

PRIMERES JORNADES SOBRE GESTIÓ DE LA INFORMACIÓ CIENTÍFICA

Barcelona, **12-13 d'abril de 2012**
Institut d'Estudis Catalans

Segona taula rodona: *Open Access i Open Data*

F. Peset (Universidad Politécnica de Valencia)

<http://blocs.iec.cat/observatori/jgic-2012/>



Guion

- Reflexiones
- Open Access
 - actualidad
 - - futuro
- Open Data
 - definiciones
 - implicaciones
- Retos



De la sociedad red a la sociedad del dato

- Rebeldías
 - Frente a la opacidad, transparencia
 - Frente al monopolio, la reutilización

Es Open todo?

- **Acceso abierto** no es una mera cuestión tecnológica, sino **un movimiento social de largo recorrido**
- PERO también es una cuestión **tecnológica**. Open y reusables son los metadatos (el contenido se sirve en origen). Metadatos descritos en DC, estandarizado y extendido; y el uso de OAI-PMH los expone para crear otros servicios

Ejemplo de descripción en DC: E-LIS Eprints in Library and Information Science

- <head>
- <title>E-LIS - Políticas editoriales en la documentación española e implantación de E-LIS (E-Prints in Library and Information Science)</title>
-
- <link href="http://purl.org/DC/elements/1.0/" rel="schema.DC" />
- <meta content="Políticas editoriales en la documentación española e implantación de E-LIS (E-Prints in Library and Information Science)" name="DC.title" />
- <meta content="Peset Mancebo, Fernanda" name="DC.creator" />
- <meta content="Subirats Coll, Imma" name="DC.creator" />
- <meta content="Barrueco Cruz, José Manuel" name="DC.creator" />
- <meta content="Noverges Doménech, Natividad" name="DC.creator" />
- <meta content="H. Information sources, supports, channels." name="DC.subject" />
- <meta content="[Spanish abstract]

La comunicación ... " name="DC.description" /><meta content="2003-01-01" name="DC.date" />
- <meta content="Conference Paper" name="DC.type" /><meta content="http://eprints.rclis.org/archive/00000416/" name="DC.identifier" />
- <meta content="pdf http://eprints.rclis.org/archive/00000416/01/CALSI.pdf" name="DC.format" />
- <meta content="es" name="DC.language" />
- </head>



Pero no nos engañemos,

- **Libre.** A menudo se mantiene en su forma española para diferenciarlo de lo simplemente gratis. Implica sin restricciones: Redistribución. Reutilización. Ausencia de restricciones tecnológicas.
 - **Abierto.** Desde su contenido inicial, de código accesible y manipulable, incluye actualmente la posibilidad técnica y legal de transformarse contenida en el propio acto de creación-distribución. Pero puede contener limitaciones como atribución
- En estos momentos podríamos llamarlo ...
 - Conocimiento entreabierto
 - Con más de una puerta en cada casa
 - Rendijas de apertura selectiva
 - Se sitúa en el terreno de lo que se podría denominar una **innovación disruptiva**, que transforma procesos y define un nuevo escenario



El Acceso abierto A LAS PUBLICACIONES está actualmente consolidado

- *Extensión del número de depósitos OAI-compliant, incluso a otros ámbitos como el cultural*
- Cambio en el sistema editorial y presión en el nivel de los policymakers y financiadores (mandatos)
 - Interés en que se devuelva a la sociedad el producto de la inversión en I+D: artículo 34 de la *Ley de la Ciencia, la Tecnología y la Innovación* de 2011



Futuro

- Aumentar el depósito efectivo = aumentar en número de items en repositorios o revistas open. Acabar con la paradoja Jekyll/Hyde
- Consolidar los repositorios dentro de las estrategias universitarias
- Acoger todo tipo de material, como docente
- Flexibilizar las formas de propiedad intelectual y declararla. La legislación es bastante más restrictiva que los deseos de los editores y autores científicos.



