



# Generación y Extracción Automática de Metadatos:

DCS

## El Proyecto DCS

---

Presentado por: Ernesto Giralt Hernández.

Madrid, 15 de Abril de 2005

**Sand!**

[www.esand.net](http://www.esand.net)

# Agenda

- Definición y objetivos
- Arquitectura y servicios
- Aplicaciones
- Futuro

# Definición y objetivos

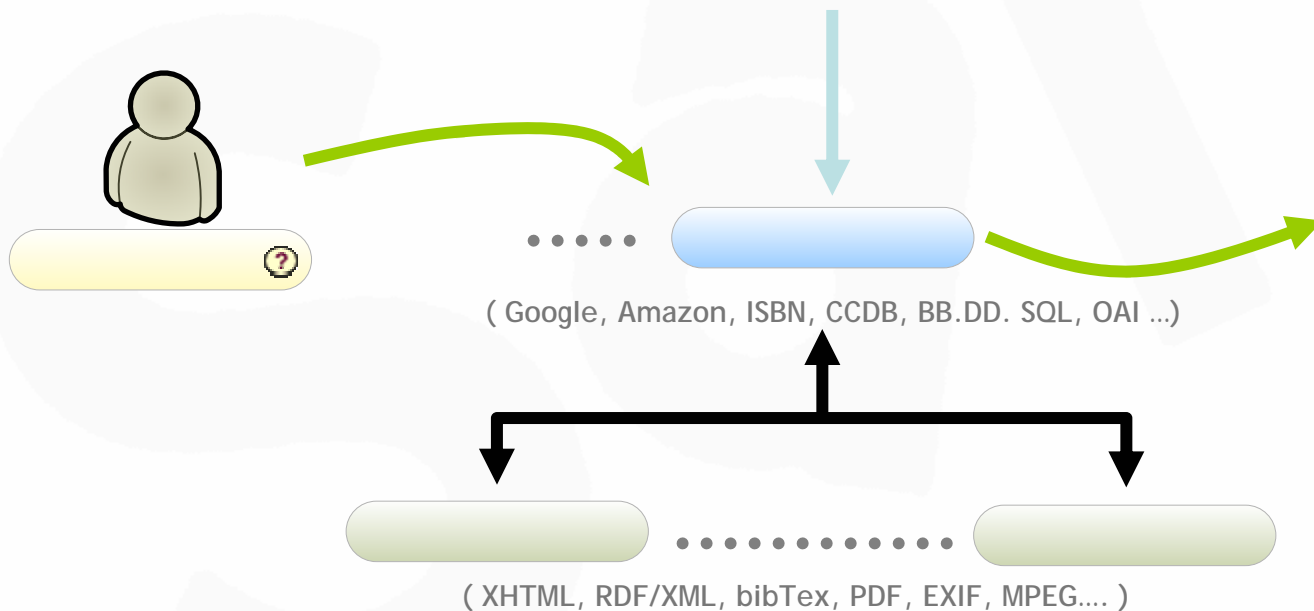
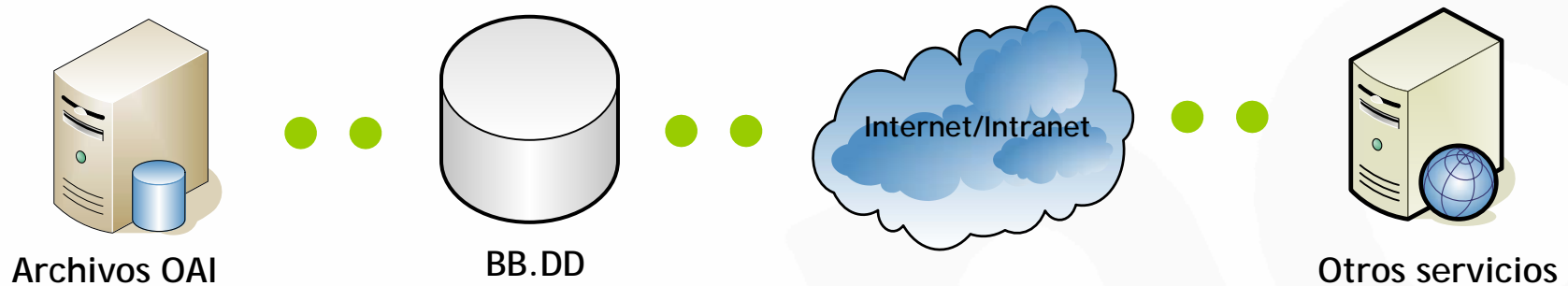
- Dublin Core Services (DCS) es un conjunto de servicios y aplicaciones para la extracción y generación automática de metadatos



