

JIIDE
2012

Madrid, del 17 al 19 de octubre

Análisis de la visibilidad global de los publicadores de los recursos geográficos estandarizados

Aneta J. Florczyk, **F. Javier López-Pellicer**, Juan Valiño-García, Javier Nogueras-Iso,
F. Javier Zarazaga-Soria,

Madrid, 18 de octubre de 2012

Índice

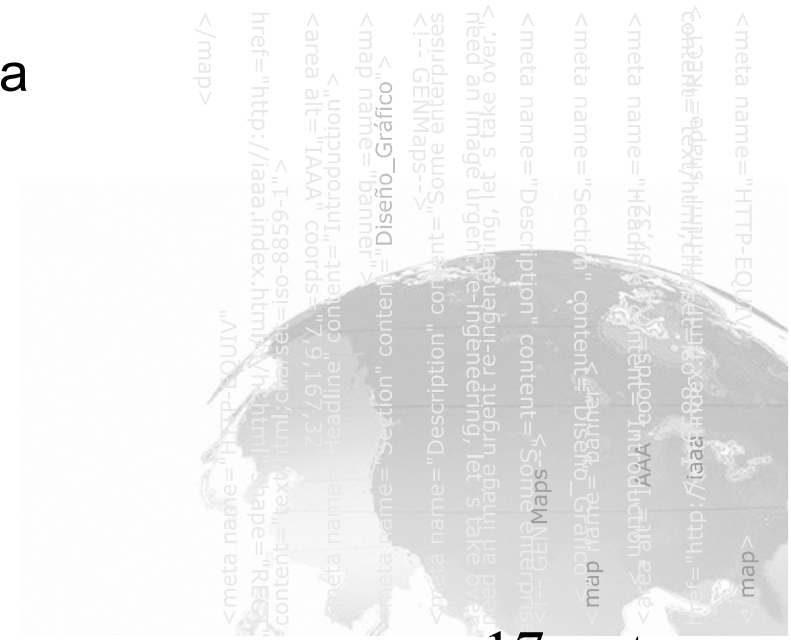
□ Introducción

□ La Web y visibilidad de las IDEs

□ Análisis de la visibilidad

- Muestreo
- Caracterización de la muestra
- Evaluación y resultados

□ Conclusiones



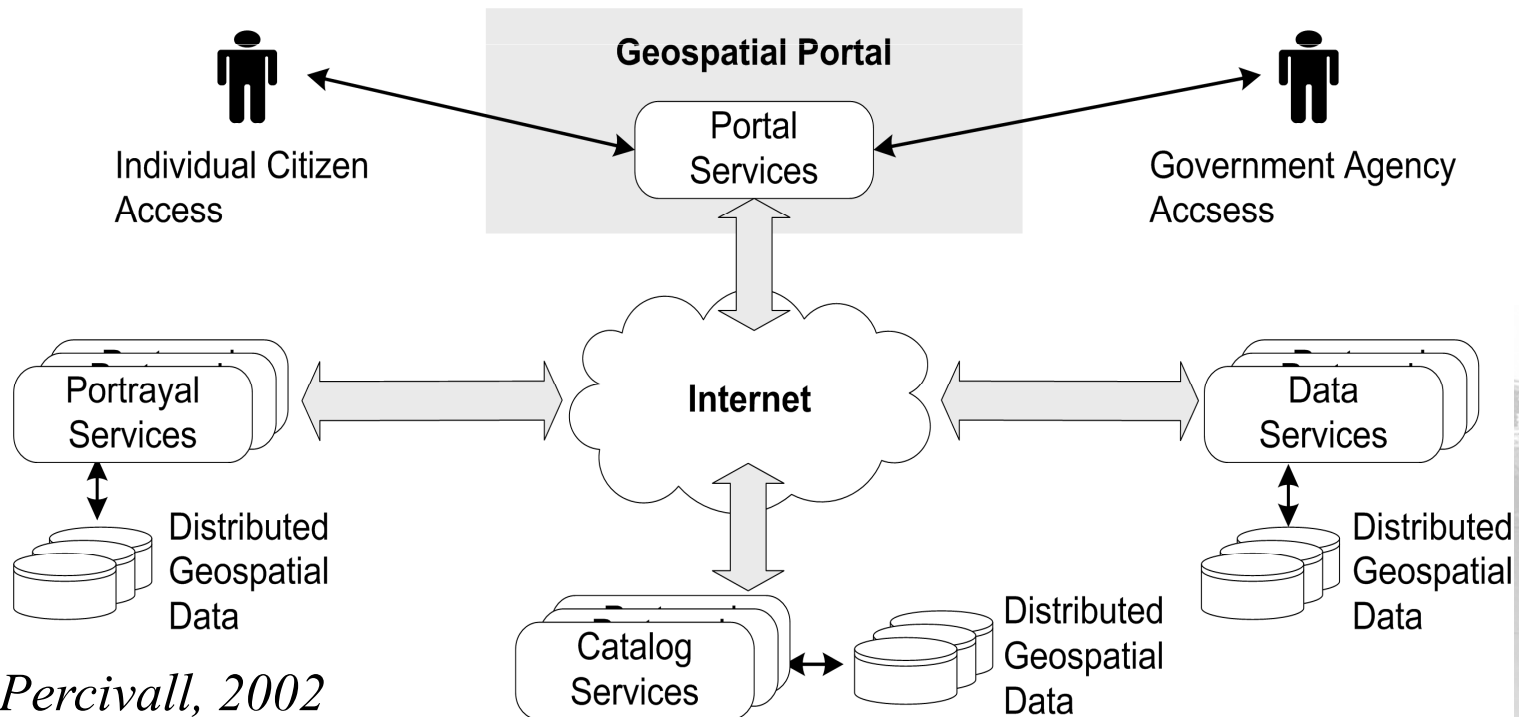
Introducción

 La Web

- ❖ Paraíso para compartir y participar
- ❖ Publicar – buscar – acceder

- Web como plataforma tecnologica para las IDEs

- ❖ Geoportal – el acceso a los contenidos de los IDEs



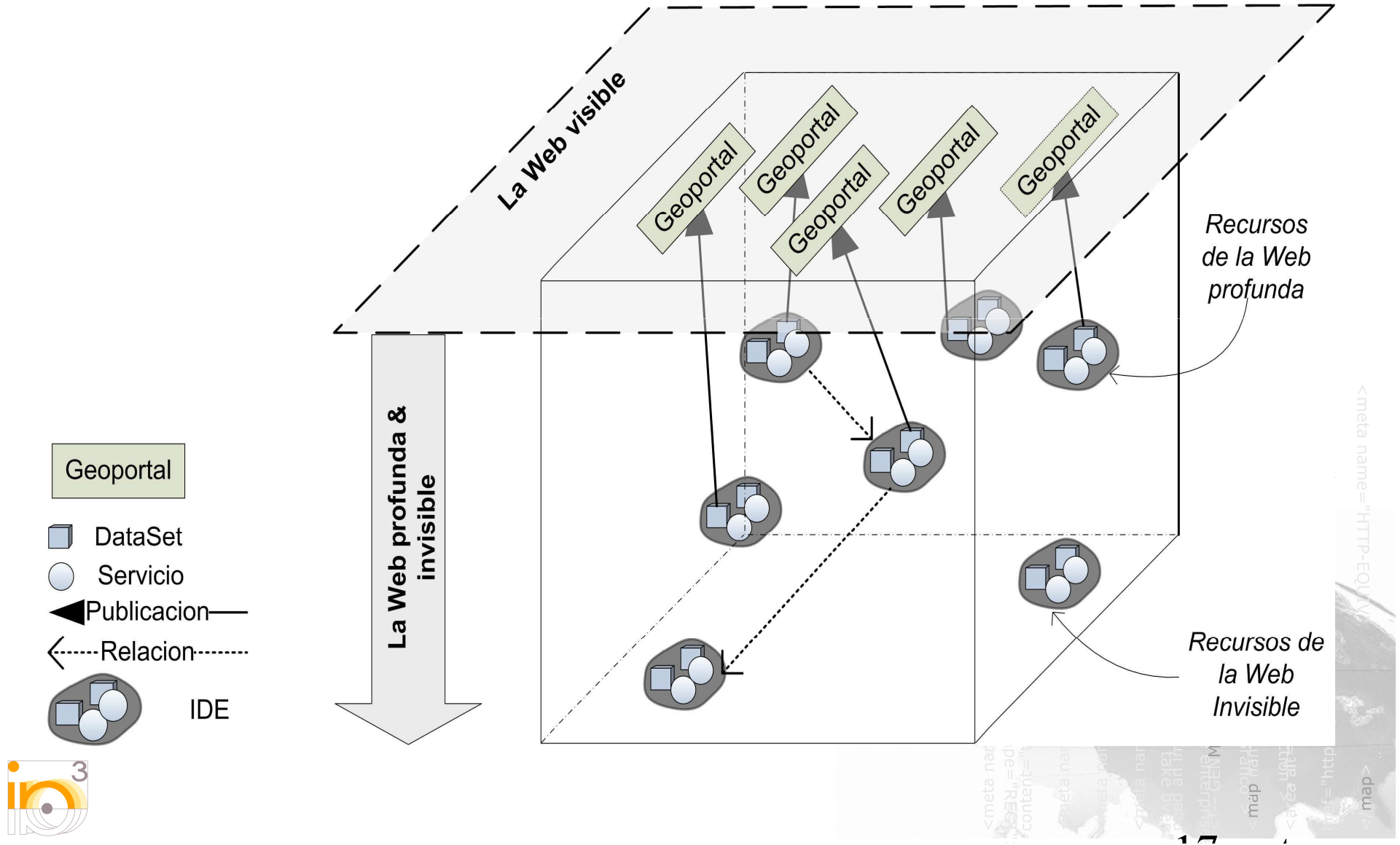
Introducción

- ❑ La Web = Libertad o libertinaje?
 - ❖ Dinamicismo de los contenidos (creación, eliminación y cambio) sin control
 - ❖ Acelerada evolución (paradigma, tecnología)
- ❑ Falta de seguimiento de recomendaciones de W3C
 - ❖ Premisividad e indulgencia (e.g. sistemas de búsqueda, navegadores)