Pero también... **DATOS DATOS DATOS**

BIG DATA

DATAHUB

Data Avalanche

Data deluge

NIH-OCDE

RIDING THE WAVE

“Los datos deben ser la infraestructura de la economía digital”

Abella (creación empleo e innovación)

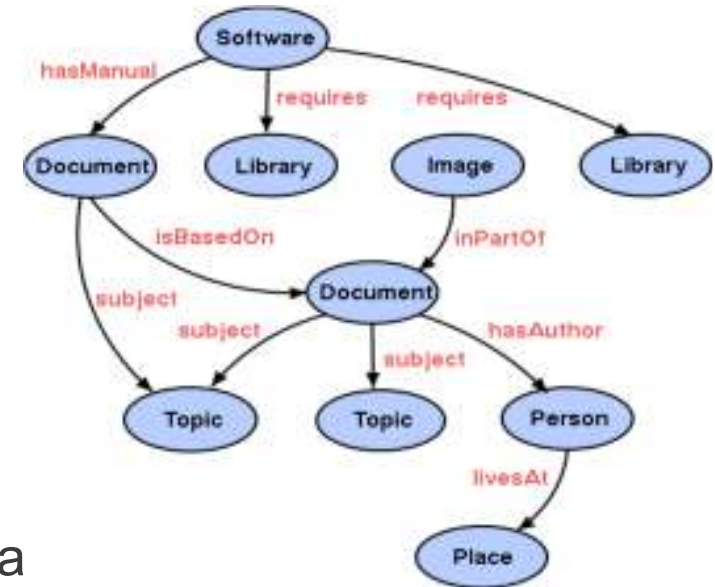


Precisiones terminológicas

- Linked data
- Open data
 - Gubernamentales
 - Derivados de la investigación (raw data)



Linked open data: nivel técnico al nivel de los metadatos, no contenidos (audio, pdf...)

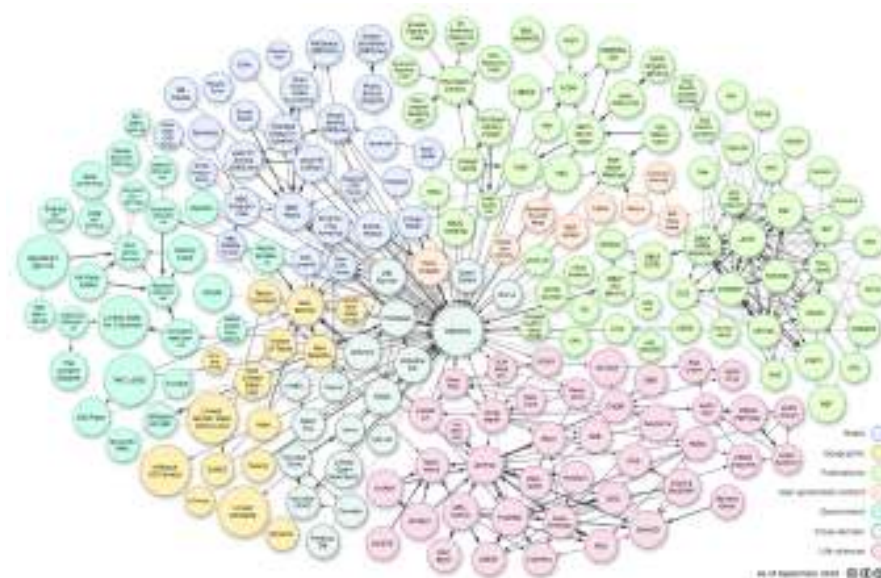


- Se relaciona con la web semántica
 - Proporcionar **acceso a los datos (METADATOS)** de manera que sean fácilmente **reusables**
 - Mejorar el **descubrimiento** de datos relevantes entre un conjunto ingente de datos/conjuntos de datos
 - Crear **aplicaciones** que permita la **integración** desde fuentes de datos inmensas y almacenadas de diferentes maneras