## Capacidades

- Genera/extrae metadatos de más de 10 tipos de formatos digitales (XML, bibTex, XHTML, PNG, entre otros)
- Procesa contenido en 11 idiomas diferentes
- Diccionario de 5300 palabras vacías/términos no preferidos
- Reconoce más de 70 variantes de elementos de metadatos usados en los documentos HTML/XHTML

# Arquitectura y servicios



Registros Dublin Core (XML,RDF,XHTML)

<b>Title</b>	XML.com: Dublin Core in the
<b>Creator</b>	Dale Dougherty
<b>Date</b>	Oct. 25, 2000
<b>Description</b>	The recent Dublin Core Met- and to explore its relationsh
<b>Subject</b>	xml , xml discussion stylus , core , web services , web , studio world , world web co guidelines , resource framev description framework , xslt standards , online computer O'Reilly, OREILLY, Metadata
<b>Format</b>	text/html
<b>Identifier</b>	<a href="http://www.xml.com/pub/a/">http://www.xml.com/pub/a/</a>

## ¿Cómo se extraen y generan los metadatos?

- Métodos de análisis de contenido (TF/IDF)
- Hojas de estilo (XSL)
- Expresiones regulares
- Traducción/filtrado directo de formatos
- Listas de términos y palabras vacías

Sitio web: [www.describethis.com](http://www.describethis.com)

The screenshot shows the DescribeThis website in a Mozilla Firefox browser window. The browser's address bar displays the URL <http://www.describethis.com/>. The website header features the DCS logo and the text "Metadata Services and Tools". A search bar with the placeholder text "describe this" and a "Go" button is prominently displayed. Below the search bar, a tip reads: "Tip: Enter any query text, or select the prefix/operator from the combo to improve your query text, and get assisted instantly." The main content area is divided into several sections: "About DCS" which explains the service's purpose; "DCS At work!" which lists various "RSS Samples", "Image Analysis Samples", "Web Sites Samples", "ISBN analysis (using Amazon's XML public services)", and a "Keywords sample". A highlighted box mentions the "DC-2005: International Conference on Dublin Core and Metadata Applications" held in Madrid, Spain, from September 12-15, 2005. At the bottom of the page, there are links for "Privacy policy" and "Terms of use", along with logos for "Dublin Core Used Here", "Creative Commons Attribution-NonCommercial-ShareAlike license", and "W3C CSS".

## Objetivos de [www.describethis.com](http://www.describethis.com)

- Apoyar el desarrollo y extensión de la aplicación del formato Dublin Core.
- Que el sitio sea usado como herramienta de apoyo a la catalogación de recursos digitales
- Ofrecer los servicios automáticos de DCS a diseñadores de bases de datos bibliográficas, de contenido y especialistas de la información en general.



## Servicios Web de DCS

- Definición del servicio en:

<http://www.describethis.com/ws/dcsmethods.asmx?WSDL>

### Fragmento de la respuesta a una consulta

```
<?xml version="1.0" encoding="utf-8" ?>
...
<verbadcs:record verbadcs:source="http://dublincore.org/" >
  <dc:identifier xsi:type="dcterms:URI" verbadcs:generated="True">http://dublincore.org/</dc:identifier>
  <dc:format xsi:type="dcterms:IMT" verbadcs:generated="True">text/html</dc:format>
  <dc:title>Dublin Core Metadata Initiative (DCMI)</dc:title>
  <dc:title>Dublin Core Metadata Initiative (DCMI) Home Page</dc:title>
  <dc:description>The Dublin Core Metadata Initiative is an open forum engaged in the development of interoperable
online metadata standards that support a broad range of purposes and business models. DCMI's activities include
consensus-driven working groups, global conferences and workshops, standards liaison, and educational efforts to
promote widespread acceptance of metadata standards and practices.</dc:description>
  <dc:date>2005-03-29</dc:date>
  <dc:format>text/html</dc:format>
...
```

**Integración de los servicios con otras aplicaciones de gestión de datos y sistemas operativos para ampliar el alcance y uso de DCS y del propio formato Dublin Core**

**Adopción de RDF/XML como lenguaje base para todos los procesos de análisis y generación de metadatos**

Transformación de los actuales servicios en una plataforma de gestión capaz de distribuir servicios más integrales y potentes de generación y extracción de metadatos

# Más información

Escribir a:

[ernesto.giralt@esand.net](mailto:ernesto.giralt@esand.net)

[support@describethis.com](mailto:support@describethis.com)

Visitar:

[www.describethis.com](http://www.describethis.com)

[www.esand.net](http://www.esand.net)

DCS