- ❑ Motores de búsqueda genéricos
 - ❖ Eficacia en un entorno dinámico y evolutivo
 - ❖ El escaneo de la superficie
 - ❖ El base del desarrollo de los motores especializados

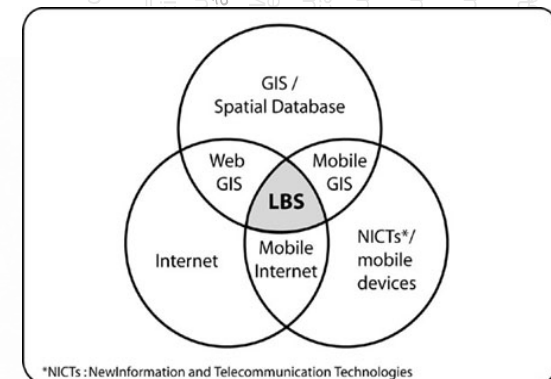


Aproximación simplificada



Visibilidad, porque?

- Visibilidad de los recursos de una IDE
 - Servicio a los ciudadanos
 - Información de referencia
- Visibilidad de los actores de una IDE
 - Posibles clientes
- El mercado
 - Oferta educativa
 - Servicios dedicados
 - LBS
 - publicidad
 - Seguridad
 - atención personalizada



LBS – intersección de las tecnologías (Brimicombe, 2002)

Visibilidad en la Web

□ Documento HTML

- *lingua franca* para publicación del hipertexto
- Metadato – descripción en la cabecera

□ SEO (e.g. google, yahoo)

- Uso del metadato
- Capacidad específica (analysis del contenido y/o del grafo de la Web)

Recomendaciones SEO

- » Uso de metadatos, pero NO SOLO para hipertexto
- » No suelen usar *keywords*
- » Importancia del elemento *title*

□ Los metadatos geográficos

- Falta del consensus (W3C, Metatags.org, Geourl.org, Dublicore)