From Data Islands to a Global Data Space

~~Data silos~~

Data cloud



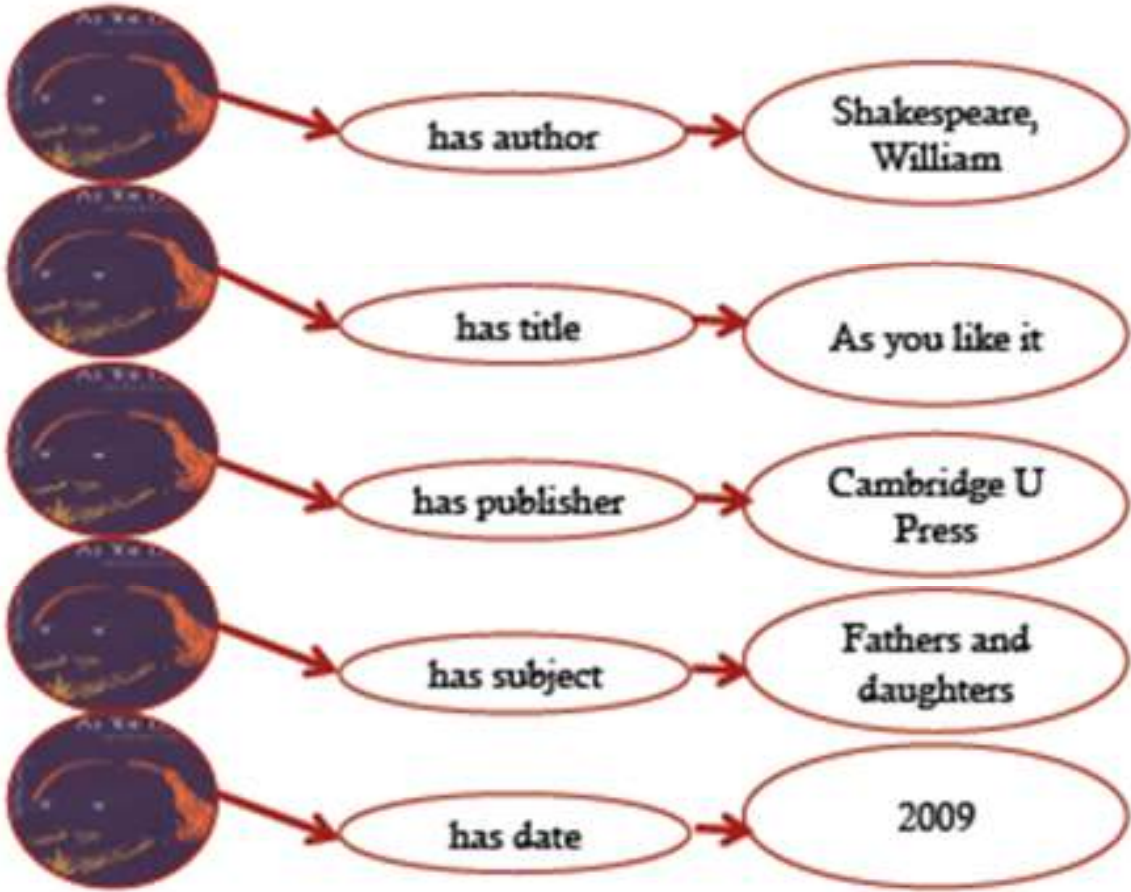
Los principios básicos del Linked Data

En 2006 Berners-Lee definió 4 reglas o recomendaciones

- Ofrecer información sobre los recursos usando RDF; **DESCRIBIR**
- Usar URIs (*Uniform Resource Identifier*) identificando los recursos de forma unívoca; **IDENTIFICAR**
- Usar URIs HTTP para que la gente pueda acceder a la información del recurso; **EXPONER**
- Incluir enlaces a otros URIs, facilitando el vínculo entre distintos datos distribuidos en la web. **ENLAZAR/DESCUBRIR**



