Estas definiciones:
  - Son opcionales.
  - Se pueden repetir.
  - Pueden aparecer en cualquier orden.
- Diseñado para proporcionar un vocabulario normalizado de características, y así describir cualquier recurso, sin que importe su formato de origen, el área de especialización o el origen cultural.



# Dublin Core - Elementos

- Contenido:
  - Título: el nombre dado a un recurso, habitualmente por el autor. Etiqueta: DC.Title
  - Claves: los tópicos del recurso. Expresa las claves o frases que describen el título o el contenido del recurso. Etiqueta: DC.Subject
  - Descripción: una descripción textual del recurso. Etiqueta: DC.Description



## Dublin Core – Elementos (cont.)

- Fuente: identifica unívocamente un trabajo a partir del cual proviene el recurso actual. Etiqueta: DC.Source
- Lengua: lenguaje del contenido intelectual del recurso. Etiqueta: DC.Language
- Relación: es un identificador de un segundo recurso y su relación con el recurso actual. Etiqueta: DC.Relation
- Cobertura: es la característica de cobertura espacial y/o temporal del contenido intelectual del recurso. Etiqueta: DC.Coverage



# Dublin Core – Elementos (cont.)

- Propiedad Intelectual:
  - Autor o Creador: la persona o organización responsable de la creación del contenido intelectual del recurso. Etiqueta: DC.Creator
  - Editor: la entidad responsable de hacer que el recurso se encuentre disponible en la red en su formato actual. Etiqueta: DC.Publisher
  - Otros Colaboradores: una persona u organización que haya tenido una contribución intelectual significativa, por ejemplo: un ilustrador o un traductor. Etiqueta: DC.Contributor



## Dublin Core – Elementos (cont.)

- Derechos: son una referencia a derechos de autor, para un servicio de gestión de derechos o para un servicio que dará información sobre términos y condiciones de acceso a un recurso. Etiqueta: DC.Rights
  
- Instanciación:
  - Fecha: una fecha en la cual el recurso se puso a disposición del usuario en su forma actual. Etiqueta: DC.Date
  
  - Tipo del Recurso: la categoría del recurso. Por ejemplo, página personal, novela, poema, diccionario, etc. Etiqueta: DC.Type



## Dublin Core – Elementos (cont.)

- Formato: es el formato de datos de un recurso, usado para identificar el software y, posiblemente, el hardware que se necesitaría para mostrar el recurso. Etiqueta: DC.Format
- Identificador del Recurso: secuencia de caracteres utilizados para identificar unívocamente un recurso. Ejemplos para recursos en línea pueden ser URLs, URIs, ISBN, ISSN. Etiqueta: DC.Identifier



# Críticas contra los Metadatos

- Son costosos y necesitan demasiado tiempo.
- Las empresas no van a producir metadatos porque no hay demanda y los usuarios no van a invertir tanto tiempo.
- Son demasiado complicados. La gente no acepta los estándares porque no los comprende y no quiere aprenderlos.





# VISIBILIDAD



# Visibilidad - Definición

- Determina qué tan fácil resulta para sus clientes potenciales, ubicar de entre las diferentes alternativas que se ofrecen, los productos o servicios que ellos requieren.
- Los buscadores tienen más de un billón de páginas indexadas y de acceso libre. Sin embargo solamente el 10% de la información es accesible al público en general.



# Visibilidad – Imágenes

- Tendemos a minimizar la importancia de las imágenes en el diseño general del *website*.
- Al incluir imágenes en una página web, debemos incluirle etiquetas (metadatos) como corresponde: nombre, descripción, texto alternativo.
- Cuidado con los PDFs de “escaneos”.



# Visibilidad – Flash

- Una página web hecha totalmente en Flash no se puede posicionar igual que una con texto y estructura HTML
- Con el Flash, se puede (y se suele) tener páginas distintas con contenido distinto pero sin *URLs* distintas.
- Texto sin estructura.



## Visibilidad – Flash (cont.)

- Un sitio por completo en Flash, tiene pocos elementos para optimizar las consultas. Resulta necesario poseer una doble versión Flash/HTML.
- Sólo se indexa contenido textual. Si se trata de imágenes o botones, sólo se descifran las descripciones textuales.
- No se indexan los FLV (formato de flash para video) porque los archivos no contienen elementos textuales.



# INTEROPERABILIDAD



# Interoperabilidad - Definición

- Habilidad de dos o más sistemas o componentes para intercambiar información y utilizar la información intercambiada. (IEEE)
- Habilidad que tiene un sistema o producto para trabajar con otros sistemas o productos sin un esfuerzo especial por parte del cliente.