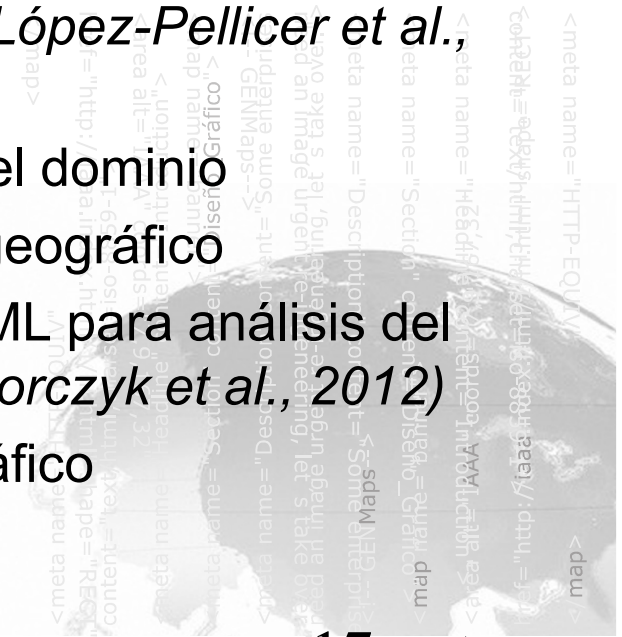
SE-like visibilidad de los recursos de las IDEs

□ Visión de los motores de la búsqueda

- Recuperación automática de los recursos típicos
- Creación del metadato descriptivo
 - funcionalidad extendida

□ Procedimiento

- Recogida de los recursos típicos
- Crawler dedicado a servicios OGC (*López-Pellicer et al., 2011*)
- Análisis manual de los sitios Web del dominio
- Creación automática del metadato geográfico
 - Herramienta dedicada a HTML para análisis del metadato y del contenido (*Florczyk et al., 2012*)
- Evaluación de los metadatos geográfico



Análisis de los sitios Web

□ La muestra de los OWS

- Recogida a lo largo de un año desde Julio 2011
- Filtrado de los servicios Ibéricos
- 50,537 entradas, 6,899 servicios únicos, y 259 dominios únicos
- 192 dominios únicos elegidos de manera aleatoria (>74% de la muestra original)

□ Dominio de un servicio OGC como base de la recuperación de un sitio Web relacionado con una IDE



Análisis de los sitios Web

❑ **Característica general de los elementos desestimados (50.5%)**

- 3.7% - un error del navegador
- 22.2% - las respuestas irróneas (por ejemplo, respuestas vacías, el listado de los ficheros, información del error del servicio o del servidor)
- 10.9% - las paginas generadas por defecto por el servidor/servicio (por ejemplo, pagina de inicio del Tomcat),
- 2.1% - las paginas de compañías dedicadas a software,
- 10.9% - las paginas repetidas