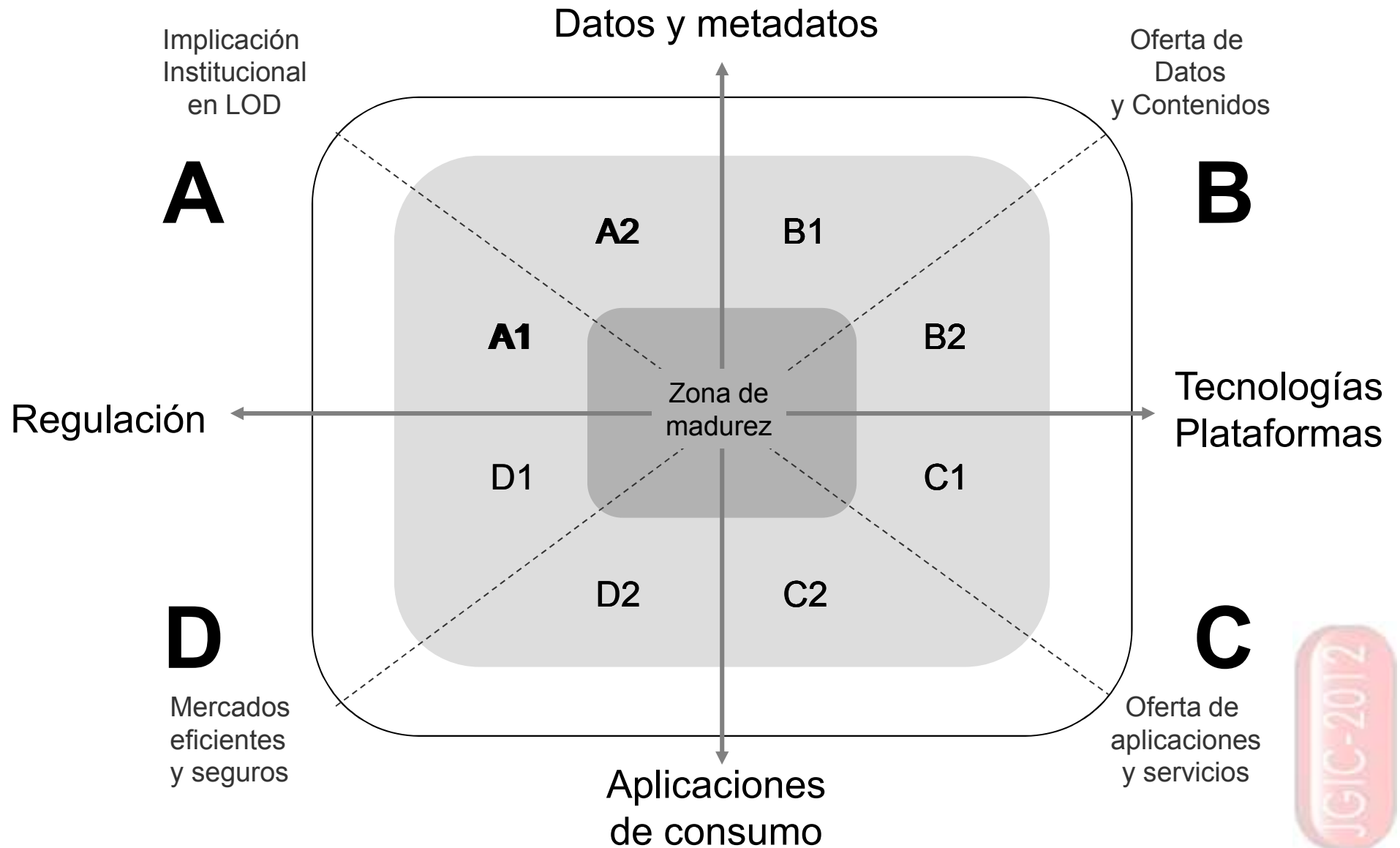
En RDF se DECLARAN sus propiedades