## Interoperabilidad – Definición (cont.)

- Capacidad de sistemas con diversas plataformas de hardware y software, estructuras de datos e interfaces, para intercambiar datos con la pérdida mínima de contenido y funcionalidad.
- Siendo estos sistemas heterogéneos, distribuidos, geográficamente distantes,
- Mediante una interconexión libre, automática y transparente, sin dejar de utilizar en ningún momento la interfaz del sistema propia





## Interoperabilidad – Definición (cont.)

- Es un concepto clave al hablar de esquemas de metadatos y la necesidad de compatibilizar todos ellos.
- La interoperabilidad entre distintos esquemas de metadatos requiere estandarización. Un ejemplo de dicha normalización es el protocolo OAI-PMH.



# Protocolo OAI – PMH

- Protocolo establecido para facilitar el intercambio de datos entre distintos servidores.
- Open Archives Initiative Protocol for Metadata Harvesting.
- Interfaz que permite el acceso a los metadatos de contenidos o recursos provenientes de distintas fuentes, plataformas y repositorios.



## Protocolo OAI – PMH (cont.)

- Utilización de estándares:
  - Transporte y Comunicación TCP/IP y Z39.50,
  - Formatos de documentos PDF y HTML,
  - Metadatos XML, MARC 21, Dublin Core, MODS, METS,
  - Enlace de citas CROSSREF, DOI



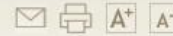
## Protocolo OAI – PMH (cont.)

- Participantes:
  - Proveedor de datos: colección de documentos que proporcionan metadatos relacionados.
  - Proveedor de servicios: solicita, recopila los metadatos de varias colecciones y proporciona búsquedas e identificación, filtrado.



## OAI – PMH (cont.)

- Peticiones:
  - GetRecord: recupera un registro concreto.
  - Identify: recupera información sobre el servidor.
  - ListIdentifiers: recupera los encabezados de los registros.
  - ListRecords: recupera los registros completos.
  - ListSets: recupera un conjunto de registros.
  - ListMetadataFormats: devuelve la lista de formatos que utiliza el servidor.



\* Búsqueda libre en los Portales . . . . .

Búsqueda avanzada

Portales

Portal
  Nombre Portal
  Revistas cosechadas
  Artículos cosechados

Portal	Nombre Portal	Revistas cosechadas	Artículos cosechados	Índice de revistas
	DIALNET	n.d.	404,242	<a href="#">Índice de revistas</a>
	e-revistas	607	142,381	<a href="#">Índice de revistas</a>
	Latin American Journals Online (LAMJOL)	26	404	<a href="#">Índice de revistas</a>
	Pepsic. Revistas electrónicas en Psicología	91	10,354	<a href="#">Índice de revistas</a>

Total de artículos indexados:

1,222,145



Noticias y Eventos



PPL en fotos



Universidad Nacional  
Autónoma de México



Red de  
Acervos  
Digitales

[Inicio](#) | [Acerca](#) | [FAQ's](#) | [¿Cómo participar?](#) | [¿Quiénes somos?](#) | [Ayuda](#)

i

**El Repositorio Institucional RAD-UNAM cuenta actualmente con 54,549 objetos digitales en 9 Repositorios Universitarios.**

### Repositorios Universitarios

- Coordinación de Difusión
- » Cultural. Reflexion de los Futuros de Mexico
- » CUAED
- » DG TIC
- » Facultad de Ciencias
- » Facultad de Filosofía y Letras
- Instituto de
- » Investigaciones Económicas
- Instituto de
- » Investigaciones Sociales
- Repositorio Universitario
- » CELE UNAM
- RU del Instituto de
- » Biología

## Repositorio Institucional RAD-UNAM

Buscador para la consulta y recuperación de recursos en formato digital producto de la actividad académica, depositados en Repositorios Universitarios digitales (RUs) de dependencias y entidades adscritas al proyecto de RAD-UNAM.

- Textos
- Imágenes
- Audios
- Videos
- Presentaciones
- Objetos de aprendizaje
- Otros

Buscar

Búsqueda avanzada

### Buscar por índices:

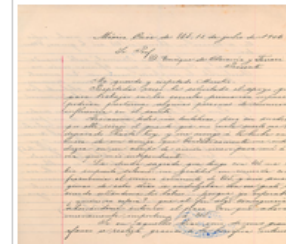
#### REPOSITORIO

- ▶ [Coordinación de Difusión Cultural. Reflexion de los Futuros de Mexico](#)
- ▶ [CUAED](#)
- ▶ [DG TIC](#)
- ▶ [Facultad de Ciencias](#)
- ▶ [Facultad de Filosofía y Letras](#)
- ▶ [Instituto de Investigaciones Económicas](#)
- ▶ [Instituto de Investigaciones Sociales](#)
- ▶ [Repositorio Universitario CELE UNAM](#)
- ▶ [RU del Instituto de Biología](#)

#### COLECCIÓN

- ▶ [00 Conferencias, cursos, cátedras, homenajes, ponencias \(lista por fecha\)](#)
- ▶ [01. ASV - Conferencias. Ética y política \(2003\)](#)
- ▶ [01. La Experiencia Literaria. 1993. Núm. 1](#)
- ▶ [Ver más](#)

### Colecciones destacadas





UNIVERSIDAD DE  
COSTA RICA

Muchas Gracias