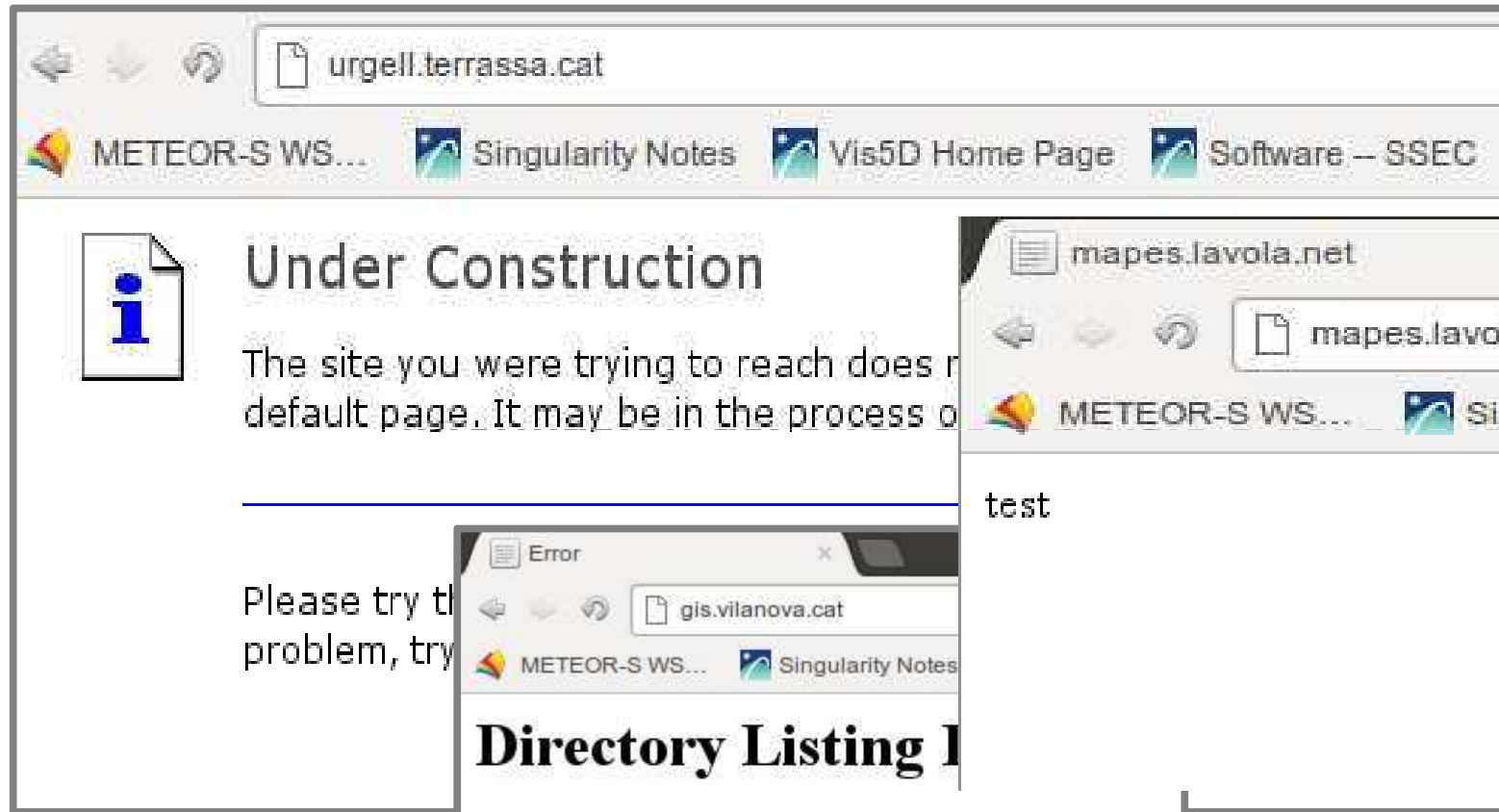
Análisis de los sitios Web

❑ **Característica general de los elementos desestimados (50.5%)**

- 3.7% - un error del navegador
- **22.2% - las respuestas irróneas (por ejemplo, respuestas vacías, el listado de los ficheros, información del error del servicio o del servidor)**
- 10.9% - las paginas generadas por defecto por el servidor/servicio (por ejemplo, pagina de inicio del Tomcat),
- 2.1% - las paginas de compañías dedicadas a software,
- 10.9% - las paginas repetidas



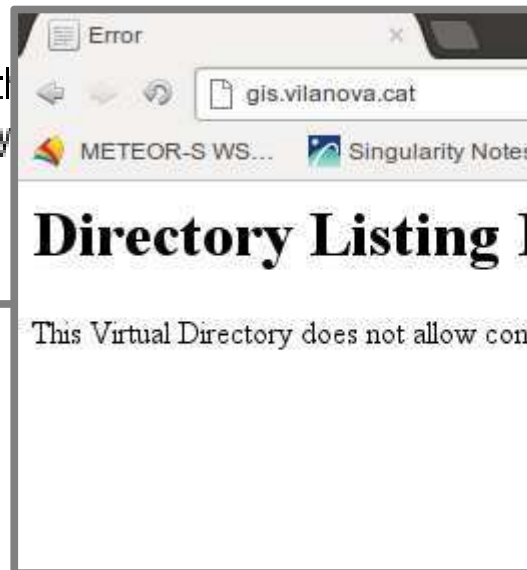
Análisis de los sitios Web



ados



test



Análisis de los sitios Web

❑ **Característica general de los elementos desestimados (50.5%)**

- 3.7% - un error del navegador
- 22.2% - las respuestas irróneas (por ejemplo, respuestas vacías, el listado de los ficheros, información del error del servicio o del servidor)
- **10.9% - las paginas generadas por defecto por el servidor/servicio (por ejemplo, pagina de inicio del Tomcat),**
- 2.1% - las paginas de compañías dedicadas a software,
- 10.9% - las paginas repetidas



Análisis de los sitios Web

Willkommen
Bienvenue
歓迎
Velkommen
Benvenuto
Welkom
Välkommen
Hoş Geldiniz
Üdvözlök
Добро пожаловать

Apache Tomcat/7.0.23

Home Documentation Configuration Examples Wiki Mailing Lists Find Help

Apache Tomcat/7.0.23

The Apache Software Foundation
<http://www.apache.org/>

If you're seeing this, you've successfully installed Tomcat. Congratulations!

Recommended Reading:

- [Security Considerations HOW-TO](#)
- [Manager Application HOW-TO](#)
- [Clustering/Session Replication HOW-TO](#)

Server Status
Manager App
Host Manager

Developer Quick Start

- [Tomcat Setup](#)
- [Realms & AAA](#)
- [Servlet Examples](#)
- [Servlet Specifications](#)
- [First Web Application](#)
- [JDBC DataSources](#)
- [JSP Examples](#)
- [Tomcat Versions](#)

Managing Tomcat

For security, access to the [manager webapp](#) is restricted. Users are defined in:

```
$CATALINA_HOME/conf/tomcat-users.xml
```

In Tomcat 7.0 access to the manager application is split between different users. [Read more...](#)

[Release Notes](#)

Test Page for Apache Installation

If you can see this, it means that the installation of the [Apache web server](#) software on this system was successful. You may now add content to this directory and replace this page.

Seeing this instead of the website you expected?

This page is here because the site administrator has changed the configuration of this web server. Please **contact the person responsible for maintaining this server with questions**. The Apache Software Foundation, which wrote the web server software this site administrator is using, has nothing to do with maintaining this site and cannot help resolve configuration issues.

The Apache [documentator](#) has been included with this distribution.

You are free to use the image below on an Apache-powered web server. Thanks for using Apache!