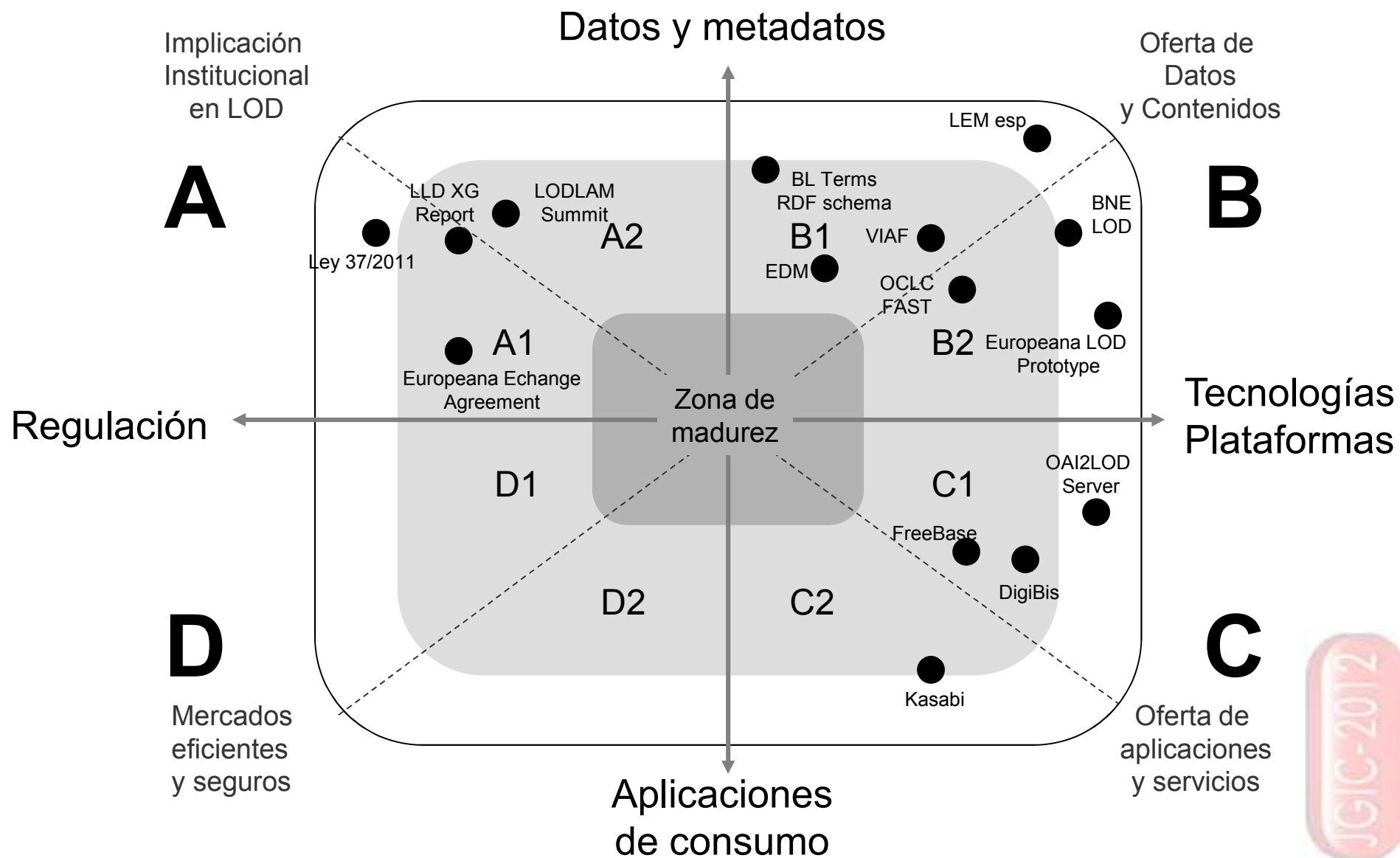
“Triples”



Implantación de linked data (enero 2012)