Powered by **APACHE**

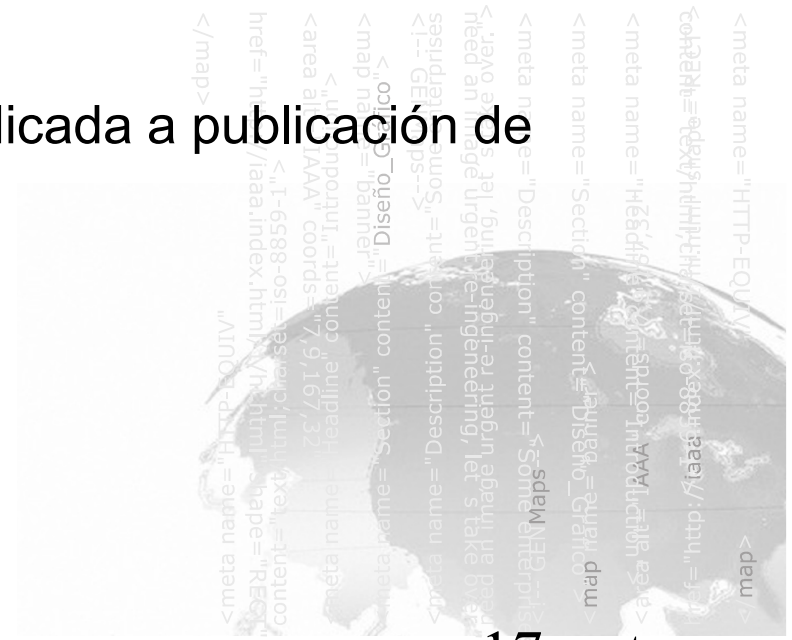
Análisis de los sitios Web

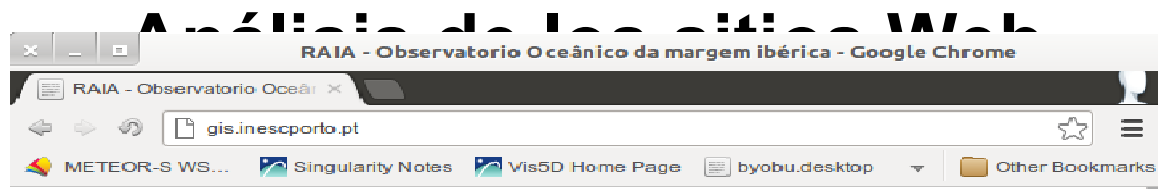
- ❑ Característica general de los elementos identificados como relevante para una IDE (49.5%)
 - 3.7% las paginas generadas de manera automática pero su contenido sugiere que están vinculadas a una IDE
 - 25.5% geoportales de una IDE - organización correspondiente a la funcionalidad
 - 8.9% un visor cartográfico
 - 11.5% parte de un portal dedicada a publicación de recursos geográficos



Análisis de los sitios Web

- ❑ Característica general de los elementos identificados como relevante para una IDE (49.5%)
 - **3.7% las paginas generadas de manera automática pero su contenido sugiere que están vinculadas a una IDE**
 - 25.5% geoportales de una IDE - organización correspondiente a la funcionalidad
 - 8.9% un visor cartográfico
 - 11.5% parte de un portal dedicada a publicación de recursos geográficos

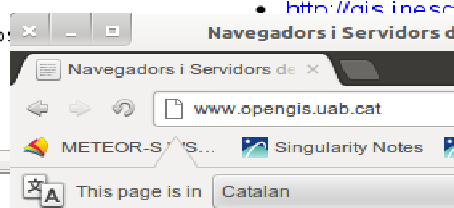




Serviços Disponibilizados

- Web Map service (ncWMS) <http://gis.inescporto.pt>
- WebMap Client Interface (Godiva) <http://gis.inescporto.pt>
- Web Map Service (Geoserver) <http://gis.inescporto.pt>
- Web Feature Service (Geoserver) <http://gis.inescporto.pt>
- Web Coverage Service (Geoserver) <http://gis.inescporto.pt>

Sensor Ob:



Banc de dades de Biodiversitat	http://biodiver.bio.ub.es/biocat/homepage.html	
Natura Digital	http://biodiver.bio.ub.es/naturadigit	
ORCA	http://biodiver.bio.ub.es/orca/	
web assignatura de Coromòfits	http://biodiver.bio.ub.es/Coromofits/home.htm	
VegAna (Vegetation edition and Analysis Tools)	http://biodiver.bio.ub.es/vegana/	



Navegadors de Mapes disponibles

Tria el navegador WMS que vols veure:

- Cartografia Oficial de Catalunya produïda per l'ICC
- Ortofotos de Catalunya
- Mapa de Cobertes del Sòl de Catalunya
- Mapa de Cultius i Aprofitaments de Catalunya
- El Món: cobertes, elevacions i límits administratius
- SATCAT: Servidor d'imatges de Satèl·lit de Catalunya
- Atles Climàtic Digital d'Aragó
- Atles Climàtic Digital de la Península Ibèrica
- Atles Climàtic Digital d'Andorra

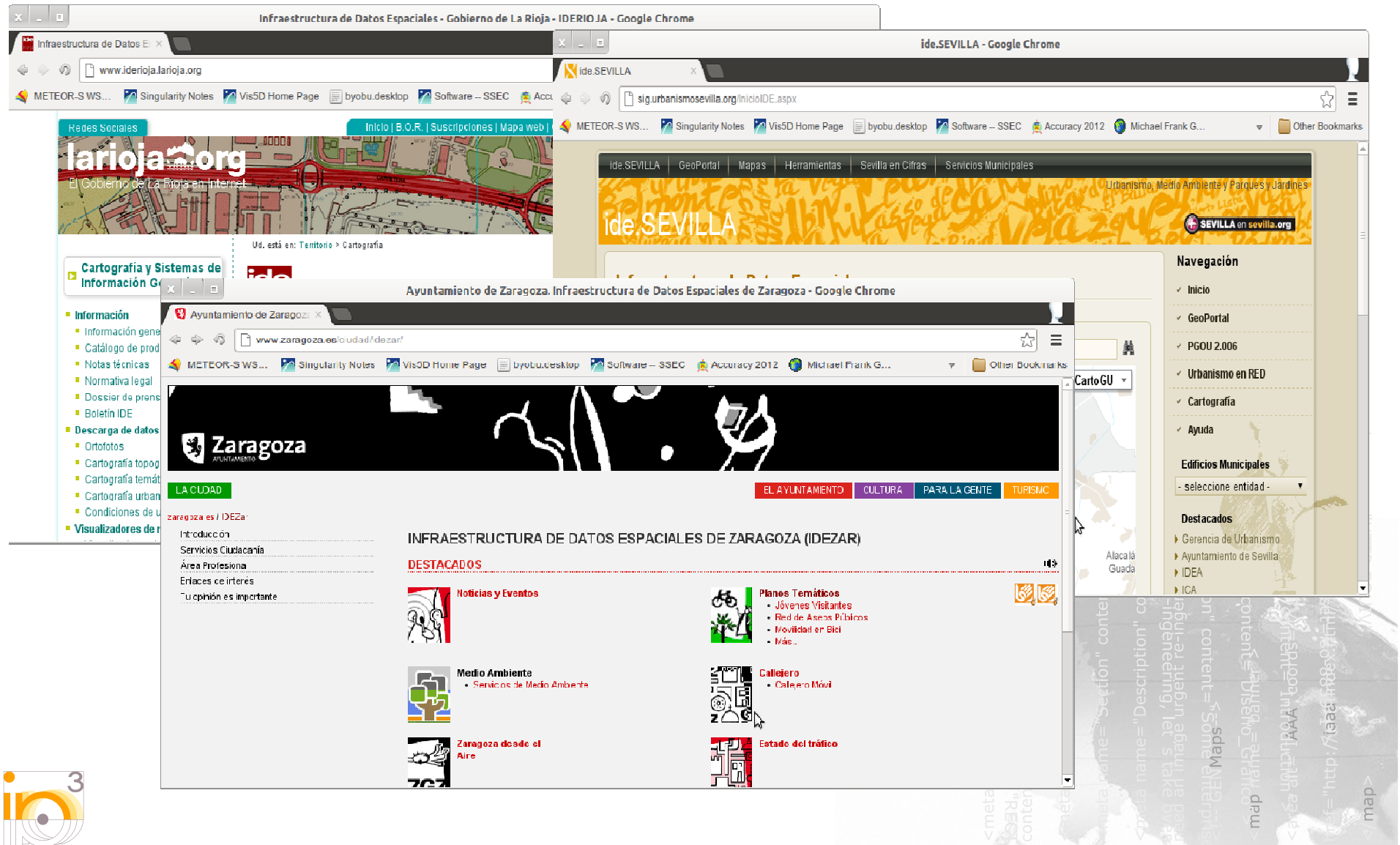


Análisis de los sitios Web

- ❑ Característica general de los elementos identificados como relevante para una IDE (49.5%)
 - 3.7% las paginas generadas de manera automática pero su contenido sugiere que están vinculadas a una IDE
 - **25.5% geoportales de una IDE - organización correspondiente a la funcionalidad**
 - 8.9% un visor cartográfico
 - 11.5% parte de un portal dedicada a publicación de recursos geográficos



Análisis de los sitios Web



Análisis de los sitios Web

The image displays a web browser window with three overlapping web portals. The top-left portal is 'Geoportal IDEC' (Google Chrome), which is a Catalan-language geoportal for the 'Infraestructura de Dades Espacials de Catalunya'. It features a navigation menu with links like 'Inici', 'Visor de Mapes', and 'Catàleg de Georecursos'. The top-right portal is 'Portal IDE-Ebro' (Google Chrome), a Spanish-language portal for the 'Confederación Hidrográfica del Ebro'. It includes a 'Presentación' section and a sidebar with links like '¿Qué es IDE-Ebro?' and '¿Qué puedo hacer?'. The bottom portal is 'GeoPortal Município de Águeda' (Google Chrome), a Portuguese-language portal for the 'Município de Águeda'. It features a map of the region and sections for 'GeoVisualizadores' and 'User login'. The browser's address bar and bookmarks are visible for each portal.

Geoportal IDEC - Google Chrome

delta.icc.cat/geoportal/cat/#slider1=1

METEOR-S... Singularity Notes Vis5D Home Page byobu.desktop Software - SSEC Accuracy 2012 Michael Frank G...

This page is in Catalan Would you like to translate it? Translate Nope

IDEC
Infraestructura de Dades
Espacials de Catalunya

SITE MAP | Cat Usuari

Inici Visor de Mapes Catàleg de Georecursos Geoportals Temàtics Iniciatives Sectorials Iniciatives Locals

Què és IDEC?
MetaD
MetaD web
Geoserveis
Participar
Enllaços
Contactar
Guia d'inici
Privacitat
Novetats v.3.0

geoinformació 225 Proveïdors 520 Serveis geogràfics Plataforma de Recursos

<?xml version="1.0" e
<!-- MetaD Ver. 3.0.4 Exportació
<xml-stylesheet type
<fileIdentifier>

Mapes

La Plataforma de Recursos de Ge...
conjunt d'eines que permeten pu...
web... [més info](#)
[Vols provar la PRG?](#)

Agüeda

Portal IDE-Ebro - Google Chrome

ide-ebro.chebro.es

METEOR-S WS Singularity Notes Vis5D Home Page hynhu.desktop Software - SSEC Accuracy 2012 Michael Frank G... Other Bookmarks

IDE-Ebro

inicio | mapa web | contactar

Bienvenidos | bienvenus | Ve comen

¿Qué es IDE-Ebro? > Presentación

Presentación

La Infraestructura de Datos Espaciales de la Confederación Hidrográfica del Ebro (IDE-Ebro) tiene por objetivo la publicación de los datos de tipo geográfico producidos en la Confederación Hidrográfica del Ebro. A través de portal de IDE-Ebro se puede acceder a dicha información de forma integrada con otro tipo de información geográfica perteneciente a territorio de la cuenca del Ebro y servida por otras IDEs a nivel nacional, regional o local.

Esta infraestructura ha sido desarrollada siguiendo los principios de la Directiva INSPIRE, cuyo objetivo final es la creación de una Infraestructura Europea de Datos Espaciales, basada en la federación de las infraestructuras de datos espaciales de los países miembros. En una primera fase, INSPIRE centra sus esfuerzos en la construcción de infraestructuras de Datos Espaciales en temas ambientales, que entre otras cosas incluyen los datos hidrológicos relacionados con la implantación de la Directiva Marco del Agua (DMA) en cada demarcación hidrográfica. Estos son el núcleo de información de IDE-Ebro, que proporciona información oficial correspondiente a los últimos informes remitidos a la Comisión Europea relativos a los artículos 5, 5 y 8 de la DMA. Junto con esta información, se proporcionan otros conjuntos de datos existentes en la Confederación Hidrográfica del Ebro, no específicos de la DMA, pero de utilidad como cartografía de apoyo.

La Confederación Hidrográfica del Ebro siempre ha procurado poner a disposición pública los datos utilizando los medios tecnológicos disponibles. Así, desde los años 90 la página Web ha servido para la consulta y la descarga de información ampliando su difusión. Con la creación de IDE-Ebro se da un paso más, adaptando esta misma voluntad a la línea tecnológica más actual.

Esta voluntad general de servir, compartir y difundir datos geográficos en el ámbito de nuestra cuenca se concreta en:

GeoPortal Município de Águeda

geoportal.cm-agueda.pt

METEOR-S WS... Singularity Notes Vis5D Home Page byobu.desktop Software - SSEC Accuracy 2012 Michael Frank G... Other Bookmarks

GeoPortal
Município de Águeda

um espaço de divulgação de informação geográfica

Início Conteúdos Mapas IG Downloads Contactos Participação pública Plantas

GeoPortal

Submitted by M on Sun, 06/28/2009 - 00:26

A Câmara Municipal de Águeda disponibiliza um novo espaço de conteúdos geográficos.