Implantación de linked data (enero 2012)



OPEN DATA



- Liberar valor social y económico de los contenidos, datos y documentos en poder de las AAPP. (*Commons o procomún: recursos de utilidad social*)
 - Directiva europea de reutilización de información en el sector público+Ley 37/2007
- Relacionado con el e-gov y con la visualización
<http://www.bbc.co.uk/news/business-15748696>
<http://xkcd.com/980/huge/#x=-6618&y=-4530&z=5>
- Los datos se proporcionan como paquetes siguiendo las pautas w3c



datos.gob.es
reutiliza la información pública

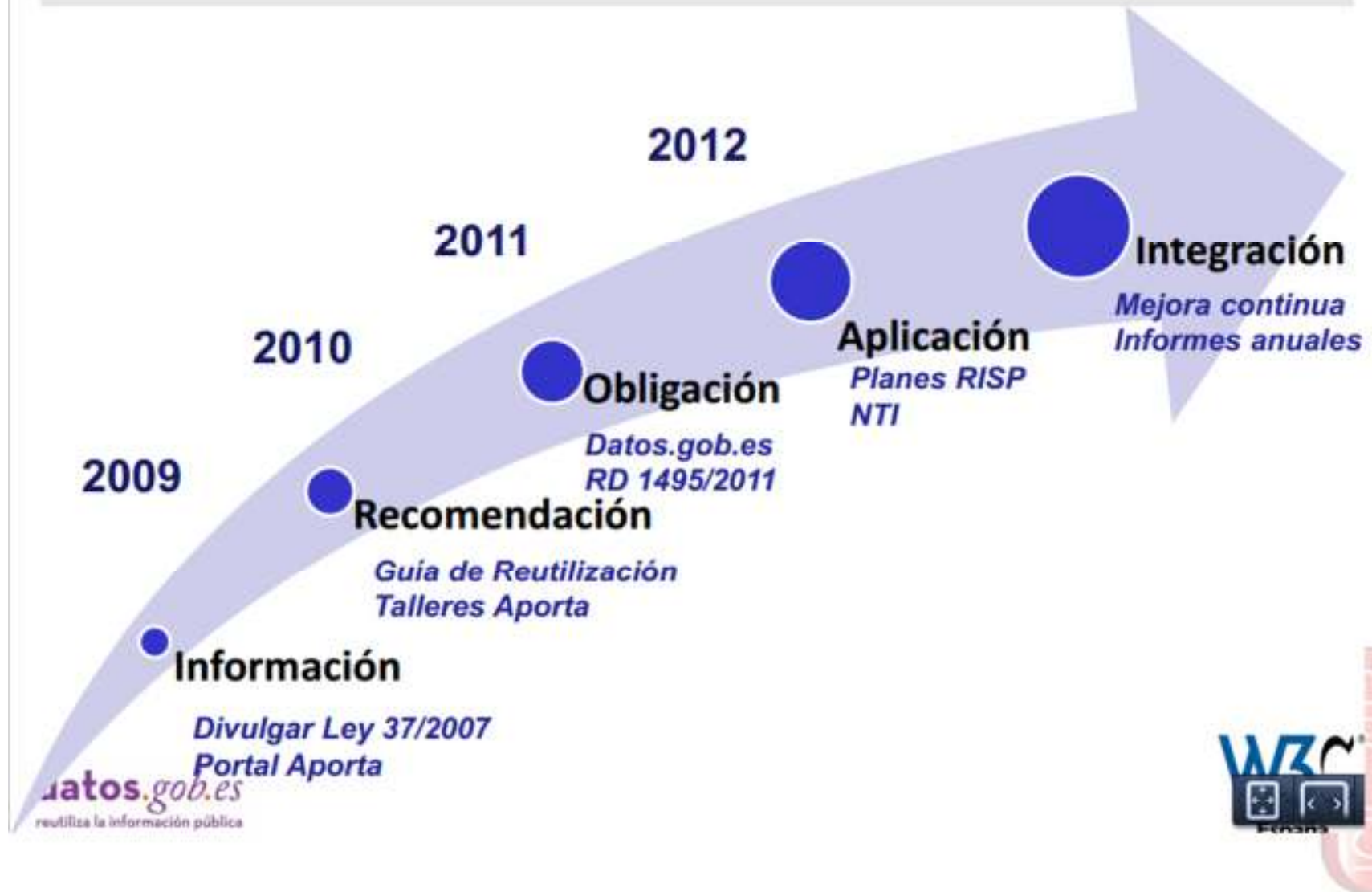
Granada 2012, <http://www.w3c.es/Eventos/2012/DiaW3C/Presentaciones/emilio.pdf>