Desenvolvido por equipa da Divisão de Tecnologias de Informação, da Câmara Municipal, e assente em software livre e de código aberto, o GeoPortal é um porto de disponibilização de conteúdos, notícias, informação geográfica, mapas e ainda serviços de carácter geoespacial.

AGUEDA
CAMARA MUNICIPAL

GeoVisualizadores

- Infraestruturas
- Participação Pública
- Plano Diretor Municipal
- Plantas de Localização
- IDE - Visualizador de Mapas

User login

Username: *

Search this site:

Search

Geocaching

- Geocaching
- Campo de Iniciação
- Cache de Carvalho
- Cache de Lourizela
- Montedalcobar
- Praia da Talhada
- Sernade do Vouga
- Sobreiro de Belazaima do

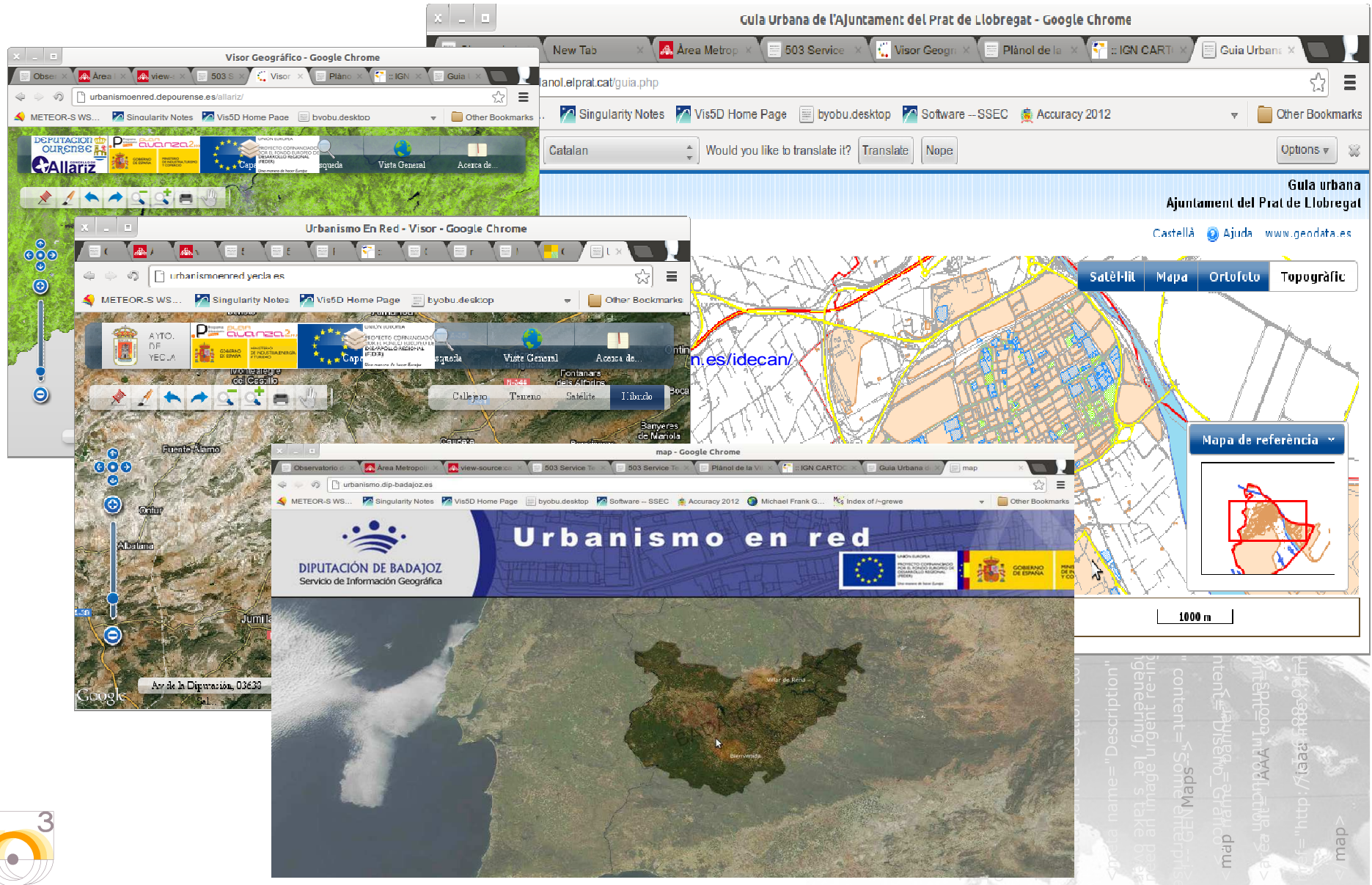
Análisis de los sitios Web

❑ Característica general de los elementos identificados como relevante para una IDE (49.5%)

- 3.7% las paginas generadas de manera automática pero su contenido sugiere que están vinculadas a una IDE
- 25.5% geoportales de una IDE - organización correspondiente a la funcionalidad
- **8.9% un visor cartográfico**
- 11.5% parte de un portal dedicada a publicación de recursos geográficos



Análisis de los sitios Web



Análisis de los sitios Web

- ❑ Característica general de los elementos identificados como relevante para una IDE (49.5%)
 - 3.7% las paginas generadas de manera automática pero su contenido sugiere que están vinculadas a una IDE
 - 25.5% geoportales de una IDE - organización correspondiente a la funcionalidad
 - 8.9% un visor cartográfico
 - **11.5% parte de un portal dedicada a publicación de recursos geográficos**



Análisis de los sitios Web



Análisis de los sitios Web

❑ Característica general de los elementos identificados como relevante para una IDE (49.5%)

- 3.7% las paginas generadas de manera automática pero su contenido sugiere que están vinculadas a una IDE
- 25.5% geoportales de una IDE - organización correspondiente a la funcionalidad
- 8.9% un visor cartográfico
- 11.5% parte de un portal dedicada a publicación de recursos geográficos