W3C
España

JGIC-2012

Las AAPP tienen monopolio natural en la producción de algunos datos, pero no tienen por qué tenerlo en exclusividad para su utilización-integración en productos de información.

Un continuo cambio de las Administraciones Públicas.



- **Problemas de los catálogos de datos**
- **77 conjuntos de ciencia y tecnología. Pero muchos son listados, por ejemplo bibliográficos. No datos de investigación (luego los veremos)**



Datos científicos: Open science?

Public investment requires that research data should be made available for verification and reuse

- Escenario en eclosión en USA, en UK, en España (caso paradigmático catalán)
- **Peligro de pérdida, necesidad de preservación estandarizada**
- En el borrador de la nueva directiva europea sobre reutilización los datos científicos se tratan aparte de los datos gubernamentales. Se vinculan al mundo del acceso abierto a las publicaciones, donde tienen su sentido: repositorios
- Todavía no están implantados en las instituciones productoras (españolas), ni los repositorios ni un sistema que gestione su **ciclo de vida integral**



Realidad

1. Datos de los investigadores (pequeños paquetes): se acrecienta la paradoja Jekill y Hide
2. Datos de los productores *públicos* (grandes paquetes): meteorológicos, geoespaciales, cristalográficos...



DIFERENTES FORMAS DE INTERCAMBIO (data-sharing)



DATASHARING formalizado

- 1. material suplementario de revistas: pequeña comunidad de usuarios.
 - Las [políticas de Open Data en ciencia](#) no parecen estar muy extendidas: estudio de las licencias sobre los datos derivados de las investigaciones publicadas revistas *OpenDataScience* (UPV-UV).
- 2. grandes repositorios temáticos o institucionales: reutilización en otros proyectos. Licencias science commons de la Open Knowledge Foundation
 - ODC Attribution License (ODC-BY): obliga a atribuir/reconocer el propietario original
 - PDDL, Public Domain Dedication and License, similar a la CC más permisiva (CC0) sin atribución ni limitación geográfica de uso.
 - Open Database License, ODbL, es más restrictiva, pero permite compartir –copiar, distribuir y usar la base de datos-, producir trabajos sobre ella y modificar, transformar y construir otros productos sobre la base de datos. Con esta última, se mantiene la mención del origen (Attribution) y obliga a ceder el nuevo producto con esta misma licencia (Share-Alike).



Beneficios

- Linked open data
 - Otra dimensión gobernada por máquinas “inteligentes”
- Open data
 - Mercado de trabajo ligado a innovación
 - Redescubrimiento
 - Igualdad de oportunidades
- Open data científico: raw data y grandes repositorios
 - Validar investigaciones
 - Innovar mediante el reuso



Conclusión

- Mercado eficiente → ampliación de agentes. Existen productores de datos (AAPP, Bibliotecas, etc.) y agentes reutilizadores, que es donde se percibe la innovación y la creación de empleo
- Productores:
 - Importancia del uso de estándares globales como RDF. Dificultad de migrar MARC-RDF
- Consumidores:
 - Tecnologías incipientes
 - Competencias en Data curation para conocer qué son Conjuntos de Datos de Calidad



The End
mpesetm@upv.es

Openess is a warm gun (John Lennon)

Awake is the new sleep