❑ Primeras observaciones

- **Casí la mitad de los sitios Web están relacionados con las IDEs**
- **Más de una tercera parte son respuestas “mal gestionadas”**



Generación del metadato geográfico

- ❑ Corpus - las paginas relacionadas con una IDE
- ❑ Característica de los metadatos generados
 - 88% contiene conjunto básico (usado por los buscadores genericos), *titulo* o *descripción*
 - 3.6% contiene metadato geográfico explicito
- ❑ El metadato geográfico generado y su evaluación según cobertura estimada
 - 20% igual
 - 69.5% igual o aceptable (al nivel del país)



Ejemplo para seguir: IDE RIOJA

<head>

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

<meta name="keywords" content="IDERIOJA, Gobierno de La Rioja, Cartografia, Rioja, IDE, Infraestructura Datos Espaciales, Mapas, Visor, Geografía, Territorio, Ortofoto, CAD, Metadatos, Saicar" />

<meta name="description" content="Infraestructura de Datos Espaciales de La Rioja." />

<meta name="title" content="Infraestructura de Datos Espaciales - Gobierno de La Rioja - IDERIOJA" />

<meta http-equiv="Pragma" content="no-cache" />

<meta http-equiv="cache-control" content="no-cache, no-store, must-revalidate"/>

<meta http-equiv="Expires" content="Mon, 01 Jan 1990 00:00:01 GMT" />

<meta name="geo.position" content="42.27189379;-2.28201889" />

<meta name="geo.placename" content="La Rioja, Logroño" />

<meta name="geo.region" content="ES-LO" />

(...)



$$(\dots)$$

```
<link rel="meta" href="http://www.w3c.es/index.rdf" />
```

<link rel="schema.DCTERMS" href="http://purl.org/dc/terms/" />

```
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
```

```
<meta name="DC.title" xml:lang="es" content="Infraestructura de Datos Espaciales - Gobierno de La Rioja - IDERIOJA" />
```

```
<meta name="DC.subject" content="IDERIOJA, Gobierno de La Rioja, Cartografia, Rioja, IDE, Infraestructura Datos Espaciales, Mapas, Visor, Geografia, Territorio, Ortofoto, CAD, Metadatos, Saicar" />
```

```
<meta name="DC.description" xml:lang="es" content="Infraestructura de Datos Espaciales de La Rioja." />
```

<meta name="DC.publisher" content="Gobierno de La Rioja" />

$$\left(\begin{array}{c} \vdots \end{array} \right)$$


Ejemplo para seguir: IDE RIOJA

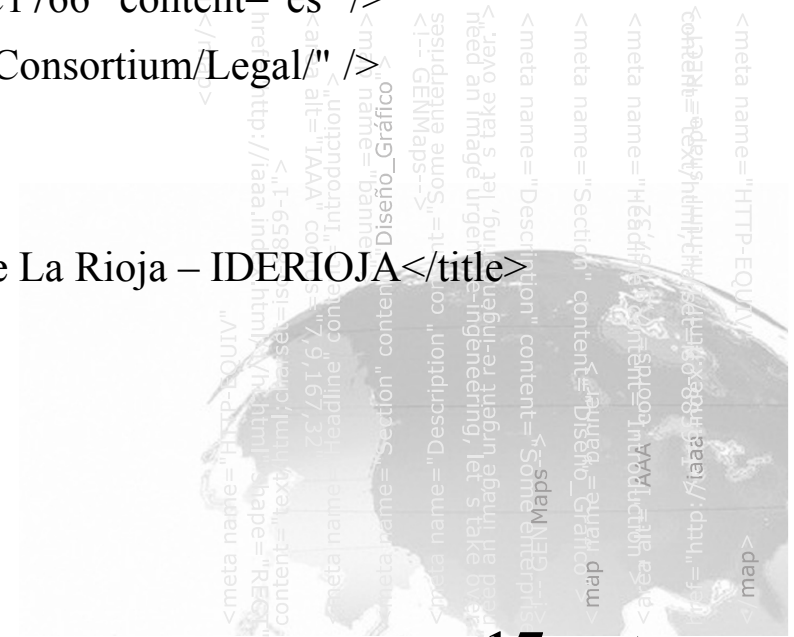
(...)

```
<meta name="DC.date" scheme="DCTERMS.W3CDTF" content="2006-03-10" />
<meta name="DC.type" scheme="DCTERMS.DCMIType" content="Text" />
<meta name="DC.creator" content="Agencia del Conocimiento y la Tecnologia" />
<meta name="DC.format" scheme="DCTERMS.IMT" content="application/xhtml+xml" />
<meta name="DC.identifier" scheme="DCTERMS.URI" content="http://www.w3c.es" />
<meta name="DC.language" scheme="DCTERMS.RFC1766" content="es" />
<meta name="DC.rights" content="http://www.w3.org/Consortium/Legal/" />
<!-- Fin de los metadatos del DC -->
```

```
<title>Infraestructura de Datos Espaciales - Gobierno de La Rioja – IDERIOJA</title>
```

(...)

```
<\head>
```



Conclusiones

❑ Perdida de visibilidad

❑ Malas practicas:

- Falta de metadato completo en la cabecera, especialmente en caso de los visores
- Falta del metadato geográfico
- falta de aprovechamiento del mecanismo de redirección

❏ Preguntas abiertas

- La política de redirección
- Sitios Web de publicación de recursos geográficos (tipos y su identificación automática)



Referencias

- ❑ Percivall, G., 2002. OpenGIS Service Architecture (Version 4.3). The OpenGIS Abstract Specification and ISO/DIS 19119. Geographic information — Services.
- ❑ Brimicombe, A. J., 2002. GIS - Where are the frontiers now?. In: Proceedings GIS 2002., Bahrain. 33-45.
- ❑ López-Pellicer, F. J., Florczyk, A. J., Béjar, R., Muro-Medrano, P. R., Zarazaga-Soria, F. J., 2011. Discovering geographic web services in search engines. Online Information Review 35 (6), 909–927.
- ❑ Florczyk, A.J., López-Pellicer, F.J., Nogueras-Iso, J., Zarazaga-Soria, F. J., 2012. Automatic Generation of Geospatial Metadata for Web Resources, In Press.



